MSc Thesis Applied Mathematics

SEMIPARAMETRIC DEPENDENCE MODELING IN ILLIQUID FINANCIAL MARKETS USING MULTIVARIATE ZERO-INFLATED GARCH-X MODELS

Case study on Voluntary Carbon Credits
in collaboration with Rabobank

Job Vlak

TU Delft

Rabobank

Semiparametric dependence modeling in illiquid financial markets using multivariate zero-inflated GARCH-X models

by

J. L. Vlak

to obtain the degree of Master of Science in Applied Mathematics, at the Delft University of Technology, to be defended publicly on the 29th of August at 12:45.

Student number:	4859758	
Study:	Applied Mathematics	
Faculty:	Faculty of Electrical Engineering, Mathematics and Computer Science	
Project duration:	January 27, 2025 – August 29, 2025	
Supervisor:	Dr. A. F. F. Derumigny	
Thesis committee:	Dr. G. F. Nane,	TU Delft, chair
	Dr. A. F. F. Derumigny,	TU Delft, responsible supervisor
	L. S. Boerop,	Rabobank, co-supervisor
	M. J. F. Van Aken,	Rabobank, co-supervisor
	E. J. H. Van de Ven,	Rabobank, co-supervisor

TU Delft

Rabobank

Semiparametric dependence modeling in illiquid financial markets using multivariate zero-inflated GARCH-X models

Cover design by Jesse Lokin

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

Acknowledgements

This thesis concludes my time as a student of Applied Mathematics at Delft University of Technology. I came to Delft in 2018, and looking back, the city has become a special place to me. This is due to the many memorable experiences, not only with friends, but also at the university. I got to know the university in a different way during my year on the student council and later while advising the executive board on student matters. In terms of my studies, a defining moment came during my bachelor thesis, also under the supervision of Alexis Derumigny, where I was first introduced to dependence modeling in financial markets. I thoroughly enjoyed the project, learned a lot, and already benefited from Alexis's exceptional guidance at that stage. The experience convinced me to continue this line of work in the master's, which I was able to do with this thesis project at Rabobank in January.

First of all, I would like to take this opportunity to thank Alexis for his outstanding supervision throughout this project. I greatly appreciated your continuous engagement and the time you have dedicated. I truly enjoyed our discussions, whether we were exploring different approaches to tackle theoretical challenges or dealing with more applied aspects of the work. Beyond mathematics, I also valued our conversations on topics such as education, politics, or the French and Dutch culture. Altogether, these moments made our collaboration both special and inspiring to me.

From the Rabobank side, I want to thank Leon Boerop, Modesta van Aken, and Emma van de Ven, who made me feel welcome at Rabo Carbon Bank from the very first moment. Our regular sparring sessions taught me a lot about the complex world of carbon markets. At the same time, you gave me great freedom to conduct my research, while still being closely involved and even attentively following the mathematical presentations. The combination of hard work and personal connection made Rabo Carbon Bank a special place, and I will truly miss the collaboration and the people there.

Furthermore, I would like to thank the last member of the thesis committee, Tina Nane. Although we had little contact during this thesis project, I followed several of your courses, which I always enjoyed. You have been very approachable throughout my studies and showed always genuine interest in students, both in mathematical and non-mathematical matters, which I really appreciated.

Finally, I would like to thank my family and friends for their continuous support. In particular, I am grateful to the friends with whom I worked hard last summer – your company made missing the summer holiday much easier. Through this project, I have become even more enthusiastic about dependence modeling and statistics in general.

Job Vlak
Amsterdam, August 2025

Abstract

Voluntary carbon markets are at an early stage of development, characterized by low and irregular trading frequency. Such limited activity results in extended periods of unchanged prices and a high incidence of zero returns, making voluntary carbon markets a typical example of illiquid financial markets. To model dependence in such settings, we propose a multivariate zero-inflated GARCH-X model. The model extends the existing zero-inflated GARCH model to a multivariate setup, incorporating both a GARCH-X component with binary trading indicators as exogenous covariates and a time-step specification that updates only when trades occur.

The multivariate extension incorporates two different types of dependence. First, cross-dependence in trading activity is modeled using Markov networks applied to binary trading indicators. Second, cross-dependence in returns is analyzed using a copula-GARCH framework applied to residuals, with residuals corresponding to zero returns treated as undefined values. To make copula methods applicable to zero-inflated data, we introduce a joint probability integral transform approach. In this construction, the univariate marginals are defined conditional on the simultaneous trading activity of each asset, rather than conditioning only on each asset's own trading activity. We prove that this method yields a consistent copula estimator when applied to the subset of observations where all assets have simultaneous non-zero trading activity. Dependence is quantified using both unconditional and conditional Kendall's tau, estimated via kernel-based methods.

Theoretical results include consistency and asymptotic normality of the quasi-maximum likelihood estimator for the multivariate model under stationary covariates. The empirical study covers seven voluntary carbon credits and six conventional financial assets. We find no significant dependence between carbon credits and conventional assets, but observe strong correlation within the carbon market, especially among nature-based credits. Results suggest that voluntary carbon markets may operate independently of more liquid assets and are influenced by peer pricing due to a lack of standardization.

Keywords: Zero-inflated GARCH, Copula-GARCH, Illiquid markets, Voluntary Carbon Credits, Dependence modeling

Glossary

List of abbreviations

Symbol	Definition
Mathematical Abbreviations	
ACF	Autocorrelation Function
a.s.	Almost surely
CDF	Cumulative Distribution Function
CLT	Central Limit Theorem
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
i.i.d.	Independent and identically distributed
LLN	Law of Large Numbers
MRF	Markov Random Field
PACF	Partial Autocorrelation Function
PDF	Probability Density Function
PIT	Probability Integral Transform
QMLE	Quasi-Maximum Likelihood Estimator
WN	White Noise
Carbon Market Abbreviations	
CDM	Clean Development Mechanism
CO_2eq	Carbon Dioxide Equivalent
CORSIA	Carbon Offsetting and Reduction Scheme for International Aviation
EU ETS	European Union Emissions Trading System
ETS	Emissions Trading Scheme
GHG	Greenhouse Gas
ICVCM	Integrity Council for the Voluntary Carbon Market
JI	Joint Implementation
MRV	Monitoring, Reporting, and Verification
NBS	Nature-Based Solutions
NDC	Nationally Determined Contribution
REDD+	Reducing Emissions from Deforestation and Forest Degradation
SBTi	Science Based Targets Initiative
VCC	Voluntary Carbon Credit
VCM	Voluntary Carbon Market
VCMI	Voluntary Carbon Markets Integrity Initiative
VCS	Verified Carbon Standard

List of symbols

Probability and statistics framework

Symbol	Definition
$\mathbb{N}, \mathbb{Z}, \mathbb{R}$	Sets of natural numbers, integers, and real numbers, respectively
Ω	Sample space, i.e. the set of all possible outcomes
$\mathcal{F}$	σ-algebra
$(\mathcal{F}_n)_{n\geq 1}$	Filtration, i.e. a sequence of σ-algebras
$\mathbb{P}$	Probability measure
X_i	General random variable for the i-th element in a sequence
$\mathbf{X}_i$	General random vector for the i-th element in a sequence
$\overline{X}$	Sample mean of a random variable
θ	General parameter
θ^*	True general parameter
Θ	General parameter space
$\text{int}(\Theta)$	Interior of a parameter space Θ
$F_X(x)$	Marginal cumulative distribution function
$F_{X,Y}(x, y)$	Joint cumulative distribution function
$\Phi(\cdot)$	Standard normal cumulative distribution function
$\phi(\cdot)$	Standard normal probability density function
$\mathcal{H}_0$	Null hypothesis
$\mathcal{H}_1$	Alternative hypothesis
$\xrightarrow{d}$	Convergence in distribution
$\xrightarrow{p}$	Convergence in probability
$\xrightarrow{a.s.}$	Almost sure convergence
$\xrightarrow{L^p}$	Convergence in L^p
κ	Kurtosis coefficient
$(M_n, \mathcal{F}_n)_{n\geq 1}$	Martingale
$(Y_n, \mathcal{F}_n)_{n\geq 1}$	Martingale Difference Sequence

General time series

Symbol	Definition
$(X_t)_{t\geq 1}$	Sequence of random variables
r_t	Return value at time t
ϵ_t	Innovation term at time t
η_t	Innovation term at time t, often i.i.d and standard normally distributed
σ_t^2	Conditional variance at time t, i.e. $\text{Var}[X_t \mid \mathcal{F}_{t-1}]$
ω	Constant term representing the long-term average level of the conditional variance
α	Coefficient measuring the impact of past squared returns
β	Coefficient measuring the persistence of past conditional variances
γ	Coefficient of the exogenous variable

Zero-inflated GARCH-X model

Symbol	Definition
$r_{j,t}$	Return value of asset j at time t
$\sigma^2_{j,t}$	Conditional variance for asset j at time t, evaluated at the true parameter
$\sigma^2_{j,t}(\cdot)$	Conditional variance for asset j at time t, defined recursively and evaluated at $\boldsymbol{\theta}_j$
$I_{j,t}$	Exogenous variable, binary trading indicator taking values in $\{0, 1\}$
$\eta_{j,t}$	Innovation term at time t for asset j, i.i.d but no distribution
$\alpha_{j,0}$	Parameter of the long-term average level of the conditional variance for asset j
$\alpha_{j,1}$	Parameter measuring the impact of past squared returns for asset j
$\alpha_{j,2}$	Parameter measuring the persistence of past conditional variances for asset j
$\alpha_{j,3}$	Parameter of the exogenous variable for asset j
$\alpha^\star_{j,k}$	True parameter for $k \in \{0, 1, 2, 3\}$ for asset j
$\boldsymbol{\theta}_j$	Parameter vector associated with asset j consisting of $\alpha_{j,0}, \alpha_{j,1}, \alpha_{j,2}, \alpha_{j,3}$
$\boldsymbol{\theta}^\star_j$	True parameter vector associated with asset j consisting of $\alpha^\star_{j,0}, \alpha^\star_{j,1}, \alpha^\star_{j,2}, \alpha^\star_{j,3}$
$\boldsymbol{\Theta}_j$	Parameter space associated with asset j
$\mathcal{T}_j$	Time index set for asset j, defined as $\{T_j, \ldots, T_{\text{final}}\}$
$\mathcal{T}^{\text{multivariate}}$	Time index set for all assets starting at the first common trading day
T_j	First time index at which asset j exhibits a non-zero return
T_{final}	Last time index in the sample, the final day of retrieved data (common to all assets)
N_j	Total number of non-zero returns for asset j, i.e. sum of returns such that $I_{\{r_{j,t}\neq 0\}} = 1$
$\mathcal{L}_{\log,j}(\cdot)$	Quasi-log-likelihood function associated with asset j

Dependence modeling

Symbol	Definition
$\eta_{j,t}$	Residual of asset j at time t
$\widehat{\eta}_{j,t}$	Estimated residual of asset j at time t
C	Copula distribution function
$\widehat{C}$	Estimated copula distribution function
c	Copula density function
$U_{j,t}$	Uniform pseudo-random variable for asset j at time t obtained via PIT
$\widehat{U}_{j,t}$	Estimated uniform pseudo-random variable for asset j at time t
$U_{j,t}^{\{I_{j,t}=1\}}$	Marginal distribution of $U_{j,t}$ conditional on trading activity of asset j
$U_{j,t}^{\{I_{i,t}=1, I_{j,t}=1\}}$	Marginal distribution of $U_{j,t}$ conditional on simultaneous trading of assets i and j
$\widehat{C}^{\text{marginal}}$	Copula estimated under the marginal conditioning approach
$\widehat{C}^{\text{pairwise}}$	Copula estimated under the joint (pairwise) conditioning approach
τ_{X_1,X_2}	Kendall's tau of X_1 and X_2
$\widehat{\tau}_{X_1,X_2}$	Estimated Kendall's tau of X_1 and X_2
$\tau_{X_1,X_2\|Z}$	Conditional Kendall's tau of X_1 and X_2 conditional on $\mathbf{Z} = \mathbf{z}$
$\widetilde{\mathcal{T}}_j$	Filtered time index including only periods with defined residuals of asset j
$\widetilde{\mathcal{T}}_{ij}$	Filtered time index including only periods with defined residuals of both assets i and j
$\widetilde{\mathcal{T}}_{\mathcal{A}}$	Filtered time index including only periods with defined residuals for simultaneous trading of all assets in set $\mathcal{A}$

Contents

List of Figures

List of Tables

1 Introduction

The global climate crisis has created an urgent need for decarbonization across all industries. Under the Paris Agreement, limiting global warming to 1.5°C requires a 50% reduction in emissions by 2030 and net-zero emissions by 2050 [5, 9]. Achieving these targets demands not only direct emission reductions but also innovative financial mechanisms that drive climate progress.

Carbon markets have emerged as a promising tool in the fight against climate change by assigning a price to carbon emissions and incentivizing projects that reduce or remove greenhouse gases. These markets facilitate the trading of carbon credits, each representing the reduction or removal of one metric ton of CO_2 equivalent elsewhere. In this way, participants can meet climate targets more cost-effectively [30].

In this context, financial institutions are becoming key players in scaling carbon markets. Rabobank, one of the largest Dutch banks, has taken a leading role with the launch of Rabo Carbon Bank in 2021. The bank engages in both the development and brokerage of voluntary carbon credits. It generates nature-based removal credits via programs like Acorn, which focus on smallholder agroforestry. At the same time, it facilitates transactions by connecting credit suppliers to corporate buyers. This research is carried out in collaboration with Rabobank. The bank provides access to proprietary trading platforms and historical transaction data, which are used in the empirical part of this study. Additionally, it offers domain expertise to support the interpretation of results, particularly in relation to market trends, regulation, and liquidity constraints.

There are two principal types of carbon markets: compliance markets and voluntary carbon markets (VCMs). Compliance markets, such as the EU Emissions Trading System, operate under legal mandates requiring emitters to hold allowances, official permits to emit one ton of CO_2, that match their verified emissions. In contrast, VCMs are not legally binding; they facilitate the trade of carbon offsets among firms and individuals pursuing voluntary climate commitments [3, 30]. Voluntary carbon credits are offsets rather than allowances, meaning they compensate for emissions elsewhere and do not grant the right to emit. Despite being smaller in scale, VCMs are rapidly growing. Demand for voluntary carbon credits is projected to reach 1.5–2 $GtCO_2$ annually by 2030, a big increase from under 0.1 Gt in 2020 [9]. Moreover, policy developments underscore the rising relevance of VCMs. For instance, the EU has announced its intention to allow up to 3% of its 2040 emissions reduction target to be met through international carbon credits, potentially amounting to 300–400 million credits [53].

Despite the growing interest, VCMs face several structural challenges that complicate financial modeling. A primary challenge is market illiquidity. Trading in carbon credits is largely conducted over-the-counter, with infrequent transactions, limited price transparency, and heterogeneous projects that vary across methodologies, locations, and certification standards. Consequently, many carbon credit types exhibit days without trading, leading to constant prices and zero returns, resulting in semi-continuous time series [9, 14]. A second issue is the limited availability and the low quality of historical data. This limitation is illustrated by the fact that the voluntary carbon market only reached an annual transaction volume of approximately $2 billion in 2021 [3], a relatively modest figure compared to other financial markets, and has since experienced a decline through 2024. Historical price data are often sparse or inconsistent, driven by fluctuations in trading volumes, evolving certification standards, and vintage-related differences. Although the average voluntary carbon credit price rose to $6.37 in 2024 - more than twice the 2020 average - overall illiquidity in the market remains [3]. While these conditions present challenges for financial modeling, they also show the need for new methods tailored to the unique characteristics of the carbon market. This represents a promising area for research.

Illiquidity in financial time series appears as zero-inflated data: a high proportion of time points where returns are exactly zero. The frequency of zero returns is often used as a proxy for asset illiquidity, with a larger fraction indicating lower market activity [44]. Let $r_{j,t}$ denote the return of asset j at time t. Then, for illiquid assets, it holds that $\mathbb{P}(r_{j,t} = 0) > 0$ for some t. Essentially, zero-inflated data have a mixed distribution: a positive probability mass at zero and a continuous distribution for non-zero values. Standard time series models such as GARCH models, which assume continuous-valued returns with no point mass at zero, are thus misspecified in this context. This misspecification can bias volatility estimation, particularly by overstating persistence or understating variance during inactive periods [25, 60].

Despite this, it has been shown that GARCH parameters can still be consistently estimated by quasi-maximum likelihood under certain conditions when zero returns are present. For instance, Escanciano proved consistency and asymptotic normality of the QMLE for a broad class of GARCH models, provided the zero-return process is stationary [21].

Several extensions to classical GARCH models have been proposed to handle zero returns. A common approach is to introduce a two-part structure: one component handles the zero returns in a certain way, and the other models the distribution of non-zero returns. For example, Sucarrat and Escribano addressed the undefined log-volatility problem in log-GARCH models by treating zero returns as missing and imputing them via an ARMA-based method [60]. More generally, Sucarrat and Grønneberg introduced a class of models in which the probability of zero returns is modeled as time-varying, rather than constant [61]. Francq and Sucarrat proposed an univariate "zero-adjusted" GARCH framework that augments the GARCH model with a binary trading indicator I_t [25]. Returns are modeled as $r_t = I_t \cdot \sigma_t(\boldsymbol{\theta}) \cdot \eta_t$, where $I_t \in \{0, 1\}$ is a binary indicator of trading activity, equal to one if a non-zero return is observed. The function $\sigma_t(\boldsymbol{\theta})$ denotes the conditional volatility evaluated at parameter vector $\boldsymbol{\theta}$, and η_t is a standardized i.i.d. innovation. Their modified QMLE remains consistent and asymptotically normal. Relatedly, Hudecová and Pešta developed hurdle-GARCH models for nonnegative time series with many zeros [39]. In their setting, returns follow a mixed discrete-continuous distribution: a structural zero may occur with a time-varying probability, and if non-zero, the return follows GARCH dynamics. Importantly, their model allows for dependent

zero occurrences, capturing the tendency of zeros to cluster in illiquid periods.

Recent work has extended GARCH modeling to multivariate settings. Xu proposed a multivariate Exponential GARCH (EGARCH) model designed to accommodate zero returns by regularizing zero observations to ensure a well-defined log-volatility process and incorporating asymmetric volatility spillovers across assets [69]. While the theoretical properties of the quasi-maximum likelihood estimator are not established, simulation results suggest consistency and asymptotic normality. Other multivariate GARCH models appear mainly in the context of count time series, particularly through Integer-GARCH (IGARCH) frameworks [57, 51].

To date, no multivariate GARCH-X framework has been developed that accounts for zero-inflated return series in illiquid markets. This thesis addresses that gap by introducing a multivariate zero-inflated multivariate GARCH-X model that captures the characteristics of illiquid financial assets such as voluntary carbon credits. Our model extends the univariate zero-adjusted GARCH framework developed by Francq and Sucarrat [25], where we have introduced three novel adaptations:

(i) Multivariate binary trading indicators $I_{j,t}$, which explicitly model the occurrence of zero returns, are modeled jointly using a Markov network to capture cross-dependence in trading activity across assets.
(ii) The lagged indicator $I_{j,t-1}$ enters the conditional variance equation as an exogenous covariate. This captures the intuition that the occurrence of a trade at time $t-1$ may influence volatility at time t.
(iii) An irregular time-step specification, denoted by t^*, in which the conditional volatility recursion is applied only at time points where trading occurs, i.e. when $I_{j,t} = 1$. This time-filtered updating mechanism ensures that volatility is not artificially smoothed over inactive periods, and that it reflects more accurately the sparse and uneven structure of return observations in illiquid markets.

Then, the returns of asset j follow a zero-inflated GARCH(1,1)-X model, specified as follows:

$$\begin{cases} r_{j,t} = I_{j,t} \cdot \sigma_{j,t}(\boldsymbol{\theta}_j^\star) \cdot \eta_{j,t}, \\ \sigma_{j,t}^2(\boldsymbol{\theta}_j^\star) = \alpha_{j,0}^\star + \alpha_{j,1}^\star \sigma_{j,t^*}^2 + \alpha_{j,2}^\star r_{j,t^*}^2 + \alpha_{j,3}^\star I_{j,t-1}, \end{cases}$$

where $I_{j,t} \in \{0, 1\}$ is a binary indicator of trading activity, equal to one if a non-zero return is observed and zero otherwise. The function $\sigma_{j,t}(\boldsymbol{\theta}_j)$ denotes the conditional volatility, evaluated at the true parameter vector $\boldsymbol{\theta}_j^\star$. The term $\eta_{j,t}$ represents a standardized i.i.d. innovation. The model is specified with respect to a set of true, real-valued parameters $\boldsymbol{\theta}_j^\star := (\alpha_{j,0}^\star, \alpha_{j,1}^\star, \alpha_{j,2}^\star, \alpha_{j,3}^\star) \subseteq (0, +\infty) \times [0, +\infty)^3$. The index t^* denotes the irregular time-step specification, which ensures that the volatility recursion updates only at time points when trading occurs. Although the current specification follows a GARCH(1,1) structure, it can be easily extended to a general GARCH(p, q) setting. The GARCH(1,1) model can thus be viewed as a special case of the more general GARCH(p, q) framework.

While theoretical asymptotic properties such as consistency and asymptotic normality have been established in the literature for univariate zero-inflated time series models, no such results exist for the multivariate setting yet. This thesis fills that gap by proving consistency and asymptotic normality of the quasi-maximum-likelihood estimator of our multivariate model.

The multivariate extension of the zero-inflated GARCH-X model is a key contribution of this thesis, which is implemented in two ways. As mentioned before, binary trading indicators are modeled jointly using a Markov network to capture cross-asset dependence in trading activity. Second, following the copula-GARCH framework, each return series is estimated univariately, and copulas are applied to the standardized residuals to model cross-asset dependence. However, modeling dependence in zero-inflated financial data introduces specific challenges. Zero returns do not generate residuals $\eta_{j,t}$, and are therefore treated as undefined values. In addition, rank-based dependence measures such as Kendall's tau and Spearman's rho become unreliable in an illiquid context due to the high frequency of ties caused by zero returns, making tie-breaking procedures inconsistent and sensitive to the chosen method [50].

To address this issue, recent research has extended both the copula framework and dependence measures to accommodate zero-inflated data. A common copula-based strategy involves a two-part structure: one component models the probability of a non-zero outcome, while the second models the distribution of positive values conditional on occurrence. Copulas are then used to capture both joint occurrence and joint magnitude across variables. In practice, this is often simplified by conditioning on time points where both variables are non-zero. In the context of illiquid assets, Raïssi proposes estimating dependence only on trading days for both assets, as zero returns provide no information on co-movement [56]. This conditional approach better captures the true dependence structure, particularly when zeros occur independently. In terms of dependence measures, Pimentel [52] introduced an adjusted Kendall's tau estimator for zero-inflated continuous variables, which was recently extended by Perrone [50] to the bivariate zero-inflated count setting. However, dependence modeling in the context of illiquid financial assets remains a relatively new area of research, with limited work conducted so far.

To apply copula methods in the presence of zero-inflated data, this thesis also adopts a two-step procedure. First, for each asset pair, we remove all observations where there is an undefined residual value for one of the assets. This corresponds to restricting attention to time points where both assets have non-zero returns, i.e., $I_{j,t} = I_{k,t} = 1$ for some asset pair (j, k), which is related to the approach of [56]. Second, we introduce a probability integral transform approach based on pairwise instead of separate conditioning. In this construction, the univariate marginals are defined conditional on the simultaneous trading activity of each asset, rather than each asset's individual trading activity. In Chapter 7, we prove that this method yields a consistent copula estimator when applied to the subset of observations with non-zero trading activity. Lastly, due to the limited sample size, the analysis is restricted to bivariate cross-dependence – both in trading activity and in asset residuals – rather than higher-dimensional dependence structures.

An outline of the thesis is as follows. Chapter 2 introduces the necessary mathematical background. We begin with classical time series models such as ARCH and GARCH, and then extend to GARCH-X. The chapter also discusses estimation procedures and theoretical properties of the estimators, including consistency and asymptotic normality. To model dependence structures, we introduce copulas and dependence measures, with particular emphasis on Kendall's tau. Additionally, we review fundamental concepts such as the Law of Large Numbers and the Central Limit Theorem, which are important for later theoretical results. The chapter concludes with a discussion of Markov networks and statistical tests for dependence in binary trading indicators. Chapter 3 turns to carbon markets. We explain how carbon markets work, and summarize relevant existing literature

on dependence modeling between carbon credits and other assets. The multivariate zero-inflated GARCH-X model is introduced in Chapter 4. Alongside model formulation, we describe the data and conduct a descriptive analysis to highlight key stylized facts that motivate modeling choices. Chapter 5 focuses on illiquidity, a central theme in this thesis. We examine how illiquidity arises in our dataset, and how cross-dependence among binary trading indicators can be captured in a bivariate framework. The univariate case of the zero-inflated GARCH-X model is the subject of Chapter 6. Here, we prove both consistency and asymptotic normality of the QMLE. In Chapter 7, we return to the multivariate setting and explore cross-dependence in the residuals using copulas. We establish the correct form of the probability integral transform for the zero-inflated multivariate framework and apply the methodology to real data to investigate dependence structures among financial and carbon assets. The conclusions of this thesis are summarized in Chapter 8.

2

Mathematical preliminaries

This chapter aims at providing the necessary fundamentals for the material discussed in subsequent chapters. We start with time series in Section 2.1 and how to deal with their statistical properties. Here, we present widely used models such as Autoregressive Moving Average (ARMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), and an extension of classical GARCH models which are called GARCH-X models. In Section 2.2, we follow up with the standard estimation of GARCH models which is quasi-maximum likelihood estimation. Next, in Section 2.3, we describe the concept of copulas in order to understand the basics of dependence in multivariate random variables. Then, we give some background on dependence measures, where we dive into rank-based correlation. We particularly focus on defining Kendall's tau, including its conditional version. In Section 2.4, we discuss conditional non-parametric estimation, particularly kernel regression. In Section 2.5, we discuss some concepts in asymptotic statistics such as stochastic convergence, consistency, asymptotic normality, the Law of Large Numbers, and the Central Limit Theorem. Then, we present Markov networks, also known as Markov random fields, as undirected graphical models for representing symmetric dependency structures among random variables. Lastly, we outline hypothesis testing, including Fisher's exact test for analyzing independence and dependence in contingency tables.

2.1 Time series

Consider a sequence of real-valued random variables $(X_t)_{t\in\mathbb{Z}}$ with index set $T \subseteq \mathbb{R}$, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Here, Ω represents the sample space, $\mathcal{F}$ is a σ-algebra of events, and $\mathbb{P}$ is a probability measure assigning probabilities to events in $\mathcal{F}$. Such sequence is called a time series. The time series is an example of a discrete-time stochastic process when $T = \mathbb{Z}$ or $T = \mathbb{N}$. On the other hand, it is a continuous time series when $T = \mathbb{R}$. The aim of time series analysis is to construct a model for the underlying stochastic process.

Prior to time series analysis, we define the notions of $\sigma-$algebras and the Borel set which are used in time series analysis.

Definition 2.1 ((Borel-)σ-algebra)**.** Let S be a set. A family $\mathcal{A} \subseteq \mathcal{P}(S)$ is called a ***σ*-algebra** if

(i) $\emptyset, S \in \mathcal{A}$

(ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
(iii) $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$.

Here, $\mathcal{P}(S)$ denotes the power set of S. That is, the set of all subsets of S, including $\emptyset$ and S itself.

Let $\mathcal{B}(S)$ be the σ-algebra generated by the open sets in S, i.e. $\mathcal{B}(S) = \sigma\{\text{open sets in } S\}$. The σ-algebra $\mathcal{B}(S)$ is called the **Borel σ-algebra** of S.

We start with the preliminary concepts of stationary and ergodic sequences, and white noise. First, a time series $(X_t)_{t\in\mathbb{Z}}$ is strictly stationary if its joint distribution remains unchanged under time shifts. There exists also a slightly weaker form which is called second-order or weak stationarity.

Definition 2.2 (Strict and weak stationarity)**.** Consider $(X_t)_{t\in\mathbb{Z}}$ is a time series.

- A time series is called **strictly stationary** if the vector $(X_{t_1}, \ldots, X_{t_k})'$ and the time-shifted vector $(X_{t_1+h}, \ldots, X_{t_k+h})'$ have the same joint distribution, for any $k \in \mathbb{N}$ and $h \in \mathbb{Z}$. Hence, for all $t_1, \ldots, t_k$, it holds that

$$\mathbb{P}(X_{t_1} \leq x_1, \ldots, X_{t_k} \leq x_k) = \mathbb{P}(X_{t_1+h} \leq x_1, \ldots, X_{t_k+h} \leq x_k).$$

- A time series is said to be **weakly stationary** if for all $h \in \mathbb{Z}$, we have that
  (i) The second moments are finite, i.e. $\mathbb{E}[X_t^2] < \infty$.
  (ii) The mean is constant over time, i.e. $\mathbb{E}[X_t] = m$ for some constant $m \in \mathbb{R}$.
  (iii) The covariance between X_t and X_{t+h} depends only on the lag h and not on t.

Note that if a process is strictly stationary, then it is also weakly stationary [24]. In general, both forms of stationarity imply that the time series exhibits similar behavior across any two equally spaced time intervals [13]. Hence, stationarity is a necessary assumption to justify aggregating data across different time periods to estimate return values [27]. Importantly, there may exist a bias when the assumption of stationarity fails, as it would suggest that the model changes over time. It is therefore common to assume that the return series of financial instruments are weakly stationary. A special case of weak stationary series is White Noise (WN), defined as a process with zero mean and uncorrelated observations.

Definition 2.3 (White Noise)**.** A weakly stationary time series (w_t) with mean $\mu_w = 0$ and autocovariance defined as

$$\gamma_W(h) = \begin{cases} \sigma_W^2, & h = 0 \\ 0, & h \neq 0, \end{cases}$$

is called a White Noise series, denoted as $w_t \sim WN(0, \sigma_w^2)$. Here, σ_W^2 represents the constant variance of the white noise process. Remark that the autocovariance is defined as $\gamma_X(h) = \text{cov}(X_{t+h}, X_t)$ which will be further discussed later this section.

Another important concept in time series analysis is ergodicity. Intuitively, ergodicity means that the time average computed along a single realization of a stochastic process converges to the expected value over the distribution of the process. Furthermore, an ergodic process must be stationary, but not all stationary processes are ergodic. Formally, a stationary sequence is called ergodic if it satisfies the Strong Law of Large Numbers [24].

Definition 2.4 (Ergodicity)**.** A real-valued strictly stationary sequence process $(X_t)_{t\in\mathbb{Z}}$ is said to be ergodic if and only if for any Borel set $\mathcal{B}$ and any $k \in \mathbb{Z}$, it holds with probability one that

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{(X_t, X_{t+1}, \ldots, X_{t+k})\in\mathcal{B}} \xrightarrow{a.s.} \mathbb{P}\Big((X_t, X_{t+1}, \ldots, X_{t+k}) \in \mathcal{B}\Big)$$

General transformations of ergodic sequences preserve ergodicity [28]. More precisely, ergodicity is preserved under measurable functions $f : \mathbb{R}^\infty \to \mathbb{R}$, where $\mathbb{R}^\infty = \mathbb{R}^{\mathbb{Z}} = \{(z_t)_{t\in\mathbb{Z}} \mid z_t \in \mathbb{R}\ \forall t \in \mathbb{Z}\}$. A point in $\mathbb{R}^\infty$ is thus an infinite two-sided sequence like $(\ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots)$. A formal statement and proof of this result can be found in for example [8].

The concepts of stationarity, ergodicity and white noise form the basis for many time series models. Now, let us introduce some traditional time series models which assume stationarity and are often constructed using Autoregressive (AR), Moving Average (MA), or a combination of both, the Autoregressive Moving Average (ARMA) model. First, let us define the AR(p) process, where X_t is expressed as a linear function of past values.

Definition 2.5 (Autoregressive Model (AR(p)))**.** An Autoregressive Model of order p is a real-valued process (X_t) of the form

$$X_t = \alpha^\star + \sum_{i=1}^{p} \phi_i^\star X_{t-i} + w_t.$$

Here, $\alpha^\star$ and $\phi_i^\star$ are the (unknown) true real-valued parameters for $i = 1, \ldots, p$, and $w_t \sim WN(0, \sigma_w^2)$.

AR models exhibit autocorrelation and can capture gradual mean-reverting dynamics. Now, let us define a MA(q) model which is a process that expresses X_t as a function of past shocks.

Definition 2.6 (Moving Average Model (MA(q)))**.** A Moving Average model of order q is a real-valued process (X_t) of the form

$$X_t = w_t + \sum_{j=1}^{q} \theta_j^\star w_{t-j}.$$

Here, $\theta_j^\star$ are the (unknown) true real-valued parameters for $j = 1, \ldots, q$, and $w_t \sim WN(0, \sigma_w^2)$.

Contrary to the AR model, the finite MA model is always stationary. Time series with constant volatility can also be described using ARMA(p, q) models, which are a combination of AR and MA models. Both time series imply a linear relationship with past observations of the stationary time series and are characterized by the orders p (number of autoregressive terms), and q (number of lagged forecast errors).

Definition 2.7 (Autoregressive Moving Average (ARMA(p,q)))**.** An autoregressive moving average model of order (p, q) is a real-valued process (X_t) of the form

$$X_t = \alpha^\star + \sum_{i=1}^{p} \phi_i^\star X_{t-i} + w_t + \sum_{j=1}^{q} \theta_j^\star w_{t-j},$$

for some (unknown) true real-valued parameters $\alpha^\star, \phi_1^\star, \cdots, \phi_p^\star, \theta_1^\star, \cdots, \theta_q^\star$, and $w_t \sim WN(0, \sigma_w^2)$.

While the ARMA models are classical for stationary time series, they fall short in describing financial data. Financial time series, such as stock prices, exchange rates, and interest rates, exhibit diverse dynamics and are recorded at varying observation frequencies, ranging from high-frequency (seconds) to low-frequency (days), which complicates statistical modeling. In a financial context, p_t denotes the asset price at time t, and the log return is defined as

$$r_t := \log\left(\frac{p_t}{p_{t-1}}\right) \approx \frac{p_t - p_{t-1}}{p_{t-1}}.$$

The main challenge lies in the statistical regularities (stylized facts) common to many financial series, which are difficult to capture using stochastic models. Below, we discuss some of the most important stylized facts, mentioned in [24]:

1. **Tiny Autocorrelation of Returns**: Stock market returns exhibit very small autocorrelation, meaning past returns provide little predictive power for future returns. This behavior is often compared to white noise. However, at intraday time scales (minutes or seconds), some autocorrelation may appear due to microstructure effects.
2. **Large Autocorrelations of Squared and Absolute Returns**: While raw returns show little autocorrelation, their squared (r_t^2) and absolute ($|r_t|$) values exhibit strong autocorrelation. This suggests that return amplitudes are somewhat predictable over short periods, contradicting the white noise assumption. The slow decay of autocorrelation in squared returns is indicative of long memory in volatility.
3. **Volatility Clustering**: Large price variations tend to be followed by further large variations, and small variations by small ones. This results in clustered periods of high and low volatility, a well-documented feature in financial time series.
4. **Fat-Tailed Distributions**: The empirical distribution of returns deviates from the normal distribution, showing a higher probability of extreme values. This leptokurtic distribution has fatter tails and a sharper peak than a Gaussian distribution. The kurtosis of stock returns is significantly higher than 3 (the value for a normal distribution), indicating a higher occurrence of extreme events.
5. **Gain/Loss Asymmetry**: Stock prices exhibit asymmetry in their up and down movements. Large drawdowns, such as those seen during financial crises, tend to be more abrupt, whereas upward trends are usually more gradual.
6. **Leverage Effect**: There is a negative correlation between volatility and stock returns: during market downturns, volatility increases significantly, whereas during market upturns, volatility tends to decrease more gradually. This effect implies that negative shocks impact volatility more than positive shocks of similar magnitude.
7. **Seasonality and Non-trading Periods**: Stock markets do not trade continuously; they are closed during weekends and holidays. As a result, volatility often spikes after nontrading periods due to the accumulation of information. Furthermore, seasonal effects, such as the day-of-the-week effect, can influence returns.
8. **Spillover Effect**: Volatility and market movements often spill over across stocks, sectors, and international markets. Events in one market can significantly impact others, as seen in global financial crises like the 2008 meltdown.

According to [24], a reliable model for financial time series must capture key stylized facts, especially

volatility clustering and time-varying variance. Traditional models like ARMA assume constant volatility. The property of non-constant conditional variance is called conditional heteroscedasticity, where return variance changes over time. To model the time-varying volatility of financial returns, Engle introduced the Autoregressive Conditional Heteroscedasticity (ARCH) model in 1982. ARCH models allow the conditional variance to depend on past squared errors, formalized as follows:

Definition 2.8 (Autoregressive Conditional Heteroscedasticity Model (ARCH(q)))**.** An ARCH(q) model for a return series (r_t) is defined as

$$\begin{cases} r_t &= \mu^\star + \epsilon_t, \\ \epsilon_t &= \sigma_t \eta_t, \\ \sigma_t^2 &= \alpha_0^\star + \sum_{i=1}^{q} \alpha_i^\star \epsilon_{t-i}^2, \end{cases} \tag{2.1}$$

where r_t represents the asset returns, $\mu^\star$ is the (unknown) true mean return, and ϵ_t is the innovation with volatility σ_t. The term η_t is an i.i.d. sequence with zero mean and unit variance, often assumed to follow a standard normal distribution but this is not necessarily the case. The (unknown) true parameters satisfy $\alpha_0^\star > 0$ and $\alpha_i^\star \geq 0$ for $i = 1, \ldots, q$.

The conditional variance σ_t^2 depends on past squared errors, and captures the empirical observation that large price movements tend to be followed by further large movements. However, one limitation of ARCH models is its reliance on a potentially large number of lagged terms q to adequately capture volatility persistence. To address the limitations of ARCH, Bollerslev introduced the Generalized ARCH (GARCH) model in 1986, which extends ARCH by incorporating lagged conditional variances. This parallels the generalization of an AR process to an ARMA process in time series analysis.

Definition 2.9 (Generalized ARCH Model (GARCH(p,q)))**.** A GARCH(p, q) model for a return series (r_t) is defined as

$$\begin{cases} r_t &= \mu^\star + \epsilon_t, \\ \epsilon_t &= \sigma_t \eta_t, \\ \sigma_t^2 &= \omega^\star + \sum_{i=1}^{q} \alpha_i^\star \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j^\star \sigma_{t-j}^2, \end{cases} \tag{2.2}$$

where r_t represents the asset return, $\mu^\star$ is the (unknown) true mean return, σ_t^2 is the conditional variance at time t, ϵ_t is the innovation, and the (unknown) true parameters are $\omega^\star > 0$, $\alpha_i^\star \geq 0$ for $i = 1, \ldots, q$, and $\beta_j^\star \geq 0$ for $j = 1, \ldots, p$. The term η_t is an i.i.d. sequence with zero mean and unit variance, often assumed to follow a standard normal distribution but this is not necessarily the case.

The constraint, $\sum_{i=1}^{\max(p,q)} (\alpha_i^\star + \beta_i^\star) < 1$, prevents the variance from diverging over time. The inclusion of past conditional variances improves the model's ability to capture long memory in volatility, reducing the need for a high number of ARCH terms [24].

Standard GARCH models describe the conditional variance of returns as a function of past squared returns and past conditional variances. In practice, however, additional information is often available that may improve the modeling of the conditional volatility. This information can take the form of an exogenous covariate such as daily trading volume, high-frequency intraday indicators, or the

returns of other related assets. To incorporate such additional information, GARCH-type models can be extended by adding such explanatory variables directly into the volatility equation. We refer to this different class as GARCH-X models.

Definition 2.10 (GARCH-X Model)**.** A GARCH-X(p, q, r) model for a return series (r_t) is defined as

$$\begin{cases} r_t &= \mu^\star + \epsilon_t, \\ \epsilon_t &= \sigma_t \eta_t, \\ \sigma_t^2 &= \omega^\star + \sum_{i=1}^{q} \alpha_i^\star \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j^\star \sigma_{t-j}^2 + \sum_{k=1}^{r} \gamma_k^\star x_{t-k}, \end{cases} \tag{2.3}$$

where r_t represents the asset return, $\mu^\star$ is the (unknown) true mean return, σ_t^2 is the conditional variance at time t. Here, x_{t-k} is an exogenous covariate observed at time $t-k$, and $\gamma_k^\star$ the corresponding (unknown) true parameter. The (unknown) true parameters satisfy $\omega^\star > 0$, $\alpha_i^\star \geq 0$ for $i = 1, \ldots, q$, $\beta_j^\star \geq 0$ for $j = 1, \ldots, p$, and $\gamma_k^\star \geq 0$ for $k = 1, \ldots, r$. The term η_t is an i.i.d. sequence with zero mean and unit variance, often assumed to follow a standard normal distribution but this is not necessarily the case.

The exogenous covariates are not necessarily dynamically modeled within the GARCH-X framework and are treated as predetermined. This imposes the practical requirement that they must be known or predictable. Note that the term *exogenous* does not refer to the concepts of weak or strong exogeneity introduced in the econometric literature. This term is used because the dynamics of the vector x_t are not specified by the GARCH-X model [24]. To ensure $\sigma_t^2 > 0$ with probability one, it is assumed that the covariates are almost surely positive and that the coefficients satisfy $\alpha \geq 0$, $\beta \geq 0$, $\omega > 0$, and $\gamma = (\gamma_1, \ldots, \gamma_r)^\top \geq 0$ component-wise. In GARCH-type models, the sequence (η_t) is traditionally assumed to be i.i.d. with mean zero and unit variance. However, this assumption is not strictly necessary for establishing the stationarity and ergodicity of the GARCH-X model. Under appropriate conditions such as the stationarity and ergodicity of the joint process (η_t, x_t), the GARCH-X model admits a unique strictly stationary and ergodic solution. A more detailed discussion can be found in Francq and Zakoïan [24].

Note that we explicitly included the (unknown) true parameters, indicated with a star symbol, in Definitions (2.5)–(2.3) to emphasize that these processes represent the true data-generating processes. In the following section, we examine in detail how to estimate these parameters. Furthermore, note that the time series models, as defined in Equations (2.1)-(2.3), generally contain a conditional mean $\mu^\star$. However, we will restrict ourselves to the case of the subclass of GARCH-X models where $\mu^\star = 0$. This assumption is formalized as follows

Assumption 2.1 $(\mu^\star = 0)$**.** *In this thesis, we restrict ourselves always to the subclass of GARCH-X models with zero conditional mean, i.e.* $\mu^\star = 0$*, resulting in the following GARCH-X model:*

$$\begin{cases} r_t &= \sigma_t \eta_t, \\ \sigma_t^2 &= \omega^\star + \sum_{i=1}^{q} \alpha_i^\star r_{t-i}^2 + \sum_{j=1}^{p} \beta_j^\star \sigma_{t-j}^2 + \sum_{k=1}^{r} \gamma_k^\star x_{t-k}. \end{cases} \tag{2.4}$$

Hence, estimation of $\mu^\star$ *later is not needed.*

Next, we would like to focus on concepts closely related to time series. Let us start with analyzing temporal dependence in the levels of a time series. We consider the concept of autocorrelation.

Autocorrelation quantifies the degree of linear dependence between observations separated by a fixed temporal lag. Formally, it is defined as follows [7]:

Definition 2.11 (Autocorrelation function (ACF))**.** For a stationary process (X_t), the theoretical autocorrelation coefficient at lag $h \in \mathbb{N}$, denoted $\rho(k)$, is defined as the Pearson correlation between X_t and X_{t-k}, and is given by

$$\rho(k) = \frac{\mathrm{Cov}(X_t, X_{t-k})}{\sigma_{X_t} \cdot \sigma_{X_{t-k}}}$$

A usual estimator of it is

$$\widehat{\rho}(h) = \frac{\sum_{i=1}^{n-h}(x_{i+h} - \bar{x})(x_i - \bar{x})}{(n-h)s_x^2},$$

where $\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean and $s_x^2 := \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the sample variance.

Positive autocorrelation implies that future values tend to move in the same direction as recent past values. On the other hand, negative serial correlation indicates a tendency for values to reverse direction relative to their recent history. In financial time series, high autocorrelations at low lags in squared or absolute returns are a characteristic of volatility clustering. Moreover, the autocorrelation function helps to determine the appropriate order when specifying $\mathrm{AR}(p)$ and $\mathrm{MA}(q)$ models.

Next to the ACF, the partial autocorrelation function (PACF) is used to separate direct lag dependencies from indirect (spurious) ones. It measures the correlation between X_t and X_{t-k} after removing the effects of intermediate lags $X_{t-1}, \ldots, X_{t-k+1}$. Formally, the PACF at lag k, denoted ϕ_{kk}, is the coefficient of X_{t-k} in the best linear prediction of X_t using $(X_{t-1}, \ldots, X_{t-k})$. While the ACF reflects overall correlation at each lag, the PACF isolates the direct influence of each lag. This makes the PACF especially useful for identifying the order of autoregressive models [24]. To estimate the partial autocorrelation at lag h, one fits an ordinary least-squares regression of X_t on its h previous observations $(X_{t-1}, X_{t-2}, \ldots, X_{t-h})$ and takes the coefficient of X_{t-h} from this regression as the estimate of the partial autocorrelation at lag h.

In time series analysis, understanding the behavior of distribution tails is essential, especially when modeling financial returns or other forms of data exhibiting volatility clustering. Previously, we have seen that financial time series typically deviate from the Gaussian assumption and display heavy tails. This behavior indicates a higher likelihood of extreme values than predicted by the normal distribution. One widely used statistical measure to capture this aspect is the kurtosis coefficient, which quantifies tail heaviness and can provide insight into the presence of volatility dynamics such as those modeled by ARCH or GARCH processes. This coefficient is defined as the ratio of the fourth central moment (assumed to exist) to the square of the second central moment. In the case of the normal distribution, the kurtosis coefficient equals 3, a value often used as a benchmark for comparison with other distributions. Mathematically, this can be expressed as follows:

Definition 2.12 (Kurtosis coefficient)**.** Let X be a random variable. The kurtosis coefficient of X is the fourth standardized moment of X, defined as

$$\kappa = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mathbb{E}\left[(X-\mu)^4\right]}{\left(\mathbb{E}\left[(X-\mu)^2\right]\right)^2} = \frac{\mu_4}{\sigma^4},$$

where $\mu = \mathbb{E}[X]$ is the mean of X, σ is the standard deviation of X, and $\mu_4 = \mathbb{E}[(X-\mu)^4]$ is its fourth central moment.

In the context of GARCH processes, it is particularly insightful to examine the distinction between the tails of the *marginal* and *conditional* distributions. Consider a strictly stationary solution (ϵ_t) of a GARCH(p,q) model. For such a process, the conditional moments of order k are proportional to σ_t^{2k}:

$$\mathbb{E}(\epsilon_t^{2k} \mid \epsilon_u, u < t) = \sigma_t^{2k}\mathbb{E}(\eta_t^{2k})$$

where η_t is, as in Definition 2.9, an i.i.d. sequence with zero mean. Consequently, the kurtosis coefficient of the conditional distribution is constant and equal to that of η_t.

For a general process of the form $\epsilon_t = \sigma_t\eta_t$, where σ_t is a function measurable with respect to the past of ϵ_t, and η_t is independent of this past and i.i.d. with zero mean and finite variance, the kurtosis coefficient of the stationary marginal distribution, provided it exists, is given by [24]:

$$\kappa_\epsilon := \frac{\mathbb{E}(\epsilon_t^4)}{\left[\mathbb{E}(\epsilon_t^2)\right]^2} = \frac{\mathbb{E}\left[\mathbb{E}(\epsilon_t^4 \mid \epsilon_u, u < t)\right]}{\left[\mathbb{E}\left[\mathbb{E}(\epsilon_t^2 \mid \epsilon_u, u < t)\right]\right]^2} = \frac{\mathbb{E}(\sigma_t^4)}{\left[\mathbb{E}(\sigma_t^2)\right]^2}\kappa_\eta,$$

where $\kappa_\eta = \mathbb{E}(\eta_t^4)$ denotes the kurtosis coefficient of the innovations (η_t). This expression reveals that the marginal distribution of (ϵ_t) exhibits heavier tails when the variability of σ_t^2 is large relative to the square of its mean [24]. The lower bound, which corresponds to the absence of ARCH effects (i.e., when σ_t^2 is almost surely constant), is precisely κ_η, implying:

$$\kappa_\epsilon \geq \kappa_\eta,$$

with equality if and only if σ_t^2 is almost surely constant.

2.2 Estimation of GARCH models

This section introduces quasi-maximum likelihood estimation, the usual method for estimating the model parameters of GARCH models and its extensions such as GARCH-X [24]. Essentially, this method allows us to estimate the conditional volatility dynamics consistently even if η_t is not normally distributed [24]. The idea is to construct a likelihood function as if η_t is Gaussian, leading to what is known as the Gaussian quasi-likelihood.

Recall, for a GARCH(p,q) model, we observe a series $r_t = \epsilon_t = \sigma_t\eta_t$, where σ_t^2 is the conditional variance and η_t is an i.i.d sequence with zero mean and unit variance. The orders p and q are assumed to be known. The parameter vector $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_{p+q+1})^\top := (\omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p)^\top$ belongs to a parameter space of the form $\boldsymbol{\Theta} \subseteq (0, +\infty) \times [0, +\infty)^{p+q}$. The true value of the parameter is unknown and is denoted by $\boldsymbol{\theta}^\star := (\theta_1^\star, \ldots, \theta_{p+q+1}^\star)^\top := (\omega^\star, \alpha_1^\star, \ldots, \alpha_q^\star, \beta_1^\star, \ldots, \beta_p^\star)^\top$. The parameter is estimated by maximizing the quasi-log-likelihood function. To write the likelihood of the model, a distribution for the i.i.d. variables η_t needs to be specified. In quasi-maximum likelihood we do not make any assumption on the distribution of these variables, as opposed to maximum likelihood. We work with a function called the (Gaussian) quasi-likelihood that - conditionally on some initial values - coincides with the likelihood when η_t is standard Gaussian. Specifically, given the initial

values $\epsilon_0, \ldots, \epsilon_{1-q}, \sigma_0^2, \ldots, \sigma_{1-p}^2$, the Gaussian quasi-log-likelihood is given by

$$L_n(\boldsymbol{\theta}) = L_n(\boldsymbol{\theta} \mid \epsilon_1, \ldots, \epsilon_n) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2(\boldsymbol{\theta})}} \exp\left(- \frac{\epsilon_t^2}{2\sigma_t^2(\boldsymbol{\theta})} \right), \tag{2.5}$$

where $\sigma_t^2(\boldsymbol{\theta})$ follows the same recursive form as the true volatility σ_t^2, but with the true parameter $\boldsymbol{\theta}^\star$ replaced by $\boldsymbol{\theta}$.

For a given value of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ under the second-order stationarity assumption, the unconditional variance is a reasonable choice for the unknown initial values [24]. These are given by

$$\epsilon_0^2 = \cdots = \epsilon_{1-q}^2 = \sigma_0^2 = \cdots = \sigma_{1-p}^2 = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j} \tag{2.6}$$

Note that such initial values are not suitable when second-order stationarity is not satisfied. In this case, Francq and Zakoïan propose alternative initial values; we refer the reader to their work for further details [24]. Then, a quasi-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_n$ is defined as any measurable solution of

$$\widehat{\boldsymbol{\theta}}_n = \arg\max_{\theta \in \Theta} L_n(\boldsymbol{\theta}),$$

Taking the logarithm, it is seen that maximizing the likelihood is equivalent to minimizing with respect to $\boldsymbol{\theta}$,

$$\widetilde{I}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{\epsilon_t^2}{\sigma_t^2(\boldsymbol{\theta})} + \log \sigma_t^2(\boldsymbol{\theta}) \right).$$

So, a quasi-maximum likelihood estimator is a measurable solution of the equation

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\theta \in \Theta} \widetilde{I}_n(\boldsymbol{\theta})$$

An important remark is the distinction between the true conditional variance process σ_t^2 and the data-dependent recursive function $\sigma_t^2(\boldsymbol{\theta})$, which is used in the quasi-likelihood estimation. The true volatility dynamics are given by

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

where $\boldsymbol{\theta}^\star = (\omega^\star, \alpha_1^\star, \ldots, \alpha_q^\star, \beta_1^\star, \ldots, \beta_p^\star)$ denotes the true (unknown) parameter vector. On the other hand, we define the data-dependent recursive conditional volatility function as

$$\sigma_t^2(\boldsymbol{\theta}) = \sigma_t^2(\omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2(\boldsymbol{\theta}),$$

which is computed recursively for any $\boldsymbol{\theta} \in \Theta$, using the same initialization scheme as in Equation (2.6). In quasi-maximum likelihood estimation, we evaluate the likelihood using this function $\sigma_t^2(\boldsymbol{\theta})$, as it depends on the data and the parameter values. We overload the notation of σ_t^2 by writing

$$\sigma_t^2 = \sigma_t^2(\boldsymbol{\theta}^\star),$$

which emphasizes that the true conditional variance σ_t^2 is equal to the data-dependent recursive function evaluated at the true parameter $\boldsymbol{\theta}^\star$. Therefore, to make the difference explicit, we denote the conditional volatility process at time t as $\sigma_t^2(\boldsymbol{\theta})$, and reserve the notation σ_t^2 for the true conditional volatility process.

Looking at the asymptotic properties of estimators, the choice of initial values is unimportant for the QMLE's limiting behavior [24]. There are two key asymptotic results: consistency and asymptotic normality. We refer to [24] for detailed proofs. Below, we state both results, including the required assumptions.

Theorem 2.1 (Consistency of the quasi-maximum likelihood method for GARCH models)**.** *Let* $\left(\widehat{\theta}_n\right)$ *be a sequence of quasi-maximum likelihood estimators that satisfy the following assumptions:*

(i) $\theta^\star \in \Theta$ *and* Θ *is compact.*

(ii) *Strictly negative Lyapunov exponent:* $\zeta(A^\star) < 0$*, where*

$$\zeta(A^\star) := \inf_{t\in\mathbb{N}\setminus\{0\}} \frac{1}{t}\,\mathbb{E}\left[\log\left\|A_{0t}A_{0t-1}\cdots A_{01}\right\|\right] = \lim_{t\to\infty}\frac{1}{t}\log\left\|A_{0t}A_{0t-1}\cdots A_{01}\right\|,$$

where $\|\cdot\|$ *denotes the matrix norm defined by* $\|A\| := \sum |a_{ij}|$ *for all matrices* $A = (a_{ij})$*, and*

$$A^\star = A_t^\star = \begin{pmatrix} \alpha_1^\star\eta_t^2 & \cdots & \alpha_q^\star\eta_t^2 & \beta_1^\star\eta_t^2 & \cdots & \beta_p^\star\eta_t^2 \\ 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ \alpha_1^\star & \cdots & \alpha_q^\star & \beta_1^\star & \cdots & \beta_p^\star \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix} \tag{2.7}$$

Then, the GARCH model, as defined in Definition (2.9)*, admits a strictly stationary solution.*

(iii) η_t^2 *has a non-degenerate distribution and* $\mathbb{E}[\eta_t^2] = 1$

(iv) *If* $p > 0$*,* $\mathcal{A}_{\theta^\star}(z)$ *and* $\mathcal{B}_{\theta^\star}(z)$ *have no common roots, where*

$$\mathcal{A}_\theta(z) = \sum_{i=1}^{q}\alpha_i z^i \quad \textit{and} \quad \mathcal{B}_\theta(z) = 1 - \sum_{j=1}^{p}\beta_j z^j.$$

Moreover, $\mathcal{A}_{\theta^\star} \neq 0$*, and* $\alpha_q^\star + \beta_p^\star \neq 0$*.*

Under assumptions $(i) - (iv)$*, almost surely*

$$\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}^\star, \quad \textit{as } n \to +\infty$$

It is not assumed that the true value of the parameter belongs to the interior of the parameter space. Thus, Theorem 2.1 allows handling cases where some coefficients are null. Next, we state asymptotic normality for GARCH processes.

Theorem 2.2 (Asymptotic normality of the quasi-maximum likelihood method for GARCH models)**.** *Let* $\left(\widehat{\boldsymbol{\theta}}_n\right)$ *be a sequence of quasi-maximum likelihood estimators satisfying the following assumptions:*

(*a*) *Assumptions* (*i*)–(*iv*) *from Theorem 2.1 hold.*
(*b*) $\boldsymbol{\theta}^\star \in \operatorname{int}(\boldsymbol{\Theta})$, *where* $\operatorname{int}(\boldsymbol{\Theta})$ *denotes the interior of the parameter space* $\boldsymbol{\Theta}$.
(*c*) *The fourth moment of the innovations exists:* $\kappa_\eta := \mathbb{E}[\eta_t^4] < \infty$.

Then, the QMLE is asymptotically normal:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\star) \xrightarrow{d} \mathcal{N}\left(0, (\kappa_\eta - 1)J^{-1}\right),$$

where the matrix J *is defined as*

$$J := \mathbb{E}_{\boldsymbol{\theta}^\star}\left(\frac{\partial^2 l_t(\boldsymbol{\theta}^\star)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) = \mathbb{E}_{\boldsymbol{\theta}^\star}\left(\frac{1}{\sigma_t^4(\boldsymbol{\theta}^\star)}\frac{\partial\sigma_t^2(\boldsymbol{\theta}^\star)}{\partial\boldsymbol{\theta}}\frac{\partial\sigma_t^2(\boldsymbol{\theta}^\star)}{\partial\boldsymbol{\theta}'}\right),$$

and is positive definite.

Instead of in the case of consistency, the asymptotic normality requires that the true value of the parameter belongs to the interior of the parameter space. This is generally not the case when components of $\boldsymbol{\theta}^\star$ are equal to zero [66].

Note that these asymptotic results do not immediately extend to models such as GARCH-X. The asymptotic properties for the GARCH-X model have been established by Kristensen and Han, see also [36]. In Chapter 5, we show that the QMLE of our proposed zero-inflated GARCH-X model is both consistent and asymptotically normal, extending on the results of [24] and [36].

2.3 Copulas

The concept of copulas, first introduced by Sklar in 1959, establishes a link between multivariate distributions and their one-dimensional marginal distributions. Copulas capture the dependence structure among the components of a random vector, and have become a powerful tool for modeling multivariate dependence. They are especially popular since they allow one to easily model and estimate the distribution of random vectors by estimating marginals and copulas separately [38].

In probabilistic terms, a d-dimensional copula is a multivariate distribution function on the hypercube $[0,1]^d$, where the marginals are uniform. In analytic terms, this can be written as follows:

Definition 2.13 (d-dimensional copula)**.** Let $d \geq 2$ be an integer. A copula is a function $C : [0,1]^d \to [0,1]$ with the following properties:

1. For any $j = 1, \ldots, d$ and all $u_j \in [0,1]$, $C(u_1, \ldots, u_{j-1}, 0, u_{j+1}, \ldots, u_d) = 0$

2. For any $j = 1, \ldots, d$ and all $u_j \in [0,1]$, $C(1, \ldots, 1, u_j, 1, \ldots, 1) = u_j$

3. C is d-increasing, i.e. for each hyper-rectangle $A = \Pi_{j=1}^{d}[a_j, b_j] \subseteq [0,1]^d$ the C-volume of A is non-negative:

$$\int_A \mathrm{d}C(\mathbf{u}) \geq 0.$$

Transforming marginal distributions to a uniform scale on $[0,1]$ is essential in copula modeling as it separates the dependence structure from individual marginal characteristics. Without this standardization, dependence measures would be influenced by specific properties of the marginals [16]. This transformation is achieved using the probability integral transform, which states that for a continuous random variable X with cumulative distribution function $F_X(x)$, the random variable $U := F_X(X)$ follows a standard uniform distribution. Once the dependence structure is modeled via a copula, the quantile transformation reverses this process, mapping uniform samples back to the original scale using the inverse CDF: $X = F_X^{-1}(U)$. The following theorem shows how a copula effectively connects the the joint CDF of any multivariate random variable with its one-dimensional marginals.

Theorem 2.3 (Sklar's theorem)**.** *Let $\boldsymbol{X}$ be a d-dimensional continuous random vector taking values in $\mathbb{R}^d$, with joint distribution function F, marginal distribution functions F_j, and marginal density functions f_j for $j = 1, \ldots, d$. Then the joint distribution function can be expressed as*

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

Conversely, given a copula C and marginals F_j, then $(x_1, \ldots, x_d) \mapsto C(F_1(x_1), \ldots, F_d(x_d))$ defines a d-dimensional cumulative distribution function of a random variable on $\mathbb{R}^d$, with marginal distributions F_j. Moreover, if the marginals are continuous, then C is unique.

Clearly, Sklar's theorem states that any multivariate joint distribution function can be decoupled in terms of univariate marginal distribution functions and a copula that describes the dependence structure between the variables [38].
Basic concepts in probability, such as densities, also apply to the concept of copulas. If a copula exists and it has a density, then it can be obtained in the usual manner as follows,

$$c(u_1, \ldots, u_d) = \frac{\partial^p C(u_1, \ldots, u_d)}{\partial u_1, \ldots, \partial u_d},$$

where c is the copula density function. Furthermore, one of the key characteristics of copulas is that they are invariant under monotonic transformations of the marginal distribution [38]. In particular, the following theorem holds.

Theorem 2.4 (Invariance of copulas)**.** *Let $\boldsymbol{X} = (X_1, \ldots, X_d)$ be a random vector with continuous marginal cumulative distribution functions $F_1, \ldots, F_d$ and a continuous copula C. Let $H_i : \mathbb{R} \to \mathbb{R}$, for $i = 1, \ldots, d$, be strictly increasing functions. Then the dependence structure of the random vector*

$$(H_1(X_1), \ldots, H_d(X_d))$$

is also given by the copula C.

For theoretical results on copulas, the following general bounds are useful. Any copula C is pointwise bounded from below by the lower Fréchet-Hoeffding bound W and from above by the upper Fréchet-Hoeffding bound M [16].

Theorem 2.5 (Fréchet-Hoeffding bounds). *For any d-dimensional copula C, it holds that*

$$W(\boldsymbol{u}) \leq C(\boldsymbol{u}) \leq M(\boldsymbol{u}), \quad \boldsymbol{u} \in [0,1]^d$$

where $W(\boldsymbol{u}) = max\left\{ \sum_{j=1}^{d} u_j - d + 1, 0 \right\}$ *and* $M(\boldsymbol{u}) = min_{i \leq j \leq d}\{u_j\}$.

Note that the upper bound M_d is a copula, whereas the lower bound W_d is a copula only for $d \geq 2$. Generally, copulas fall into three families: elliptical copulas, Archimedean copulas, and quadratic copulas [2]. First, elliptical copulas are widely favored in finance literature due to their straightforward implementation [38]. This family includes the normal and Student-t copulas, which are based on elliptical distributions like the multivariate Gaussian or Student-t distribution. The Gaussian copula is symmetric and lacks tail dependence, whereas the Student-t copula can capture extreme dependence between variables. In contrast, Archimedean copulas, such as the Gumbel and Clayton copulas, are not derived from multivariate distribution functions and are useful for capturing asymmetry between lower and upper tail dependencies. The Clayton copula shows greater dependence in the negative tail than in the positive tail, while the Gumbel copula exhibits stronger dependence in the upper tail than in the lower tail. All Archimedean copulas are asymmetric, except for the Frank copula, which can capture the full range of dependence for marginals with weak tail dependence [46]. Similar to the Frank copula, the Plackett copula family is symmetric and can accommodate all possible positive and negative dependence. This is an example of a quadratic copula. However, in this thesis, we will not further focus on Archimedean and quadratic copulas.

Let us now transition to a conditional framework by introducing a d-dimensional conditioning variable $\mathbf{Z}$, formally introduced by [48]. In this context, we can extend the definition of copulas and Sklar's theorem to accommodate the conditional setting as follows using the work of [58].

Definition 2.14. Let $p \geq 2$ and $\mathbf{Z}$ be a conditioning vector taking values in $\mathcal{Z} \subset \mathbb{R}^d$. A conditional copula is a measurable function $C : [0,1]^p \times \mathcal{Z} \to [0,1]$ such that for $\mathbb{P}_{\mathbf{Z}}$-almost every $\mathbf{z} \in \mathcal{Z}$ the following properties are satisfied:

1. For any $j = 1, \ldots, p$ and all $u_j \in [0,1]$,

$$C(u_1, \ldots, u_{j-1}, 0, u_{j+1}, \ldots, u_p \mid \mathbf{Z} = \mathbf{z}) = 0.$$

2. For any $j = 1, \ldots, p$ and all $u_j \in [0,1]$,

$$C(1, \ldots, 1, u_j, 1, \ldots, 1 \mid \mathbf{Z} = \mathbf{z}) = u_j.$$

3. For each hyperrectangle $A = \prod_{j=1}^{p} [a_j, b_j] \subseteq [0,1]^p$, the C-volume of A is non-negative:

$$\int_A dC(\mathbf{u} \mid \mathbf{Z} = \mathbf{z}) \geq 0.$$

Clearly, conditional copulas are defined as the conditional joint CDF of a multivariate random vector on the unit cube with uniform marginals (conditionally to $\mathbf{Z}$). Next, let us simply state the conditional version of Sklar's theorem. For a proof we refer to [48].

Theorem 2.6 (Sklar's Theorem for Conditional Copulas). *Let* $\mathbf{X}$ *and* $\mathbf{Z}$ *be random vectors taking values in respectively* $\mathbb{R}^p$ *and* $\mathcal{Z} \subset \mathbb{R}^d$. *Let the conditional joint CDF of* $\mathbf{X}$ *given* $\mathbf{Z} = \mathbf{z}$, *denoted by* $F_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}$, *have*

conditional marginals $F_{1|\mathbf{Z}=\mathbf{z}}, \ldots, F_{p|\mathbf{Z}=\mathbf{z}}$. *Then, there exists a conditional copula, denoted by* $C_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}$, *such that for all* $\mathbf{x} \in \mathbb{R}^p$ *and all* $\mathbf{z} \in \mathcal{Z}$,

$$C_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}(u_1, \ldots, u_p) = F_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}\left(F^{-1}_{1|\mathbf{Z}=\mathbf{z}}(u_1), \ldots, F^{-1}_{p|\mathbf{Z}=\mathbf{z}}(u_p)\right), \tag{2.9}$$

where $F^{-1}_{j|\mathbf{Z}=\mathbf{z}}$ *denotes the generalized inverse of* $F_{j|\mathbf{Z}=\mathbf{z}}$ *for* $j = 1, \ldots, p$. *Therefore, if* $\mathbf{X}|\mathbf{Z} = \mathbf{z}$ *is continuous, then* $C_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}$ *is unique.*

Similarly to the unconditional case, we can combine any set of continuous conditional marginal CDFs with any conditional copula to form a well-defined conditional CDF. Lastly, it is important to highlight that the dependence of the conditional copula $C_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}$ on the conditioning variable $\mathbf{z}$ introduces challenges in both model specification and inference. To address this, the so-called simplifying assumption is often imposed. This assumption tells that $C_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}$ is independent of $\mathbf{z}$. Note that this does not imply that the conditional copula coincides with the unconditional copula; rather, it means that the conditional dependence structure does not vary with the conditioning point. For a comprehensive discussion of the simplifying assumption, and methods for testing it, we refer to [17].

2.4 Dependence measures

With applications in mind, it is often desirable to summarize the dependence between components of a random vector by a real number. Such numerical summaries of dependence are known as dependence measures and are mostly studied in bivariate cases [38].

Pearson's rho is a widely used measure of linear dependence. Essentially, it is a normalized version of the covariance with values within $[-1, 1]$. Its mathematical definition and sample version are given in the following definition.

Definition 2.15 (Pearson's rho)**.** Let X_1 and X_2 be real-valued random variables. The population Pearson correlation coefficient of X_1 and X_2 is defined by

$$\rho_{X_1,X_2} = \frac{\operatorname{cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}}$$

where σ denotes the standard deviation. Further, let $\{(X_{1,1}, X_{2,1}), \ldots, (X_{1,n}, X_{2,n})\}$ be paired observations, then a classical sample Pearson correlation coefficient is defined by

$$\widehat{\rho}_{X_1,X_2} = \frac{\sum_{i=1}^n (X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2)}{\sqrt{\sum_{i=1}^n (X_{1,i} - \bar{X}_1)^2 \sum_{i=1}^n (X_{2,i} - \bar{X}_2)^2}}.$$

Despite its popularity, it has several key limitations. First, it is only defined for random vectors with finite second moments, meaning it does not exist for all distributions. Additionally, even for continuous distributions, it depends on the marginal distribution of the random vector and cannot be expressed solely in terms of the underlying copula. This makes it less suitable for studying dependence structures in a copula-based framework. Furthermore, Pearson's correlation is invariant only under strictly increasing linear transformations but not under strictly increasing transformations in general. These limitations highlight its restricted applicability, particularly in cases where nonlinear dependencies or varying marginal distributions play a significant role.

Contrary to usual linear correlations, rank correlation measures have the advantage of being defined without any condition on moments and being invariant to changes in the underlying marginal distributions [18]. Rank-based correlation measures provide a broader characterization of dependence by focusing on the monotonic association between variables. More precisely, measures of rank correlation indicate the similarity of the orderings of the data when ranked by each of the quantiles [58]. Two widely used measures of rank correlation are Spearman's rho and Kendall's tau. Additional rank-based dependence measures can be found in [46].

Let us start with Kendall's tau which is based on the concepts of concordance and discordance between data points. Informally, a pair of random variables is concordant if large values of one variable tend to be associated with large values of the other and, conversely, small values of one variable are associated with small values of the other [46]. Formally, the definitions of concordance and discordance are as follows:

Definition 2.16 (Concordance and discordance)**.** Let $(X_{1,1}, X_{2,1})$ and $(X_{1,2}, X_{2,2})$ be two independent copies of a random vector $\mathbf{X} \in \mathbb{R}^2$. A pair is:

- **Concordant** if $(X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) > 0$,
- **Discordant** if $(X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) < 0$.

A pair of bivariate observations is concordant when both elements of one observation are either greater than or less than the corresponding elements of another observation. A pair is discordant when only one of the elements in one observation is greater than the corresponding element in the other [46]. Using these concepts, Kendall's tau is defined as the difference between the probability of concordance and the probability of discordance between two independent versions of (X_1, X_2). Formally, this can be expressed as follows:

Definition 2.17 (Kendall's tau)**.** Let X_1 and X_2 be real-valued random variables. The population Kendall's tau is given by:

$$\tau_{X_1,X_2} = \mathbb{P}((X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) > 0) - \mathbb{P}((X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) < 0),$$

where $(X_{1,i}, X_{2,i})_{i=1,2}$ are two independent copies of (X_1, X_2). A classical sample version can be expressed as,

$$\widehat{\tau}_{X_1,X_2} = \frac{2}{n(n-1)} \sum_{i_1<i_2} \text{sign}((X_{1,i_1} - X_{1,i_2})(X_{2,i_1} - X_{2,i_2})).$$

Hence, Kendall's tau always falls within the interval $[-1, 1]$. A value of -1 indicates perfect negative correlation, while a value of 1 signifies perfect positive correlation. If the value of Kendall's tau is equal to zero, the variables are independent. However, the converse does not necessarily hold, i.e. having a Kendall's tau of zero does not imply that the variables are completely independent [7]. Furthermore, Kendall's tau is often linked to copulas. Denoting by $C_{1,2}$ the unique underlying copula

of (X_1, X_2), which is assumed to be continuous, Kendall's tau has several equivalent representations:

$$\begin{aligned}\tau_{X_1,X_2} &= 2\mathbb{P}((X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) > 0) - 1 \\ &= 4\mathbb{P}(X_{1,1} < X_{2,1}, X_{1,2} < X_{2,2}) - 1 \\ &= 1 - 4\mathbb{P}(X_{1,1} < X_{2,1}, X_{1,2} > X_{2,2}) \\ &= 4\int_{[0,1]^2} C(u_1, u_2)dC(u_1, u_2) - 1.\end{aligned}$$

Kendall's tau provides significant insights into the underlying dependence structure. For convenience, the notation $\tau_{1,2}$ will be used instead of τ_{X_1,X_2} whenever the variables are clear from the context.

Now, let us discuss the other rank-based correlation coefficient, Spearman's rho. Consider random variables X_1, X_2 with marginals F_1, F_2. Essentially, Spearman's rho is the linear correlation coefficient of the random vector $(F(X_1), F(X_2))$ obtained from (X_1, X_2) by marginally applying the probability integral transform [38]. In comparison with the linear correlation coefficient, Spearman's rho thus always exists. Since the distribution of $(F(X_1), F(X_2))$ is in fact the copula C, it is clear Spearman's rho depends on the underlying copula, and not on the marginals F_1 and F_2. Formally, Spearman's rho is defined as follows:

Definition 2.18 (Spearman's rho)**.** Let X_1 and X_2 be real-valued continuous random variables. The population Spearman's rho of X_1 and X_2 is defined as:

$$\rho_{X_1,X_2} := \rho_{U_1,U_2} = \frac{\mathrm{Cov}(U_1, U_2)}{\sigma_{U_1}\sigma_{U_2}} = \frac{E[U_1U_2 - \frac{1}{4}]}{\frac{1}{12}}$$

which simplifies to:

$$\rho_{X_1,X_2} = 12\int_{[0,1]^2} u_1u_2dC(u_1, u_2) - 3,$$

where $\sigma_{U_1}, \sigma_{U_2}$ denote the standard deviations of U_1, U_2 and C is the copula corresponding to (X_1, X_2). Further, given paired observations $\{(X_{1,1}, X_{2,1}), \ldots, (X_{1,n}, X_{2,n})\}$ with distinct integer ranks, the sample Spearman's rho is defined as:

$$\widehat{\rho}_{X_1,X_2} = 1 - \frac{6\sum_{i=1}^{n}(R(X_{1,i}) - R(X_{2,i}))^2}{n(n^2 - 1)},$$

where $R(X_{1,i})$ and $R(X_{2,i})$ denote the ranks of observations $X_{1,i}$ and $X_{2,i}$ for $i = 1, \ldots, n$.

The corresponding sample statistic is robust to outliers since it depends only on ranked data. The same robustness property also applies to other rank correlation coefficients, such as Kendall's rank correlation coefficient [58].

Lastly, we will discuss the conditional setup for Kendall's tau, as we will only use this conditional dependence measure throughout this thesis. Generally, conditional dependence measures are of interest because they allow us to summarize the evolution of the dependence between X_1 and X_2, when the covariate is changing [18]. Indeed, the goal is now to measure the dependence between the two components X_1 and X_2, given the vector of covariates $\mathbf{Z}$. Starting with the conditional Kendall's tau, this is a conditional dependence measure used to predict whether a pair of random variables

is concordant or discordant conditionally on $\mathbf{Z}$. We define the conditional Kendall's tau in a similar manner as its unconditional counterpart.

Definition 2.19 (Conditional Kendall's tau)**.** Let X_1 and X_2 be real-valued random variables and $\mathbf{Z}$ be a random vector taking values in $\mathbf{Z} \subset \mathbb{R}^d$. For any point $\mathbf{z} \in \mathbf{Z}$, we define Kendall's tau of X_1 and X_2 conditional on $\mathbf{Z} = \mathbf{z}$ by

$$\begin{aligned}\tau_{1,2|\mathbf{Z}=\mathbf{z}} = \mathbb{P}\left((X_{1,1} - X_{1,2})(X_{2,1} - X_{2,2}) > 0 \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\right) \\ - \mathbb{P}\left((X_{1,1} - X_{1,2})(X_{2,1} - X_{2,2}) < 0 \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\right),\end{aligned}$$

where $(X_{1,i}, X_{2,i}, Z_i)_{i=1,2}$ are two independent versions of $(X_1, X_2, \mathbf{Z})$.

For every point $\mathbf{z} \in \mathbf{Z}$, the conditional Kendall's tau takes values in $[-1, 1]$, while the underlying conditional copula is a bivariate function for each $\mathbf{z} \in \mathbf{Z}$. Similarly as in the unconditional setup, when the conditional marginal distributions of X_1 and X_2 given $\mathbf{Z} = \mathbf{z}$ are continuous, Definition 2.19 is equivalent to any of the following expressions:

$$\begin{aligned}\tau_{1,2|\mathbf{Z}=\mathbf{z}} &= 4\mathbb{P}(X_{1,1} < X_{1,2}, X_{2,1} < X_{2,2} \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}) - 1, \\ &= 1 - 4\mathbb{P}(X_{1,1} < X_{1,2}, X_{2,1} > X_{2,2} \mid \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}), \\ &= 4\int_{[0,1]^2} C_{1,2|\mathbf{Z}=\mathbf{z}}(u_1, u_2)\, dC_{1,2|\mathbf{Z}=\mathbf{z}}(u_1, u_2) - 1.\end{aligned}$$

Note that, as conditional copulas themselves, conditional Kendall's tau are invariant with respect to monotonic transformations [18]. Of course, if $\mathbf{Z}$ is independent of (X_1, X_2) then, for every $\mathbf{z} \in \mathbb{R}^d$ the conditional Kendall's tau is equal to the unconditional Kendall's tau.

2.5 Conditional dependence estimation

In this section, we introduce the notion of kernel regression for construction of non-parametric conditional estimators. Let us be interested in the estimation of the expectation of a random variable Y conditional on some covariate $\mathbf{Z} \in \mathbb{R}^d$, taking values in a measurable state space $\mathcal{Z} \subseteq \mathbb{R}^d$. Then, in the case of a continuously distributed covariate we will almost surely never observe that $\mathbf{Z} = \mathbf{z}$, for any $\mathbf{z} \in \mathcal{Z}$. Therefore a method is needed with which it is still possible to compute reasonable estimates. Most natural is the concept of kernel regression, or also known as kernel smoothing. That is, to consider adjacent observations in which the variate $\mathbf{Z}$ is close to the points $\mathbf{z}$ at which we want the estimate. Recall the definition of the conditional expectation:

$$m(\mathbf{z}) = \mathbb{E}[Y \mid \mathbf{Z} = \mathbf{z}] = \int y f_{Y|\mathbf{Z}=\mathbf{z}}(y)\mathrm{d}y = \frac{\int f_{\mathbf{Z},Y(\mathbf{z},y)}(y)\mathrm{d}y}{f_{\mathbf{Z}}(\mathbf{z})} \tag{2.10}$$

In kernel regression, the estimates of $f_{\mathbf{Z},Y}(\mathbf{z}, y)$ and $f_{\mathbf{Z}}(\mathbf{z})$ are computed by kernel density estimation. That is, the density functions are approximated by adjacent observations of Z in the following way

$$\begin{aligned}\hat{f}_{\mathbf{Z},Y}(\mathbf{z}, y; h) &:= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d} K_1\left(\frac{\mathbf{z} - \mathbf{Z}_i}{h}\right) \frac{1}{h} K_2\left(\frac{y - Y_i}{h}\right), \\ \hat{f}_{Z}(\mathbf{z}; h) &:= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d} K_1\left(\frac{\mathbf{z} - \mathbf{Z}_i}{h}\right),\end{aligned} \tag{2.11}$$

where $K_1 : \mathbb{R}^d \to \mathbb{R}$ and $K_2 : \mathbb{R} \to \mathbb{R}$ are kernel functions, assumed to be symmetric, unimodal at zero, and satisfying $\int K = 1$. The parameter h is the bandwidth and n is the sample size. The bandwidth controls the sensitivity of the density estimates towards observations further away from $\mathbf{z}$, whereas the kernel function defines the form of the dependency. Commonly used kernels are the Gaussian and Epanechnikov kernels. Then, the smoothed estimate of $m(\mathbf{z})$ is obtained by replacing $f_{\mathbf{Z},Y}(\mathbf{z}, y)$ and $f_{\mathbf{Z}}(\mathbf{z})$ in (2.10) with their respective kernel density estimates. By using the respective kernel density estimates from (2.11), we obtain

$$\begin{aligned}\widehat{m}(\mathbf{z}; h) &:= \frac{\int y \widehat{f}_{\mathbf{Z},Y}(\mathbf{z}, y; h)\, dy}{\widehat{f}_{\mathbf{Z}}(\mathbf{z}; h)} \\ &= \frac{\int y \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d} K_1\left(\frac{\mathbf{z}-\mathbf{Z}_i}{h}\right) \frac{1}{h} K_2\left(\frac{y-Y_i}{h}\right) dy}{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d} K_1\left(\frac{\mathbf{z}-\mathbf{Z}_i}{h}\right)} \\ &= \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d} K_1\left(\frac{\mathbf{z}-\mathbf{Z}_i}{h}\right) \int y \frac{1}{h} K_2\left(\frac{y-Y_i}{h}\right) dy}{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d} K_1\left(\frac{\mathbf{z}-\mathbf{Z}_i}{h}\right)}.\end{aligned} \tag{2.12}$$

We can work out (2.12) by a change of variable $u_i = \frac{y-Y_i}{h}$. This derivation is originated from [32] and also mentioned in [58]. We find that

$$\begin{aligned}\int y \frac{1}{h} K_2\left(\frac{y - Y_i}{h}\right) dy &= \int (hu + Y_i) K_2(u) du \\ &= h \int u K_2(u) du + Y_i \int K_2(u) du \\ &= Y_i,\end{aligned} \tag{2.13}$$

where we used that K_2 is symmetric and that $\int K_2 = 1$. By combining Equations (2.13) and (2.12), we obtain

$$\begin{aligned}\widehat{m}(z; h) &= \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d} K_1\left(\frac{z-Z_i}{h}\right) Y_i}{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{h^d} K_1\left(\frac{z-Z_i}{h}\right)} \\ &= \sum_{i=1}^{n} w_{i,n}(z) Y_i,\end{aligned} \tag{2.14}$$

where

$$w_{i,n}(z) := \frac{K_h(Z_i - z)}{\sum_{k=1}^{n} K_h(Z_k - z)}, \tag{2.15}$$

with $K_h := h^{-d} K(\cdot/h)$. The resulting estimator $\widehat{m}(\mathbf{z}; h)$ can be seen as a weighted average of $Y_1, \cdots, Y_n$ by means of the so-called Nadaraya-Watson weights $w_{i,n}(\mathbf{z})$. Furthermore, note again that the bandwidth h controls the estimator's sensitivity towards observations $\mathbf{Z}_i$ that are further away from point z. It will have a strong influence on the characteristics of the estimator and is closely related to the sample size. Reducing the bandwidth will decrease the estimator's bias and increase its variance, which is known as the bias-variance tradeoff [7] [19]. Hereby, larger sample sizes will allow for a smaller choice of the bandwidth. Further note that the volume of the space $\mathcal{Z}$ grows exponentially

fast when increasing the dimensionality of $\mathcal{Z}$. As such, the density of observations within that space decreases at the same rate, calling for an exponentially increasing bandwidth. In practice, this so-called curse of dimensionality means that we can only consider covariates of a few dimensions at most.

In this thesis, kernel estimation is carried out on the basis of the Epanechnikov kernel and simulations are performed with the help of the `R` package `CondCopulas` [20]. Specifically, the function `CKT.kernel` implements a kernel-based nonparametric estimator for the conditional Kendall's tau $\tau_{1,2|\mathbf{z}}$ as proposed by Derumigny and Fermanian [18]. This estimator is defined by

$$\widehat{\tau}_{1,2|z} = \frac{\widehat{\tau}^*_{1,2|z}}{1 - s_n},$$

where,

$$\widehat{\tau}^*_{1,2|\mathbf{z}} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}(\mathbf{z}) w_{j,n}(\mathbf{z}) \Big[\mathbf{1}\Big\{(X_{i,1} - X_{j,1})(X_{i,2} - X_{j,2}) > 0\Big\} - \mathbf{1}\{(X_{i,1} - X_{j,1})(X_{i,2} - X_{j,2}) < 0\Big\} \Big]$$

$$s_n = \sum_{i=1}^{n} w^2_{i,n}(\mathbf{z}).$$

Here, the weights $w_{i,n}(z)$ are defined as Nadaraya–Watson kernel weights. The estimator $\widehat{\tau}_{1,2|z}$ includes a correction factor $1 - s_n$ such that the estimate only takes values in $[-1, 1]$. In fact, $\widehat{\tau}^*_{1,2|\mathbf{z}}$ takes values in $[-1 + s_n, 1 + s_n]$. The implementation supports both univariate and multivariate conditioning variables $\mathbf{Z}$, and allows for the choice of kernel functions (e.g. Epanechnikov or Gaussian) and bandwidth h.

2.6 Asymptotic statistics

Asymptotic statistics studies the limiting behavior of statistical procedures as the sample size tends to infinity. Central to this theory is stochastic convergence, which describes how sequences of random variables behave in the limit. While the classical limit $f_n(x) \to f(x)$ as $n \to \infty$ refers to pointwise or uniform convergence of deterministic functions on a fixed domain, stochastic convergence formalizes convergence in a probabilistic sense. There are several distinct modes of stochastic convergence. These include convergence in distribution, in probability, almost surely, and in L^p. Each captures a different level of probabilistic control over the sequence and forms part of a natural hierarchy.

This section begins with the formal definitions of these convergence concepts. We then develop the notions of consistency and asymptotic normality. These properties are important for the theoretical analysis of estimators and play a role in large-sample inference. All definitions and results in this section are based on the work of [7] and [66].

A random vector in $\mathbb{R}^k$ is a vector $\mathbf{X} = (X_1, \ldots, X_k)$ of real random variables. Recall that the distribution function of $\mathbf{X}$ is the map $\mathbf{x} \mapsto \mathbb{P}(\mathbf{X} \leq \mathbf{x})$. Let $(\mathbf{X}_n)_{n\in\mathbb{N}}$ be a sequence of random vectors defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $d(x, y)$ be a distance function on $\mathbb{R}^k$, for example the

Euclidean distance:

$$d(x, y) = \| x - y \| = \left(\sum_{i=1}^{k} (x_i - y_i)^2 \right)^{\frac{1}{2}}.$$

Then, we define the following modes of stochastic convergence:

Definition 2.20 (Modes of stochastic convergence)**.**

(i) **Convergence in distribution:** A sequence of random vectors $(\mathbf{X}_n)_{n\geq 1}$ is said to converge in distribution to a random vector $\mathbf{X}$, denoted by $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$, if

$$\mathbb{P}(\mathbf{X}_n \leq \mathbf{x}) \to \mathbb{P}(\mathbf{X} \leq \mathbf{x}), \tag{2.16}$$

for every $\mathbf{x}$ at which the limit distribution function $\mathbf{x} \mapsto \mathbb{P}(\mathbf{X} \leq \mathbf{x})$ is continuous.

(ii) **Convergence in probability:** A sequence of random variables $(\mathbf{X}_n)_{n\geq 1}$ is said to converge in probability, denoted by $X_n \xrightarrow{p} X$, if

$$\forall \epsilon > 0, \quad \mathbb{P}(d(\mathbf{X}_n, \mathbf{X}) > \epsilon) \underset{n\to\infty}{\longrightarrow} 0. \tag{2.17}$$

(iii) **Almost sure convergence:** A sequence of random variables $(\mathbf{X}_n)_{n\in\mathbb{N}}$ is said to converge almost surely to $\mathbf{X}$, denoted by $\mathbf{X}_n \xrightarrow{a.s.} \mathbf{X}$, if

$$\mathbb{P}(\lim_{n\to\infty} d(\mathbf{X}_n, \mathbf{X}) = 0) = 1. \tag{2.18}$$

(iv) **L^p-convergence:** A sequence of random variables $(\mathbf{X}_n)_{n\geq 1}$ is said to converge in L^p to $\mathbf{X}$ for $p \geq 1$, denoted by $\mathbf{X}_n \xrightarrow{L^p} \mathbf{X}$, if

$$\lim_{n\to+\infty} \mathbb{E}\left[|\mathbf{X}_n - \mathbf{X}|^p\right] = 0, \tag{2.19}$$

given that the p-th absolute moments $\mathbb{E}[|X_n|^p]$ and $\mathbb{E}[|X|^p]$ of X_n and X exist for all n.

There exist relationships between the different modes of stochastic convergence. Specifically, almost sure convergence implies convergence in probability, which in turn implies convergence in distribution. Similarly, convergence in L^p for $p \geq 1$ implies convergence in probability. However, these implications are not generally reversible. For example, convergence in distribution does not imply convergence in probability, unless there are additional assumptions. For proofs of the mentioned implications, we refer to [66].

A useful result in stochastic convergence is the continuous mapping theorem. Essentially, it states that if a sequence of random vectors $(\mathbf{X}_n)_{n\geq 1}$ converges to a random vector $\mathbf{X}$, and g is a continuous function, then the sequence of transformed random vectors $(g(\mathbf{X}_n))_{n\geq 1}$ converges to $g(\mathbf{X})$. Importantly, this result holds for each of the three principal modes of stochastic convergence.

Theorem 2.7 (Continuous Mapping Theorem)**.** *Let $g : \mathbb{R}^k \to \mathbb{R}^m$ be a function that is continuous at every point of a set $C \subset \mathbb{R}^k$, with $\mathbb{P}(\boldsymbol{X} \in C) = 1$. Then:*

(i) If $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$, then $g(\boldsymbol{X}_n) \xrightarrow{d} g(\boldsymbol{X})$.

(*ii*) *If* $\boldsymbol{X}_n \xrightarrow{p} \boldsymbol{X}$, *then* $g(\boldsymbol{X}_n) \xrightarrow{p} g(\boldsymbol{X})$.

(*iii*) *If* $\boldsymbol{X}_n \xrightarrow{a.s.} \boldsymbol{X}$, *then* $g(\boldsymbol{X}_n) \xrightarrow{a.s.} g(\boldsymbol{X})$.

For a proof of Theorem 2.7, we refer to [66]. From these definitions of stochastic convergences, one can define corresponding properties about estimators themselves: consistency and asymptotic normality. Both of these properties will be proved for the estimators that we will propose in this thesis, specifically in Chapter 5. In this setting, let Θ denote the parameter space, and $\theta^\star$ denote the true parameter.

Definition 2.21 (Consistency)**.** An estimator $\widehat{\boldsymbol{\theta}}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$ of the true parameter $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta}$ is said to be

(i) **weakly consistent**, if

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}^\star, \quad \text{as } n \to +\infty.$$

(ii) **strongly consistent**, if

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}^\star, \quad \text{as } n \to +\infty.$$

Consistency guarantees that for large n, the estimator will be arbitrarily close to the parameter value. Related to the definition of consistency is the following proposition of Kristensen and Rahbek, adapted from [42, p. 961] to the notation used in this thesis. It shows that, under suitable conditions such as the existence of a limiting criterion function defined via the stationary distribution of the process, strong consistency of the estimator holds. This proposition forms the basis of the consistency proof for the quasi-maximum likelihood estimator of zero-inflated GARCH-X models proposed in Chapter 6. Formally, let $(r_t)_{t\geq 1}$ denote a geometrically ergodic Markov sequence such that a stationary version $(r_t^*)_{t\geq 1}$ exists, representing for example financial returns of some asset j. Recall that $L_n(\boldsymbol{\theta})$ is the average Gaussian quasi-log-likelihood function given by

$$L_{j,n}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} l(r_t \mid r_{t-1}; \boldsymbol{\theta}), \quad l(r_{j,t} \mid r_{j,t-1}; \boldsymbol{\theta}_j) = \log(\sigma_t) + \frac{r_{j,t}^2}{\sigma_t}.$$

We now state Proposition 2 from Kristensen and Rahbek [42].

Proposition 2.1 (Kristensen & Rahbek)**.** *Make the following assumptions for some asset* $j \in \{1, \ldots, N_{assets}\}$*:*

(*i*) *The parameter space* Θ_j *is a compact Euclidean space, and the true parameter* $\theta^\star$ *belongs to* Θ.
(*ii*) *The mapping* $\boldsymbol{\theta} \mapsto l(r_{j,t} \mid r_{j,t-1}; \boldsymbol{\theta})$ *is continuous for all* $x, y \in \mathbb{R}^d$.
(*iii*) *For all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, *the following exists*

$$L(\boldsymbol{\theta}) := \mathbb{E}\big[l(r_t^* \mid r_{t-1}^*; \boldsymbol{\theta})\big].$$

(*iv*) *The true parameter minimizes uniquely, that is,* $L(\boldsymbol{\theta}^\star) < L(\boldsymbol{\theta})$ *for all* $\boldsymbol{\theta} \neq \boldsymbol{\theta}^\star$.
(*v*) *For any compact set* $\mathcal{D} \subset \boldsymbol{\Theta}$ *with* $\boldsymbol{\theta}^\star \notin \mathcal{D}$, *we have*

$$\mathbb{E}\big[\inf_{\boldsymbol{\theta}\in D} l(r_t \mid r_{t-1}; \boldsymbol{\theta})\big] > -\infty.$$

Then, the estimator $\widehat{\boldsymbol{\theta}}_n$ *defined by* $\widehat{\boldsymbol{\theta}}_n := \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{n}\sum_{i=1}^{n} l(r_i \mid r_{i-i}; \boldsymbol{\theta})$ *is strongly consistent, that is,*

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}^\star \quad as\ n \to \infty. \tag{2.20}$$

In asymptotic theory, it is convenient to have short expressions for terms that converge in probability to zero. The notations $o_p(\cdot)$ and $O_p(\cdot)$ are used to describe the limiting behavior of sequences of random variables. The notation $o_p(1)$ is short for a sequence of random vectors that converge to zero in probability. On the other hand, the expression $O_p(1)$ denotes a sequence that is bounded in probability [66]. More generally, we define this concept as follows.

Definition 2.22 (Big-O_p and Small-o_p)**.** Let (X_n) be a sequence of random variables and (a_n) a sequence of positive constants. Then we have:

(i) **Small-o_p.**

$$X_n = o_p(a_n),\ \text{if}\ \frac{X_n}{a_n} \xrightarrow{p} 0 \quad \text{as}\ n \to +\infty.$$

Equivalently, this may be written as $X_n/a_n = o_p(1)$.

(ii) **Big-O_p.**

$$X_n = O_p(a_n),\ \text{if}\ \forall\varepsilon > 0,\ \exists M > 0,\ \exists N < +\infty,\ \forall n > N,\ \mathbb{P}\left(\left|\frac{X_n}{a_n}\right| > M\right) < \varepsilon.$$

That is, the sequence X_n/a_n is stochastically bounded.

The difference between $o_p(a_n)$ and $O_p(a_n)$ lies in the strength of convergence. In addition to the formal Definition 2.22, the notation $X_n = O_p(a_n)$ intuitively means that, with high probability, X_n does not grow faster than some constant multiple of a_n as $n \to +\infty$. In contrast, the notation $X_n = o_p(a_n)$ expresses that X_n/a_n converges to zero in probability. So while $O_p(a_n)$ only requires boundedness in probability, $o_p(a_n)$ requires the stronger condition of convergence in probability toward zero. As a result, $o_p(a_n)$ always implies $O_p(a_n)$, but the reverse implication does not hold [66].

We use the O_p and o_p notations in the following result by Kristensen and Shin. The theorem of Kristensen and Shin is the second theorem used in the consistency proof of the zero-inflated GARCH-X models proposed in Chapter 6, next to Proposition 2.1. Before stating the result, we introduce additional notation: the score S_T and the hessian H_T, which are defined as the first and second derivatives of the quasi-log-likelihood function with respect to the parameter vector $\boldsymbol{\theta}$ respectively, and are given by:

$$S_T(\boldsymbol{\theta}) = \frac{\partial L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\nu_T}\sum_{t=1}^{T} \frac{\partial l(r_t \mid r_{t-1}; \theta)}{\partial \boldsymbol{\theta}} \in \mathbb{R}^d,$$

$$H_T(\boldsymbol{\theta}) = \frac{\partial^2 L_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = \frac{1}{\nu_T}\sum_{t=1}^{T} \frac{\partial^2 l(r_t \mid r_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} \in \mathbb{R}^{d\times d},$$

$$G_{T,i}(\boldsymbol{\theta}) = \frac{\partial^3 L_T(\boldsymbol{\theta}}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'\partial \boldsymbol{\theta}_i} = \frac{1}{\nu_T}\sum_{t=1}^{T} \frac{\partial^3 l(r_t \mid r_{t-1}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'\partial \boldsymbol{\theta}_i} \in \mathbb{R}^{d\times d}$$

In addition, we define the diagonal matrix $\mathcal{I}_T(\boldsymbol{\theta}) = \operatorname{diag}\{i_T(\boldsymbol{\theta})\} \in \mathbb{R}^{d\times d}$, where $i_T(\boldsymbol{\theta})$ denotes the vector of diagonal entries corresponding to the observed Fisher information. The Fisher information measures the amount of information that an observable random variable carries about an unknown parameter $\boldsymbol{\theta}$. Formally, it is defined as the variance of the score function. In Kristensen and Shin, it is defined as

$$i_T(\boldsymbol{\theta}) = \frac{1}{\nu_T}\sum_{t=1}^{T}\mathbb{E}\left[\frac{\partial^2 L_T(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = \mathbb{E}[H_T(\boldsymbol{\theta})] \in \mathbb{R}^{d\times d}.$$

Based on this, we define the standardized versions of the score and Hessian as

$$U_T(\boldsymbol{\theta}) = l_T^{-1/2}(\boldsymbol{\theta})S_T(\boldsymbol{\theta}), \quad V_T(\boldsymbol{\theta}) = l_T^{-1/2}(\boldsymbol{\theta})H_T(\boldsymbol{\theta})l_T^{-1/2}(\boldsymbol{\theta}), \quad W_{T,i} = l_T^{-1/2}G_{T,i}(\boldsymbol{\theta})l_T^{-1/2}$$

The following theorem, originally from Kristensen and Shin [43], is adapted to the notation used in this thesis. For the original formulation and the proof, we refer to [43].

Theorem 2.8 (Kristensen & Shin)**.** *Let $L_T(\boldsymbol{\theta})$ be the log-likelihood and let ν_T be a normalizing factor to ensure $L_T(\boldsymbol{\theta})$ is well-behaved asymptotically. Assume that the following conditions hold:*

(i) For some $\epsilon > 0$ and $l_T^{-1} = O_p(1)$, the parameter space is given by a sequence of local neighborhoods,

$$\boldsymbol{\Theta} = \left\{\boldsymbol{\theta} : \left\| l_T^{1/2}(\boldsymbol{\theta}^\star - \boldsymbol{\theta}) \right\| \le \epsilon\right\} \subseteq \mathbb{R}^d$$

(ii) $L_T(\boldsymbol{\theta})$ is three times continuously differentiable with its derivatives satisfying:

(a) $\left(\sqrt{\nu_T}\cdot U_T(\boldsymbol{\theta}^\star),\ V_T(\boldsymbol{\theta}^\star)\right) \xrightarrow{d} (S_{+\infty},\ H_{+\infty})$, with $H_{+\infty} < 0$ a.s.

(b) $\max_{j=1,\dots,d}\sup_{\boldsymbol{\theta}\in\Theta_T}\|W_T(\boldsymbol{\theta})\| = o_p(1)$

(iii) For some sequence $\nu_T \to +\infty$ it holds that

$$\frac{1}{\nu_T}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\widehat{L}_T(\boldsymbol{\theta}) - L_T(\boldsymbol{\theta})| = o_p(1/\sqrt{\nu_T}),$$

Equivalently, this can be written as

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\widehat{L}_T(\boldsymbol{\theta}) - L_T(\boldsymbol{\theta})| = o_p(\sqrt{\nu_T})$$

Then, for the quasi-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ and the asymptotically equivalent estimator $\boldsymbol{\theta}^*$, we have:

$$\sqrt{\nu_T}\cdot l_T^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = o_p(1).$$

Note that normalizing factor ν_T is only important for the theoretical derivations and not relevant for the actual implementation of our estimator since ν_T does not depend on $\boldsymbol{\theta}$. The standard choice is $\nu_T = T$, as is the case when the model is stationary [43].

Next, we will give the defintion of asymptotic normality and follow up with the Law of Large Numbers and the Central Limit Theorem for different cases.

Definition 2.23 (Asymptotic normality)**.** An estimator $\widehat{\boldsymbol{\theta}}_n = \widehat{\theta}_n(X_1,\dots,X_n)$ is said to be asymptoti-

cally normally distributed, if

$$a_n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\star) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\boldsymbol{\theta}^\star)), \text{ as } n \to +\infty,$$

where a_n is a scaling sequence that tends to $+\infty$, and $\sigma^2(\theta^\star)$ denotes the so-called asymptotic variance, assumed to be strictly positive.

Note that in most cases $a_n = \sqrt{n}$. This form of convergence means that, for large n, the distribution of $\hat{\boldsymbol{\theta}}_n$ can be approximated by a normal distribution centered at the true parameter $\boldsymbol{\theta}^\star$, with variance $\sigma^2(\boldsymbol{\theta}^\star)/n$.

Consistency and asymptotic normality of estimators are often justified by the Law of Large Numbers and the Central Limit Theorem respectively [66]. For the Law of Large Numbers, we distinguish three cases.

Theorem 2.9 (Law of Large Numbers)**.**

(*i*) ***Weak Law of Large Numbers***
Let $(\boldsymbol{X}_n)_{n\geq 1}$ be a sequence of independent, identically distributed random vectors with a marginal distribution that has finite expected value μ. For every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|\overline{\boldsymbol{X}}_n - \mu| > \varepsilon\right) = 0. \tag{2.21}$$

In other words, $\overline{\boldsymbol{X}}_n$ converges in probability to μ.

(*ii*) ***Strong Law of Large Numbers***
Let $(\boldsymbol{X}_n)_{n\geq 1}$ be a sequence of independent, identically distributed random variables with a marginal distribution that has finite expected value μ. Then,

$$\mathbb{P}\left(\lim_{n\to\infty} \overline{\boldsymbol{X}}_n = \mu\right) = 1. \tag{2.22}$$

In other words, $\overline{\boldsymbol{X}}_n$ converges almost surely to μ.

(*iii*) ***Law of Large Numbers for stationary and ergodic sequences:***
Let $(\boldsymbol{X}_n)_{n\geq 1}$ be a stationary and ergodic sequence of random variables with a marginal distribution that has finite expected value μ, and $\mathbb{E}[|X_n|] < +\infty$ for all n. Then,

$$\mathbb{P}\left(\lim_{n\to\infty} \overline{\boldsymbol{X}}_n = \mu\right) = 1. \tag{2.23}$$

In other words, $\overline{\boldsymbol{X}}_n$ converges almost surely to μ.

Note that the Law of Large Numbers for stationary and ergodic sequences is a generalization of the Strong Law of Large Numbers. Indeed, the Strong Law of Large Numbers is a special case where the random vectors are independent and identically distributed. A generalization of the Law of Large Numbers is the ergodic theorem, where the sequence (X_t) is not independent, but stationary and ergodic.

Theorem 2.10 (Ergodic theorem for stationary sequences)**.** *Let $(X_t)_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic sequence. Let f be a measurable function and $\mathbb{E}\Big[\Big|f(\dots, X_{t-1}, X_t, X_{t+1}, \dots)\Big|\Big] < +\infty$. Then,*

$$\frac{1}{n}\sum_{t=1}^{n} f(\dots, X_{t-1}, X_t, X_{t+1}, \dots) \xrightarrow{a.s.} \mathbb{E}\Big[f(\dots, X_{t-1}, X_t, X_{t+1}, \dots)\Big]$$

A related concept is the generalization of the Glivenko-Cantelli theorem to the stationary and ergodic setting, see [64] for a more general statement. This theorem will be used to prove asymptotic results on copula estimators in Chapter 7.

Theorem 2.11 (Glivenko-Cantelli Theorem from Tucker)**.** *Let $(X_n)_{n\geq 1}$ be a strictly stationary and ergodic sequence of random variables. If $(\widehat{F}_n(x))_{n\geq 1}$ denotes the associated sequence of empirical distribution functions, then*

$$\mathbb{P}\Big(\sup_{x\in(-\infty,+\infty)} |\widehat{F}_n(x) - F(x)| \to 0\Big) = 1,$$

where $F(x)$ denotes the CDF of X_1, and the empirical CDF is defined as $\widehat{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i\leq x\}}$.

Next, we present the classical Central Limit Theorem.

Theorem 2.12 (Central Limit Theorem)**.** *Let $(X_n)_{n\geq 1}$ be independent and identically distributed random variables with finite mean μ and finite variance σ^2. Then,*

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Equivalently, for all $z \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq z\right) = \Phi(z),$$

where $\Phi(z)$ denotes the cumulative distribution function of the standard normal distribution.

A generalization of the classical CLT, as stated in Theorem 2.12, is the CLT for martingales. This will be used in proving asymptotic normality for our proposed estimator in Chapter 6. Prior to stating the martingale CLT, we introduce the necessary supporting theory on martingales based on [11] and [31].

Let $(\Omega, \mathcal{F}, \mathbb{P}; (\mathcal{F}_n)_{n\geq 0})$ be a filtered probability space. This means that we equip the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_n)_{n\geq 0}$. A filtration is defined as a non-decreasing sequence of sub-σ-algebras of the σ-algebra $\mathcal{F}$, i.e. $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}$. The σ-algebra $\mathcal{F}_n$ represents the information available up to time n.

Definition 2.24 (Martingale)**.** A sequence of real-valued random variables $(M_n)_{n\geq 0}$ is called a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 1}$ if it satisfies:

$$(i)\ M_n \in \mathcal{F}_n, \quad (ii)\ \mathbb{E}[|M_n|] < \infty, \quad (iii)\ \mathbb{E}[M_n \mid \mathcal{F}_{n-1}] = M_{n-1}, \quad \text{for all } n \geq 1.$$

A related notion is that of a martingale difference sequence, which forms the basis for martingale central limit theorems.

Definition 2.25 (Martingale difference sequence)**.** A sequence $(Y_n)_{n\geq 1}$ of real-valued random variables is called a martingale difference sequence with respect to $(\mathcal{F}_n)_{n\geq 1}$ if it satisfies:

$$(i)\ Y_n \in \mathcal{F}_n, \quad (ii)\ \mathbb{E}[|Y_n|] < \infty, \quad (iii)\ \mathbb{E}[Y_n \mid \mathcal{F}_{n-1}] = 0, \quad \text{for all } n \geq 1.$$

An example of a martingale difference process is given by

$$Y_n := M_n - M_{n-1},$$

where M_n is a martingale. In this case, the sequence (Y_n) captures the increments of the martingale and satisfies the conditions of a martingale difference sequence. To analyze the variation of a martingale or martingale difference sequence, we introduce the concept of predictable quadratic variation.

Definition 2.26 (Predicable quadratic variation)**.** Let $(Y_n)_{n\geq 1}$ be a martingale difference sequence with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$, and let M_n be a martingale such that $Y_n := M_n - M_{n-1}$. The predictable quadratic variation of $(M_n)_{n\geq 0}$ is defined by

$$V_n^2 := \langle M \rangle_n := \sum_{j=1}^{n} \mathbb{E}[Y_j^2 \mid \mathcal{F}_{j-1}],$$

where each $\mathcal{F}_{j-1}$ represents the information available up to time $j-1$. The process $\langle M \rangle_{n\geq 0}$ is a predictable, right-continuous, and increasing process starting at zero.

Moreover, the unconditional variance is given by

$$s_n^2 := \mathbb{E}[V_n^2] = \mathbb{E}[\langle M \rangle_n] = \sum_{j=1}^{n} \mathbb{E}[Y_j^2].$$

We now state the Central Limit Theorem for martingales. This formulation follows from the foundational work of Brown [11].

Theorem 2.13 (Central Limit Theorem for Martingales)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a filtration $(\mathcal{F}_n)_{n\geq 0}$). Let M_n be a martingale, and $(Y_n)_{n\geq 1}$ be a martingale difference sequence both with respect to $(\mathcal{F}_n)_{n\geq 1}$ such that $S_n = M_n - M_{n-1}$ for all n. Let V_n^2 denote the predictable quadratic variation, and s_n^2 denote the variance. Assume the following two conditions hold:*

(i) ***Lindeberg condition.*** *For all $\varepsilon > 0$,*

$$\frac{1}{s_n^2} \sum_{j=1}^{n} \mathbb{E}\left[Y_j^2 \cdot \mathbf{1}_{\{|S_j| > \varepsilon s_n\}} \right] \longrightarrow 0 \quad \text{as } n \to \infty.$$

(ii) ***Stable predictable quadratic variation.***

$$\frac{V_n^2}{s_n^2} \xrightarrow{p} 1 \quad \text{as } n \to \infty.$$

Then the normalized martingale converges in distribution to the standard normal distribution:

$$\frac{M_n}{s_n} \xrightarrow{d} \mathcal{N}(0,1).$$

2.7 Markov Networks

In probabilistic modeling, understanding the dependencies and independencies among random variables is important. Markov networks, also known as Markov random fields, are a framework to represent these relationships using graph theory. This section introduces Markov networks as a tool that will be used modeling the dependency structures underlying the trade/no-trade behavior of the voluntary carbon asset. In Chapter 5, we will discuss the trade/no-trade behavior among the voluntary carbon credits in more detail. Unlike Bayesian networks, which are directed and are used for causal inference, Markov networks use undirected graphs to capture symmetric dependency relationships [49].

Let $U = \{X_1, X_2, \dots, X_n\}$ represent a finite set of elements corresponding to random variables. A dependency model M is a collection of conditional independence statements among subsets of U. Specifically, for disjoint subsets $X, Y, Z \subseteq U$, we write $I(X, Z, Y)_M$ to denote that the variables in X are conditionally independent of those in Y given the variables Z, according to the model M. Here, I denotes a subset of triplets (X, Z, Y) for which the assertion of conditional dependence holds as defined previously. Any joint probability distribution over U implicitly defines such a dependency model, since we can test the truth of each independence statement using the criterion $\mathbb{P}(X \mid Y, Z) = \mathbb{P}(X \mid Z)$ whenever $\mathbb{P}(Y, Z) > 0$.

To formally approach Markov networks, we introduce some basic elements of graph theory. Let $G = (V, E)$ be an undirected graph, where V is a finite set of vertices and $E \subseteq V \times V$ is a set of edges. Here, each vertex $v \in V$ corresponds to a random variable $X_v \in U$ and an edge $(u, w) \in E$ represents a direct dependency between X_v and X_w. Once this correspondence is established, we will no longer distinguish between U and V, and will denote the graph as $G = (U, E)$, and interpret each edge $(X_v, X_w) \in E$ as representing a symmetric probabilistic dependency between v and w [49]. A central concept in this graphical framework is separation. For subsets $X, Y, Z \subseteq U$, we say that Z separates X and Y in the graph G, denoted by $\langle X \mid Z \mid Y \rangle_G$, if every path from a node in X to a node in Y passes through at least one node in Z [49]. This graph-theoretic notion corresponds to the probabilistic notion of conditional independence. This is also known as the global Markov property [12]. Furthermore, the local Markov property means that each variable is conditionally independent of all other non-neighboring variables given its neighbors. To formalize the graphical encoding of dependencies, we define three types of maps.

Definition 2.27 (I-map, D-map, and P-map)**.** Let G be an undirected graph, M be a dependency model, then we distinguish the following maps:

- An undirected graph G is an **I-map** (independency map) of the dependency model M if:

$$\langle X \mid Z \mid Y \rangle_G \Rightarrow I(X, Z, Y)_M.$$

- Conversely, an undirected graph G is a **D-map** (dependency map) of M if:

$$I(X, Z, Y)_M \Rightarrow \langle X \mid Z \mid Y \rangle_G.$$

- Lastly, an undirected graph G is a **P-map** (perfect map), if G is both a D-map and an I-map, or equivalently:

$$I(X, Z, Y)_M \Leftrightarrow \langle X \mid Z \mid Y \rangle_G.$$

These implications in Definition 2.27 define whether the graph captures all independencies (I-map), all dependencies (D-map), or both. Note that the empty graph is a trivial D-map, and a complete graph is a trivial I-map [49]. Moreover, it is important to mention that in a dependency map, neighboring nodes represent variables that are dependent, though not all dependent variables are connected in the graph. Similarly, in an independency map, non-neighboring nodes represent variables that are independent, but not all independent variables appear as non-neighbors in the graph [49]. Now, let us formalize the concept of Markov Networks.

Definition 2.28 (Markov Network)**.** A Markov network, or Markov random field, is a pair (G, P), where $G = (U, E)$ is an undirected graph and P is a joint probability distribution over the variables U, such that G is a minimal I-map of P:

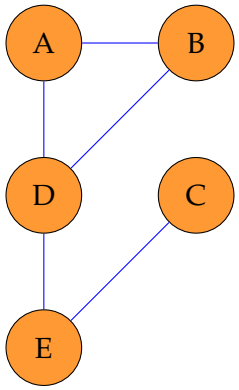
$$\langle X \mid Z \mid Y \rangle_G \Rightarrow I(X, Z, Y)_P,$$

for all disjoint $X, Y, Z \subseteq U$, and the removal of any edge in G would invalidate this property.

Example 1 (Markov Network)**.** *A Markov Random Field can be represented by an undirected graph* $G = (V, E)$*. In this example, the set of vertices* V*, and the set of edges* E *are given by*

$$V = \{A, B, C, D, E\}, \quad E = \{\{A, B\}, \{A, D\}, \{B, D\}, \{D, E\}, \{C, E\}\},$$

where each edge indicates a conditional dependency between the connected variables. Here, we have that A *depends on* B *and* D*.* B *depends on* A *and* D*.* D *depends on* A*,* B*, and* E*.* E *depends on* D *and* C*.* C *depends only on* E*.*



However, undirected graphs have a limitation in representing certain dependency structures, such as induced dependencies. Therefore, Pearl and Paz introduced five axioms that characterize a class of models that can be faithfully represented graphically [23].

Theorem 2.14 (Graph-isomorphism)**.** *Let $W, X, Y, Z \subseteq U$ be disjoint sets of random variables and let $\gamma \in U$ be a random variable. A necessary and sufficient condition for a dependency model M to be graph-isomorph is that $I(X, Z, Y)_M$ satisfies the following five independent axioms. The subscript M is removed for clarity purposes:*

(*i*) ***Symmetry:*** $I(X, Z, Y) \iff I(Y, Z, X)$*, independence between X and Y given Z mutual.*
(*ii*) ***Decomposition:*** $I(X, Z, Y \cup W) \Rightarrow I(X, Z, Y) \wedge I(X, Z, W)$*, independence from a union implies independence from each part.*
(*iii*) ***Intersection:*** $I(X, Z \cup W, Y) \wedge I(X, Z \cup Y, W) \Rightarrow I(X, Z, Y \cup W)$
(*iv*) ***Strong Union:*** $I(X, Z, Y) \Rightarrow I(X, Z \cup W, Y)$*, adding more disjoint subsets to the conditioning set preserves independence.*
(*v*) ***Transitivity:*** $I(X, Z, Y) \Rightarrow I(X, Z, \gamma)$ *or* $I(\gamma, Z, Y)$

Since exact isomorphic representations of informational dependencies are often unattainable, we restrict attention to I-maps that capture all independencies. While some true independencies may be omitted, we require that such omissions be minimized to ensure the graph contains no superfluous edges.

Two related concepts that are important are a Markov blanket $BL_I(\alpha)$ and a Markov boundary of α denoted by $B_I(\alpha)$. Formally, they are defined as

Definition 2.29 (Markov blanket and Markov boundary)**.** A Markov blanket $BL_I(\alpha)$ of a real-valued random variable $\alpha \in U$ is any subset $S \subseteq U$ of elements for which

$$I(\alpha, S, U - S - \alpha)). \tag{2.24}$$

A set is called a Markov boundary of α, denoted $B_I(\alpha)$ if it is a minimal Markov blanket of α, i.e. none of its proper subsets satisfy (2.24). Essentially, the Markov blanket is equal to the neighbourhood of a node.

The Markov boundary $B_I(\alpha)$ can be seen as the smallest set of elements that shields α from the influence of all other elements. Note that the Markov boundary is nonempty [49]. The following result links Markov networks to Markov boundaries:

Corollary 2.1. *The Markov network G_0 for any strictly positive distribution can be constructed by linking each real-valued random variable α to all members of its Markov boundary $B_I(\alpha)$.*

Corollary 2.1 is particularly useful because Markov boundaries $B_I(\alpha)$ often represent the most direct influences on a variable α. These influences may correspond to immediate effects, underlying causes, or defining features of a concept. If the Markov boundary is known for each variable, the overall structure of the distribution can be reconstructed by connecting each variable to the members of its boundary. Intuitively, this is like assembling a global map from local neighborhoods. Importantly, the resulting graph constitutes a valid I-map, meaning that many global independencies can be inferred through local separation tests on the graph structure [49]. There are two standard methods for constructing a minimal I-map given a dependency model:

- **Edge deletion:** Begin with a complete graph and remove any edge (α, β) if there exists a separating set Z such that $I(\alpha, Z, \beta)_M$ holds.

- **Markov boundary aggregation:** For each node α, determine $B_G(\alpha)$ and connect α to each of its boundary nodes.

Once the graph structure is specified, the next step is to quantify the Markov network. This is achieved by defining a Gibbs distribution over the network and using the concept of cliques. A clique in graph theory refers to a subset of vertices such that every two distinct vertices are adjacent. That is, a clique is a subset that is complete. A probability distribution over a set of variables X forms a Gibbs distribution relative to a graph G if it can be factorized over the maximal cliques of G as:

$$\mathbb{P}(X = x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \phi_C(x_C),$$

where $\mathcal{C}$ denotes the set of maximal cliques in G, $\phi_C(x_C) : X_C \to (0, \infty)$ are non-negative potential functions defined over the variables within each clique C, and Z is a normalizing constant called the partition function. Each potential function $\phi_C(x_C)$ captures the interaction strength among variables within the clique C. The potentials need not integrate to one individually and are defined only up to a multiplicative constant, because normalization is achieved globally through Z. Specifically, the partition function is given by:

$$Z = \sum_{x \in \mathcal{X}} \prod_{C \in \mathcal{C}} \phi_C(x_C),$$

where $\mathcal{X}$ denotes the full state space of all variables, i.e. the space of all possible configurations of the random variables. From Equation (2.7), it becomes clear that computing the joint probability distribution is computationally intensive.

The Hammersley-Clifford theorem provides the theoretical bridge between the topological structure of a Markov network and a consistent probabilistic interpretation.

Theorem 2.15 (Hammersley-Clifford)**.** *Let $G = (V, E)$ be an undirected graph and $X_1, \ldots, X_n$ be random are random variables that take on a finite number of values. If P a strictly positive probability distribution and satisfies the local Markov property with respect to G, then it factors with respect to G. In other words, G is an I-map of P.*

The important result following from Theorem 2.15 is that for strictly positive distributions, the global Markov property, the local Markov property, and factoring with respect to the graph G are equivalent [12]. In other words, the Hammersley-Clifford theorem formalizes the connection between factorization and conditional independence in Markov random fields [49]. This factorization captures both the structure and strength of dependencies: the graph encodes which variables interact directly, while the potentials quantify how different joint configurations are weighted relative to each other. In this way, Markov networks provide a representation of multivariate dependencies, avoiding the need to specify the full joint distribution explicitly.

2.8 Hypothesis testing for contingency tables

Hypothesis testing is a fundamental statistical procedure used to make inferences about populations based on sample data. In its basic form, it involves the formulation of two competing hypotheses. The null hypothesis $\mathcal{H}_0$, which represents a baseline or default assumption, and the alternative hypothesis $\mathcal{H}_1$, which represents the presence of an effect or association [7]. One example of hypothesis

testing that we will use in this research is Fisher's exact test used in the analysis of contingency tables [7]. In the context of this research, we will use Fisher's exact test to study temporal dependence within assets and cross-dependence between assets in trade/no-trade data, as discussed in Chapter 4.

Fisher's exact test is a method for testing the null hypothesis that two categorical variables are independent, conditioned on fixed marginal totals. Let us consider the following generalized contingency table:

	Category I	Category II	
Group 1	x_{11}	x_{12}	$n_{1\cdot} = x_{11} + x_{12}$
Group 2	x_{21}	x_{22}	$n_{2\cdot} = x_{21} + x_{22}$
	$n_{\cdot 1} = x_{11} + x_{21}$	$n_{\cdot 2} = x_{12} + x_{22}$	$n_{\cdot\cdot} = n_{1\cdot} + n_{2\cdot} = n_{\cdot 1} + n_{\cdot 2}$

Table 2.1: *A general* 2×2 *contingency table*

In the context of a 2×2 contingency table, the test statistic is a function of the observed cell counts. To derive its asymptotic distribution, we must specify assumptions regarding the structure of the table. In this setting, we treat the row totals $n_{1\cdot}$ and $n_{2\cdot}$ as fixed, while considering the cell counts x_{11} and x_{21} as realizations of random variables X_{11} and X_{21}, respectively. Knowledge of x_{11} and x_{21}, together with the fixed row totals, fully determines the entire table [41]. To model the randomness, we assume that $X_{11} \sim \text{Binomial}(n_{1\cdot}, \theta_1)$ and $X_{21} \sim \text{Binomial}(n_{2\cdot}, \theta_2)$, where we denote the unknown parameters as $\theta = (\theta_1, \theta_2) \in \Theta = [0, 1]^2$. This formulation allows us to write the following hypothesis test setup, where we test $\mathcal{H}_0 : \theta \in \Theta_0$ against $\mathcal{H}_1 : \theta \in \Theta_1$, with $\Theta_0 = \{(\theta_1, \theta_2) \in \Theta : \theta_1 = \theta_2\}$ and $\Theta_1 = \Theta \setminus \Theta_0$ [7]. Equivalently, the hypotheses can be stated as a two-sided test:

$$\mathcal{H}_0 : \theta_1 = \theta_2 = \theta \text{ against } \mathcal{H}_1 : \theta_1 \neq \theta_2, \tag{2.25}$$

Although this binomial assumption seems quite natural, one can argue this is actually quite a strong claim to make. Indeed, recall that for a binomial trial, one should need a set of independent Bernoulli experiments with the same probability of success in each experiment [41].

In general, for small sample sizes, the usual asymptotic methods such as the chi-square test may not give accurate results. In these cases, exact inference methods provide a more reliable alternative [70]. These methods are called exact because they are derived from the true sampling distribution without relying on asymptotic approximations [41]. One of the most well-known exact methods is Fisher's exact test, originally introduced by Ronald Fisher in 1934 [22].

We now give a derivation of Fisher's exact test in the context of a 2×2 contingency table, as defined in Table 2.1. Assuming a binomial sampling framework, Fisher argues that under the null hypothesis $H_0 : \theta_1 = \theta_2 = \theta$, the probability of observing x_{11} successes in group 1 is:

$$\mathbb{P}_\theta(X_{11} = x_{11}) = \binom{n_{1\cdot}}{x_{11}} \theta^{x_{11}} (1-\theta)^{x_{12}}. \tag{2.26}$$

An analogous expression holds for group 2. Thus, the joint probability of observing the table is:

$$\mathbb{P}_\theta(\mathbf{X} = \mathbf{x}) = \mathbb{P}_\theta\left(\begin{pmatrix} X_{11} \\ X_{21} \end{pmatrix} = \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix}\right) = \binom{n_{1\cdot}}{x_{11}} \binom{n_{2\cdot}}{x_{21}} \theta^{n_{\cdot 1}} (1-\theta)^{n_{\cdot 2}}, \tag{2.27}$$

where $n_{\cdot 1} = x_{11} + x_{21}$ and $n_{\cdot 2} = x_{12} + x_{22}$. In this derivation, Fisher implicitly assumed that the row margins $(n_{1\cdot}, n_{2\cdot})$ are fixed. Fisher noted that the factor $\theta^{n_{\cdot 1}}(1-\theta)^{n_{\cdot 2}}$ is constant across all tables with the same column margins [41]. This leads to a conditional probability under the null hypothesis, given the row margins:

$$\mathbb{P}_\theta(\mathbf{X} = \mathbf{x} \mid X_{11} + X_{21} = n_{\cdot 1}) = \frac{\mathbb{P}_\theta(X_{11} = x_{11}, X_{21} = n_{\cdot 1} - x_{11})}{\mathbb{P}_\theta(X_{11} + X_{21} = n_{\cdot 1})}.$$

Substituting from Equation (2.27), and using the fact that the sum of independent binomial variables with common success probability θ is again binomial, we obtain:

$$\begin{aligned}\mathbb{P}_\theta(\mathbf{X} = \mathbf{x} \mid X_{11} + X_{21} = n_{\cdot 1}) &= \frac{\binom{n_{1\cdot}}{x_{11}}\binom{n_{2\cdot}}{n_{\cdot 1}-x_{11}}\theta^{n_{\cdot 1}}(1-\theta)^{n_{\cdot\cdot}-n_{\cdot 1}}}{\binom{n_{\cdot\cdot}}{n_{\cdot 1}}\theta^{n_{\cdot 1}}(1-\theta)^{n_{\cdot\cdot}-n_{\cdot 1}}} \\ &= \frac{\binom{n_{1\cdot}}{x_{11}}\binom{n_{2\cdot}}{n_{\cdot 1}-x_{11}}}{\binom{n_{\cdot\cdot}}{n_{\cdot 1}}}.\end{aligned} \tag{2.28}$$

Equation (2.28) corresponds to the probability mass function of the hypergeometric distribution with parameters $(n_{\cdot\cdot}, n_{1\cdot}, n_{\cdot 1})$, and serves as the basis for Fisher's exact test. Clearly, the probability distribution now no longer depends on θ, and therefore we can denote this expression more simply by:

$$\mathbb{P}_\theta(\mathbf{X} = \mathbf{x} \mid X_{11} + X_{21} = n_{\cdot 1}) = \frac{\binom{n_{1\cdot}}{x_{11}}\binom{n_{2\cdot}}{n_{\cdot 1}-x_{11}}}{\binom{n_{\cdot\cdot}}{n_{\cdot 1}}}.$$

The test statistic $T(X) = X_{11} + X_{21}$ is thus a sufficient statistic for θ [7].

An important challenge in applying Fisher's test arises when defining p-values for two-sided tests such as (2.25). The discrete and possibly asymmetric nature of the hypergeometric distribution complicates the notion of "more extreme" outcomes under the null hypothesis [70]. However, there exists a general form for the two-sided p-value:

$$p_F(x) = p_F(x_{11}, x_{21}) = \sum_{i \in E_x} \mathbb{P}(X_{11} = i \mid X_{11} + X_{21} = n_{\cdot 1}),$$

where the set E_x of outcomes contains all the table outcomes for x_{11} that can be considered as more extreme than x_{11}. The definition of this set depends on the chosen criterion of extremeness. One common definition, used in the `R` implementation of Fisher's exact test, is:

$$E_x := \{i \in \{0, \ldots, n_{1\cdot}\} : T(i, n_{\cdot 1} - i) \leq T(x_{11}, x_{21})\}, \tag{2.29}$$

where $T(x_{11}, x_{21}) = \mathbb{P}(X_{11} = x_{11} \mid X_{11} + X_{21} = n_{\cdot 1})$ is used as the test statistic. In this framework, the null hypothesis is rejected when observing $\mathbf{x}$ such that $T(\mathbf{x})$ is small [41].

An important concept in the context of the Fisher exact test is the odds ratio. In the context of a general 2×2 contingency table as shown in Table 2.1, the odds ratio (OR) provides a natural measure of association between the two groups and the two categories. Formally, it is defined as

$$\text{OR} = \frac{\frac{x_{11}}{x_{12}}}{\frac{x_{21}}{x_{22}}} = \frac{x_{11} \cdot x_{22}}{x_{12} \cdot x_{21}},$$

where x_{11} and x_{12} denote the counts of Group 1 in Categories I and II, and x_{21} and x_{22} the corresponding counts for Group 2. An odds ratio equal to one indicates independence, values greater than one suggest that the odds of belonging to Category I are higher for Group 1 than for Group 2, and values below one imply the opposite [62]. Note that the odds ratio is symmetric in the two events: whether we describe it in terms of groups or categories, the interpretation remains the same, and no causal direction is implied [62]. In conclusion, in a Fisher test, the p-value assesses whether the observed association is statistically significant, while the odds ratio quantifies its strength and direction.

3 Carbon Markets

Carbon credits are tradable permits representing the offset of a specific amount of carbon dioxide or other greenhouse gases (GHGs). They are a financially sustainable mechanism to handle climate change. It allows organizations to offset their emissions by buying projects that reduce, remove, or avoid GHG emissions elsewhere. The concept is rooted in the principle of carbon offsetting, where companies, governments, and individuals invest in environmental projects to balance out their own carbon footprints. In this chapter, we will explain carbon markets, specifically the Voluntary Carbon Market (VCM), the pricing of carbon credits, and its dependency with other asset classes.

The evolution of carbon markets began with the Kyoto Protocol in 1997, which introduced mechanisms like the Clean Development Mechanism (CDM) and Joint Implementation (JI) [65]. These enabled emission reductions in one country to be used towards the targets of another. The Paris Agreement in 2015 shifted the global approach to a more inclusive system of national pledges and introduced Article 6, which provides a framework for international carbon trading. The VCM developed in parallel to the CDM and Article 6, which were always intended to only serve countries for compliance purposes. The VCM did build on the CDM methodologies (some are still used today), but made the project interventions available to private actors instead of governments.

Over time, the voluntary carbon market has seen rapid growth, especially post-2015, driven by corporate net-zero commitments and the rise of ESG-investing [5]. This growth has been enabled by technological innovations such as digital monitoring and satellite-based verification tools that enhance market transparency and efficiency. Despite their benefits, carbon markets face criticism over issues such as additionality, permanence, leakage, double-counting, and greenwashing [3]. To address these concerns, initiatives like the Integrity Council for the Voluntary Carbon Market (ICVCM) and the Voluntary Carbon Markets Integrity Initiative (VCMI) are working to enhance market integrity and transparency [65].

3.1 The structure of Carbon Markets

Formally, the carbon market consists of the compliance carbon markets and voluntary carbon markets. The compliance carbon market is a government-regulated system where companies must adhere to legally binding emission limits. It typically operates under cap-and-trade programs, where a regulatory authority sets an overall cap on emissions and distributes or auctions allowances to emit-

ters. These allowances represent the legal right to emit a specified amount of CO_2, which is 1000kg CO_2 equivalent. Firms that emit less than their quota can sell surplus allowances, while those exceeding their cap must purchase additional credits or face penalties [3]. Key compliance markets include the European Union Emissions Trading System (EU ETS), California's Cap-and-Trade Program, and China's National ETS. Each of these systems has distinct rules for allowance distribution, credit issuance, and trading, but all aim to drive emissions reductions in line with policy goals.

The voluntary carbon market operates independently of regulatory mandates, allowing entities to purchase carbon credits as part of their sustainability strategies. Participation is typically driven by corporate social responsibility, consumer expectations, investor pressure, and anticipation of future regulations [3]. The key distinction between allowances in the compliance carbon market and carbon credits in the voluntary carbon market lies in their function. Allowances in the compliance carbon market grant the right to emit one ton of CO_2eq, whereas carbon credits in the voluntary market represent the offset of one ton of CO_2eq elsewhere to compensate for emissions. Voluntary carbon credits fall into three broad categories: removal, reduction, and avoidance. First, removal credits come from projects that physically extract CO_2 from the atmosphere, such as afforestation and direct air capture. As of the end of 2023, these account for approximately 3% of certified credits on the VCM [30]. Second, reduction credits are generated from activities that lower emissions at their source—for example, through energy efficiency measures or fuel switching. They represent about 22% of certified credits in the market [30]. Lastly, avoidance credits are associated with projects that prevent emissions from occurring, such as forest conservation efforts that avert deforestation and the release of stored carbon. These make up the remaining 75% of certified credits [30].

The VCM can be further categorized into nature-based and technology-based solutions [5]. Nature-based solutions include ARR (afforestation, reforestation, and revegetation), REDD+ (Reducing Emissions from Deforestation and Forest Degradation), mangrove restoration, IFM (improved forest management), and biochar projects. In addition to carbon sequestration, they offer biodiversity and ecosystem co-benefits but face challenges like land-use change and impermanence. Biochar is contested as a nature-based solution because of the technology needed to create biochar. Technology-based solutions, on the other hand, such as direct air capture, carbon capture and storage, enhanced weathering, mineralization, and waste-to-energy projects, typically offer greater permanence but are often more expensive and less scalable in the short term.

Voluntary carbon credits (VCCs) are intangible assets, and unlike compliance allowances, they are not standardized [3]. Each credit reflects the characteristics of its underlying project, including co-benefits, vintage, verification standards, and geography. This heterogeneity results in over-the-counter transactions being the dominant mechanism in the VCM, leading to limited price transparency and liquidity [3]. We will discuss carbon credit pricing and trading in the next subsection.

Market participants include project developers, certification bodies (e.g. Verra, Gold Standard, ACR), rating agencies (e.g. Sylvera, BeZero, Calyx Global), brokers, corporate buyers, and intermediaries such as offsetting platforms. The lifecycle of a credit typically involves project design, validation, issuance, trading, and eventual retirement.

3.2 Pricing and trading of Carbon Credits

Pricing in the carbon credit market is influenced by both market dynamics and project characteristics. The two dominant pricing mechanisms are [3]:

- Over-the-counter (OTC) trading: Negotiated directly between project developers and end-users, offering flexibility in pricing and contract terms but resulting in low transparency. In fact, we do not know the price of an OTC carbon credit since this is not market-driven. Importantly, the majority of high-quality carbon credits are traded over-the-counter.
- Exchange-based trading: Conducted on standardized platforms such as the Chicago Climate Exchange (CCX), Xpansiv CBL, and AirCarbon Exchange. These exchanges provide greater transparency, standardization, and liquidity. Note that these platforms also often use standardized contracts, which is the key instrument that facilitates scalable and transparent exchange.

Voluntary credits are increasingly being treated as commodities, but they differ due to their lack of standardization. Pricing is influenced by a mix of factors:

1. Credit quality: Verified and well-rated credits command higher prices due to trust in their environmental integrity. Key components of quality include permanence, additionality, leakage and co-benefits.
2. Vintage: Recent credits are preferred as they reflect current standards and methodologies.
3. Project type: Each type of credit has different characteristics. For example, nature-based solutions typically have higher co-benefits, and technology-based removals (like BECCS) can be more expensive due to high implementation costs. However, tech-based credits often have a longer permanence.
4. Location: Projects in jurisdictions with strong governance tend to command higher prices, as they are perceived as more reliable. Additionally, projects in the Global North (EU, US, AUS/NZ) often face higher opportunity costs, input prices, and labor costs, which contribute to higher credit prices.
5. Supply and demand: Growing corporate demand against a limited supply of high-quality credits creates upward price pressure. Moreover, there is a shift in demand from low to high quality credits.
6. Policy landscape: Regulatory developments influence credit valuation.
7. Co-benefits and ESG alignment: Projects offering biodiversity, social, and economic benefits tend to be valued higher by ESG-focused investors.

The primary challenge in pricing voluntary carbon credits is the market's diversity, including various projects, methodologies, and geographies [3] [47]. This fragmentation makes it difficult to establish a uniform price for voluntary carbon credits. To successfully compare the Voluntary Carbon Market with other commodity markets, it is important to establish a standardized pricing approach. We achieve this by categorizing VCM credits into removal and avoidance credits, recognizing the price differentiation between these types [5]. We further subdivide these categories into three groups: nature-based solutions, hybrid solutions, and technology-based solutions, as outlined in [5]. See Figure 3.1 for a visualization of this categorization.

In Chapter 4 and Appendix A, we describe which voluntary carbon credits we have chosen for our research and provide some characteristics, also based on the categorization of Figure 3.1.

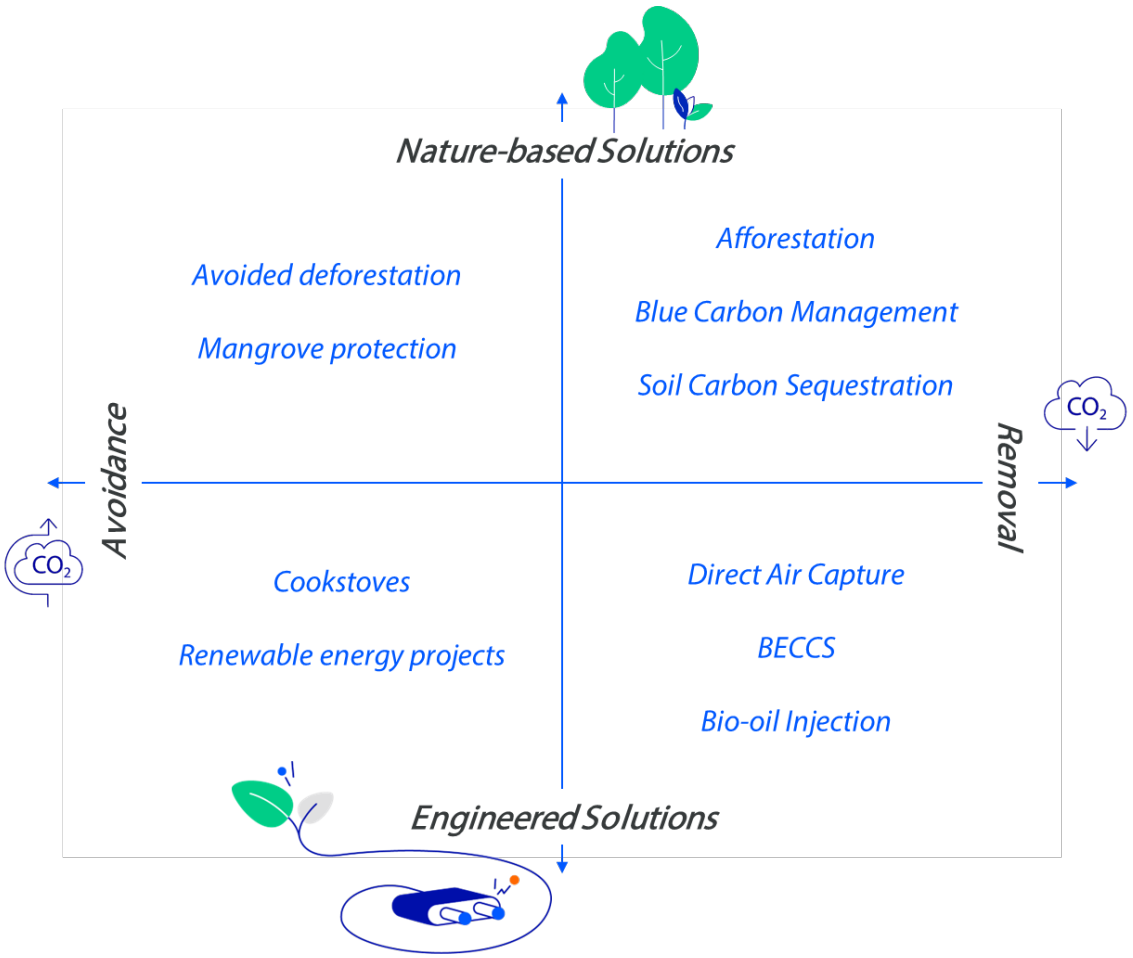


Figure 3.1: *Non-exhaustive categorization of the Voluntary Carbon Credit market*

3.3 Market dependencies between carbon credits and commodities

With carbon credits increasingly framed as a rising commodity class, a natural question arises: how does this emerging market interact with traditional commodity and financial markets? This section reviews existing literature and hypotheses regarding the dependence of voluntary carbon credit prices on other commodity prices and macroeconomic variables. Given the relative novelty of the voluntary carbon market, rigorous empirical studies remain limited, but academic and institutional interest is growing. Researchers seek to understand whether carbon credits behave like conventional financial assets — such as correlating with energy prices or exhibiting systemic risk traits — or if they remain largely idiosyncratic.

Research of MSCI shows there exists a positive seasonal correlation between EU ETS prices and oil, ranging from –0.1 to 0.6 [68]. However, these conclusions primarily relate to compliance markets. For voluntary credits, the extent of financial integration and price co-movement with other asset classes remains an open question. A leading hypothesis is that as voluntary credits become more standardized and liquid, their price behavior may begin to reflect broader market trends, such as energy cycles or investor sentiment. In the following paragraphs, we discuss energy, agricultural,

and metal commodities in relation with carbon credits.

Energy commodities such as oil, natural gas, and coal are affecting emissions and are known to influence compliance markets. For example, rising gas prices may lead power producers to substitute toward coal, increasing emissions and driving up carbon allowance prices [68]. In 2022, high oil and gas prices contributed to a notable shift in EU ETS market dynamics, with correlation patterns changing accordingly [1]. On the other hand, in the voluntary market, the relationship with energy prices is more indirect. For example, higher fossil fuel prices may increase the financial attractiveness of offset projects (such as renewables), or stimulate climate-related investments. On the other hand, high energy costs may reduce corporate discretionary budgets, which decreases demand for voluntary offsets. Empirical evidence of correlation between VCM prices and energy commodities is scarce. The VCM has remained segmented from compliance and energy-linked markets—at least thus far. However, as voluntary credits begin trading on commodity exchanges, these linkages may intensify. This presents an important empirical hypothesis for research.

Shifting to agricultural commodities, many voluntary offset projects are inherently tied to land use, such as REDD+, afforestation, and soil carbon initiatives. This suggests a potential connection to agricultural commodity markets. The viability of forest conservation projects, for example, may depend on the price of competing land uses such as soy, beef, or palm oil. If commodity prices for these drivers are high, the financial incentive to deforest increases, undermining offset projects or increasing the required carbon price to maintain conservation. Conversely, if agricultural commodity prices decline, landowners may be more willing to engage in carbon offsetting, increasing supply of credits and potentially lowering credit prices. However, it is not easy to switch frequently between agricultural commodities and carbon credits since a project developer has to commit to a carbon program for several years. Although the link between carbon credits and agricultural commodities is recognized qualitatively - especially in the context of project leakage or permanence - quantitative studies between agricultural commodity prices and voluntary carbon credits remain absent from the literature. However, analogies can be drawn from fields like weather derivatives or agricultural yield-linked instruments, where environmental variables and commodity economics are intrinsically linked.

Following to metal as a commodity, voluntary carbon markets, particularly in their early form, have exhibited characteristics of speculative trading environments. In thin, unregulated markets, price spikes can be driven more by sentiment than fundamentals. A notable example occurred in 2021 when blockchain projects such as KlimaDAO contributed to a sharp rise in credit prices [4]. Carbon credits in this context functioned more like speculative tokens than environmental instruments.

This raises questions about the potential correlation between carbon credits and other speculative assets such as cryptocurrencies, or even with safe-haven commodities like gold. Some investors argue that carbon credits can function as a hedge—similar to gold, but for climate transition risk. Research on compliance markets has suggested that carbon allowances have low correlation with both gold and equities, implying they may act as a separate asset class [67]. Although similar empirical studies for voluntary credits are lacking due to data constraints, the hypothesis of carbon credits as low-correlation instruments remains. This is increasingly attractive to institutional investors seeking ESG-aligned, diversifying assets.

3.3.1 Existing research and knowledge gaps

Academic research on the relation between voluntary carbon credits and conventional commodities remains limited. Conte and Kotchen conducted one of the few early econometric studies of voluntary carbon credit prices. They used hedonic regression to identify drivers such as project type, certification standard, and geographic region [14]. However, their study did not examine macroeconomic or cross-market dependencies. More recently, Trencher et al. found that a substantial share of corporate offset purchases was linked to low-quality credits that undermine the climate integrity of the VCM [63]. The study suggests that there was a demand-side preference for cheaper credits, rather than a sophisticated, quality-driven procurement strategy. However, currently, there is a shift to high-quality carbon credit demand. **In conclusion, this demand-driven nature may explain why VCM prices have not tracked commodity cycles or broader financial markets.**

As of our understanding, there are no existing studies researching the dependency between voluntary carbon credits and cross-market assets such as energy, agricultural or metal commodities. This creates a valuable opportunity for further empirical work. Interesting research directions we will dive into are:

- Testing for correlation or cointegration between VCM prices and energy commodities (e.g. oil, coal, gas).
- Assessing linkages with agricultural commodities relevant to nature-based projects.
- Evaluating the impact of macroeconomic variables (e.g. equity indices, volatility indices) on voluntary carbon demand.
- Investigating price convergence or divergence between voluntary and compliance carbon markets.
- Comparing volatility levels across VCM, compliance markets, and traditional commodities.

Understanding these relationships will be interesting for investors, regulators, and climate policymakers. This thesis contributes to this upcoming field by examining the statistical properties and dependencies of voluntary carbon credits in relation to broader commodity and financial markets.

3.4 Risks and trends in the voluntary market

The voluntary carbon credit market has experienced changes, transitioning to a more mature and quality-driven market. However, the VCM can still be seen as an early-stage market dealing with notable risks. One major concern is market uncertainty and price volatility, which have become particularly pronounced due to the widening price gap between removal and avoidance credits. Removal credit prices have remained stable at around $21 per tonne, whereas avoidance credit prices have continued to decline, falling below $4 per tonne [47]. This divergence signals a fundamental revaluation of credit types; buyers are more looking into high-quality removal credits. However, this situation also creates hesitancy among corporate buyers who seek price stability and predictable returns.

Another persistent challenge is ensuring the quality and integrity of carbon credits. An investigation by Guardian in 2023 revealed that over 90% of rainforest carbon offsets certified by Verra may be worthless [34]. This significantly undermined market confidence and contributed to a drop in

carbon credit prices. With increased scrutiny on project additionality, permanence, and co-benefits, many buyers are wary of accusations of greenwashing [3]. The introduction of frameworks such as the ICVCM's Core Carbon Principles and VCMI's Claims Code of Practice aims to address these concerns, yet enforcement remains inconsistent. The ongoing development of methodologies for monitoring, reporting, and verification is essential to building trust in the market and preventing the issuance of low-quality credits.

Regulatory uncertainty further complicates market participation. One of the key debates revolves around the role of voluntary credits in Scope 3 emissions accounting [47]. While voluntary markets offer companies a tool to offset emissions beyond their immediate operations, regulatory bodies such as the Science-Based Targets initiative (SBTi) have yet to establish clear guidelines on how these credits can be applied within net-zero commitments. At the same time, the integration of voluntary credits into compliance markets through Article 6.4 of the Paris Agreement remains a work in progress, raising questions about how voluntary credits can align with national decarbonization strategies.

Despite these risks, several trends signal the continued evolution of the voluntary carbon market. One of the most notable shifts is the growing preference for carbon removal credits. Permanent removal solutions such as direct air capture (DAC) and bioenergy with carbon capture and storage (BECCS) have seen a 300% increase in purchased volumes, reflecting corporate demand for long-term climate strategies. Similarly, nature-based removal solutions, such as afforestation and enhanced weathering, are becoming more attractive due to their relative affordability and scalability.

Another key trend is the diversification of buyer portfolios. Companies are moving away from short-term, one-off purchases and instead adopting multi-year procurement strategies that blend high-quality removal and avoidance credits. This transition ensures a balanced approach to immediate emission reductions while securing long-term access to durable removal solutions. Additionally, many buyers are integrating risk diversification strategies by sourcing credits from multiple geographies and technologies, mitigating potential project failures or policy shifts.

Market infrastructure and quality standards are also improving. The adoption of ICVCM's CCP has set a benchmark for high-integrity credits. Also, there have been several advancements in digital monitoring technologies, such as satellite imagery and blockchain tracking, to increase transparency and accountability. These improvements are important in differentiating high-quality projects from those with weaker verification processes.

Furthermore, the potential integration of voluntary and compliance markets continues to be explored. Regulatory bodies are examining ways to bridge these markets under Article 6.4 of the Paris Agreement, potentially allowing voluntary credits to be used within compliance frameworks. While this transition could bring greater liquidity and price stability, it also introduces new complexities in standardization and governance.

Corporate engagement in supply chain decarbonization represents another major development. With growing pressure to address Scope 3 emissions, companies are increasingly leveraging carbon markets as a tool for supplier engagement. New digital tools and market platforms are making it easier for suppliers to participate, enabling more seamless collaboration on emissions reduction initiatives.

4

Model formulation and financial asset data analysis

This chapter provides the data and modeling framework used to study the dynamics of voluntary carbon credits and conventional financial assets. We begin in Section 4.1 with a description of the dataset. It consists of seven voluntary carbon credit assets and six conventional financial assets. The carbon credits differ by project type and vintage. The conventional assets include equity indices, commodities, and fixed income instruments, chosen to represent mature and liquid markets. Asset selection was conducted in collaboration with TU Delft and Rabobank.

Section 4.2 introduces the multivariate zero-inflated GARCH-X (MZIGARCH-X) model, which is developed to accommodate zero-inflated time series data as found in illiquid financial markets. In such markets, including voluntary carbon credit markets, trading is irregular and many days show no price movement, resulting in zero returns. Standard GARCH models are not well-suited to this context, as they assume continuously evolving return series. The MZIGARCH-X model addresses this limitation by combining a multivariate trading indicator function with a multivariate GARCH-X structure. This allows us to model both return dynamics and the presence or absence of trading activity within a unified framework. The focus is on a GARCH(1,1)-X specification, but the extension to a general GARCH(p,q)-X model is also discussed.

In Section 4.3, we study the different types of dependence across assets. This is important for the multivariate modeling in later chapters. Finally, Section 4.4 presents a descriptive analysis of the voluntary carbon credit return series. We examine key time series properties such as stationarity and volatility clustering.

4.1 Structure of the data

The dataset consisting of the historical spot prices for voluntary carbon credits is obtained from Quantum Commodity Intelligence [54]. In this thesis, we consider the following voluntary carbon credits, categorized according to the classification shown in Figure 3.1:

Table 4.1: *Overview of voluntary carbon credit assets*

Asset	Type	Description
Nature-Based Offset Index	Nature-based, avoidance	An index aggregates land-use sector credits like REDD+ and wetland restoration.
REDD+	Nature-based, avoidance	A credit linked to projects that prevent deforestation across tropical regions in Africa, Asia and Latin America
Cookstoves Africa V2021	Tech-based, avoidance	A credit generated from clean cooking technologies that reduce fuel use and emissions.
Water Filtration Africa V2021	Tech-based, avoidance	A credit generated from water filtration projects that cut emissions by replacing traditional water boiling.
US IFM V2020	Nature-based, hybrid	A credit generated from improved forest management in U.S. forests, improving carbon sequestration.
Latin America ARR V2021	Nature-based, removal	A credit generated from afforestation and ecosystem restoration projects across Latin America.
Uruguay ARR V2021	Nature-based, removal	A credit generated from afforestation and reforestation efforts in Uruguay.

Additional information on the selected carbon credits can be found in Appendix A. As mentioned in Chapter 3, the voluntary carbon market is still in an early stage of development and remains relatively illiquid compared to established asset classes. Prices are not centrally regulated and often lack the standardization observed in traditional markets. As such, the dataset used in this thesis represents, to our knowledge, the most comprehensive set of market-based VCC price observations currently available. The number of observations per VCC asset ranges from approximately 450 to 1,050 historical price points. The price series have different starting dates, as trading for each VCC asset begins on different dates based on when the underlying offset projects were issued and verified. The earliest available carbon credit price series begins in 2021. To ensure consistency in our analysis, the series of the conventional assets are aligned with the earliest available date among the first traded VCC asset. Observations from the conventional assets that occur before this date are excluded, as no carbon credit data are available for that period.

The price data for the conventional assets are obtained from Yahoo Finance. In contrast to the VCCs, these conventional assets exhibit almost no zero returns, which indicates regular trading. The historical prices are available from 2007 to the present. This difference highlights the greater maturity of traditional financial markets compared to the emerging voluntary carbon market. The conventional assets included in our analysis are listed in Table 4.2. Additional information on the selected conventional assets can be found in Appendix A.

The six conventional assets listed in Table 4.2 are selected to investigate potential relationships between voluntary carbon credits and more liquid financial markets. The selection is grounded in

Table 4.2: *Overview of conventional financial assets*

Asset	Type	Description
EU ETS	Financial, carbon market	The European Union Emissions Trading System is the largest carbon market in the world, and serves as a benchmark for compliance carbon pricing.
NASDAQ	Financial, equity index	A stock market index including over 3,000 companies, serving as a barometer for U.S. equity performance, especially within the technology sector.
Natural Gas Futures	Energy, fossil fuel	Reflects spot market prices for natural gas, which is widely used in power generation and heating. Prices are sensitive to seasonal, geopolitical, and demand-supply dynamics.
Soybean Futures	Agricultural, soft commodity	Soybeans are a major global crop traded on commodity exchanges. Prices reflect agricultural yields, export demand, and biofuel policy interactions.
Renewable Energy Contract	Energy, green market	Represents a synthetic index of renewable energy sources such as wind and solar, indicating shifts in clean energy investment and policy incentives.
Volatility Index (VIX)	Financial, market sentiment	The CBOE Volatility Index measures market expectations of near-term volatility conveyed by S&P 500 option prices. It is widely used as a proxy for market uncertainty.

hypotheses developed in collaboration with Rabobank, based on macroeconomic trends. On the demand side, the cost of decarbonization plays an important role: as companies face rising costs for internal emission reductions, the use of VCCs becomes a relatively more attractive alternative. This substitution mechanism suggests potential links with macro indicators such as the VIX, which reflects risk sentiment, and with the EU ETS, the largest compliance carbon market. In addition, indices like the NASDAQ may also show potential links since this index captures large U.S. technology companies, many of which set voluntary climate targets. A declining stock performance may signal reduced capacity or willingness to invest in VCCs, whereas strong performance may support higher participation. As discussed in Chapter 3, demand-side shocks are expected to have the most immediate impact on VCC pricing, as supply adjusts only gradually due to long project development cycles, typically ranging from five to seven years.

On the supply side, agricultural commodity prices affect the relative attractiveness of land-based carbon projects: high prices for crops such as soy can reduce the economic appeal of conservation or reforestation, thereby limiting the supply of nature-based credits. Conversely, lower agricultural returns may increase the availability of such credits. Similarly, the development of renewable energy technologies may influence the issuance of tech-based credits, aligning them with clean energy ETFs.

Each selected asset is thus linked to voluntary carbon credits through a specific hypothesis: soybean futures are expected to correlate with nature-based credits due to their influence on land-use economics; the NASDAQ index and a clean energy ETF are associated with tech-based credits through their shared exposure to developments in clean technology; the EU ETS and natural gas are included based on their well-documented interdependence and their broader signaling effects on carbon markets; and the VIX is used to capture fluctuations in market risk sentiment that may influence corpo-

rate demand for offsets.

Recall that $p_{j,t}$ denotes the price of asset j at time t, where $j = 1, \ldots, N_{\text{asset}}$, and $t = T_j, \ldots, T_{\text{final}}$. Here, N_{asset} represents the number of assets and T_j is the first date for which we have an observation of asset j. Hence, $T_{\text{final}} - T_j + 1$ is the number of observations for asset j. To ensure comparability and stationarity in the time series analysis, we consider logarithmic returns, defined as

$$r_{j,t} = \log\left(\frac{p_{j,t}}{p_{j,t-1}}\right) \approx \frac{p_{j,t} - p_{j,t-1}}{p_{j,t-1}},$$

which are commonly used in financial applications due to their desirable mathematical properties [24].

4.2 Multivariate model setup and estimation

In an univariate case, each asset $j \in \{1, \ldots, N_{\text{assets}}\}$ is associated with a time index set $\mathcal{T}_j := \{T_j, \ldots, T_{\text{final}}\}$, where T_j denotes the first time at which returns for asset j are observed, and T_{final} is the common final time across all assets. To define a common multivariate time index set, we consider only time steps for which all assets have observed returns. Let

$$\mathcal{T}^{\text{multivariate}} := \{T_{\text{start}}, \ldots, T_{\text{final}}\}, \tag{4.1}$$

where $T_{\text{start}} := \max_j T_j$ is the latest starting point among all assets. For notational convenience and without loss of generality, we define $N := N_{\text{assets}}$ and $T_{\text{start}} = 0$. We now define the multivariate zero-inflated GARCH(1,1)-X model as follows.

Definition 4.1 (Multivariate zero-inflated GARCH(1,1)-X model)**.** Let N denote the number of assets and let $\mathcal{T}^{\text{multivariate}} = \{0, \ldots, T^{\text{final}}\}$ denote the common time index set over which all assets are observed. A multivariate zero-inflated GARCH(1,1)-X model for N assets is defined by

$$\begin{cases} \mathbf{r}_t = \mathbf{I}_t \odot \boldsymbol{\sigma}_t(\boldsymbol{\theta}^\star) \odot \boldsymbol{\eta}_t, \\ \sigma^2_{j,t}(\boldsymbol{\theta}^\star_j) = \alpha^\star_{j,0} + \alpha^\star_{j,1}\, \sigma^2_{j,t^*} + \alpha^\star_{j,2}\, r^2_{j,t^*} + \alpha^\star_{j,3}\, I_{j,t-1}, \quad \text{for all } j \in \{1, \ldots, N\}, \end{cases} \tag{4.2}$$

for all $t \in \mathcal{T}^{\text{multivariate}}$, where $\odot$ denotes the Hadamard (element-wise) product. The components of the model are defined as follows:

- $\mathbf{r}_t = (r_{1,t}, \ldots, r_{N,t})^\top \in \mathbb{R}^N$ is the vector of asset returns,
- $\boldsymbol{\sigma}_t = (\sigma_{1,t}, \ldots, \sigma_{N,t})^\top \in \mathbb{R}^N_+$ is the vector of conditional standard deviations,
- $\mathbf{I}_t = (I_{1,t}, \ldots, I_{N,t})^\top \in \{0,1\}^N$ is the vector of binary trading indicators,
- $\boldsymbol{\eta}_t = (\eta_{1,t}, \ldots, \eta_{N,t})^\top \in \mathbb{R}^N$ is the vector of standardized innovations,

- $t^* := t^*(j,t) := \max\{s < t : I_{j,s} = 1\}$ is the most recent trading time before t at which asset j was traded,

- $\boldsymbol{\theta}^\star := (\boldsymbol{\theta}_1^{\star\top}, \ldots, \boldsymbol{\theta}_N^{\star\top})^\top \in \boldsymbol{\Theta}$ is the true (unknown) parameter vector, where each asset-specific parameter vector is given by $\boldsymbol{\theta}_j^\star := (\alpha_{j,0}^\star, \alpha_{j,1}^\star, \alpha_{j,2}^\star, \alpha_{j,3}^\star)^\top$ for $j \in \{1, \ldots, N\}$.

The model can also be expressed in matrix form over the full time horizon. Let $\mathbf{R} \in \mathbb{R}^{(T^{\text{final}}+1)\times N}$ be the matrix of returns, defined by $\mathbf{R} = \mathbf{I} \odot \boldsymbol{\Sigma}(\boldsymbol{\theta}^\star) \odot \mathbf{E}$. This can be explicitly written as

$$\mathbf{R} = \underbrace{\begin{bmatrix} I_{1,0} & \cdots & I_{N,0} \\ \vdots & \ddots & \vdots \\ I_{1,T^{\text{final}}} & \cdots & I_{N,T^{\text{final}}} \end{bmatrix}}_{\in \{0,1\}^{(T^{\text{final}}+1)\times N}} \odot \underbrace{\begin{bmatrix} \sigma_{1,0} & \cdots & \sigma_{N,0} \\ \vdots & \ddots & \vdots \\ \sigma_{1,T^{\text{final}}} & \cdots & \sigma_{N,T^{\text{final}}} \end{bmatrix}}_{\in (0,+\infty)^{(T^{\text{final}}+1)\times N}} \odot \underbrace{\begin{bmatrix} \eta_{1,0} & \cdots & \eta_{N,0} \\ \vdots & \ddots & \vdots \\ \eta_{1,T^{\text{final}}} & \cdots & \eta_{N,T^{\text{final}}} \end{bmatrix}}_{\in (\mathbb{R}\sqcup\{\text{undefined}\})^{(T^{\text{final}}+1)\times N}} \tag{4.3}$$

This expression should be interpreted element-wise. That is, for all $j \in \{1, \ldots, N\}$ and $t \in \mathcal{T}^{\text{multivariate}}$,

$$r_{j,t} = I_{j,t} \cdot \sigma_{j,t} \cdot \eta_{j,t},$$

with the convention $0 \times \text{undefined} = 0$.

As in a univariate setting, we see that if asset j is not traded at time t the corresponding return $r_{j,t}$ is zero. Otherwise, the return equals the product of volatility and innovation, as in the standard GARCH framework. Each element of the conditional variance vector $\boldsymbol{\sigma}_t^2 = (\sigma_{1,t}^2, \ldots, \sigma_{N,t}^2)^\top$ evolves according to a component-wise zero-inflated GARCH(1,1)-X specification:

$$\sigma_{j,t}^2 = \alpha_{j,0} + \alpha_{j,1}\,\sigma_{j,t^*}^2 + \alpha_{j,2}\,r_{j,t^*}^2 + \alpha_{j,3}\,I_{j,t-1}, \quad \forall j \in \{1, \ldots, N\},$$

Thus, volatility updates depend only on non-zero return observations. Compared to a classical GARCH specification, we make two adjustments in Equation 4.2 to ensure that volatility dynamics evolve only when relevant information becomes available, which is essential in illiquid markets where trading is infrequent but highly informative. First, we include the binary trading indicator $I_{j,t-1}$ directly in the variance equation as a GARCH-X component. Second, we account for the fact that trading does not occur at every time step. We therefore define $t^* := \max\{s < t : I_{j,s} = 1\}$, which is the most recent time before t at which asset j was traded.

Parameter estimation is carried out using quasi-maximum likelihood (QMLE), which is the standard method in the GARCH literature [24]. Chapter 6 provides a detailed description of the univariate version of the zero-inflated GARCH-X model, how QMLE is applied, and its asymptotic properties. For further details, the reader is referred to that chapter.

A key feature of the zero-inflated GARCH-X model from Definition 4.2 is its multivariate extension. The multivariate nature of the model is defined by the cross-asset dependencies in both $\mathbf{I}_t$ and $\mathbf{E}_t$. To fully explain this extension, Chapter 5 focuses on the trading indicator functions $\mathbf{I}_t$ and their contemporaneous cross-dependence in trading activity. The dependencies among trading indicators are modeled using Markov Networks, as introduced in Section 2.7. In addition, Chapter 7

is dedicated to cross-dependence among assets based on their return innovations. Here, a copula-GARCH framework is employed, in which the dependencies among the residuals $\boldsymbol{\eta}_t$ are quantified using (conditional) Kendall's tau, estimated via kernel-based methods as described in Sections 2.4 and 2.5.

4.2.1 Extension to general models

Up to this point, we have considered only the multivariate zero-inflated GARCH(1,1)-X specification. However, the model can be naturally extended to a general multivariate zero-inflated GARCH(p, q, r)-X framework. This leads to the following definition:

Definition 4.2 (Multivariate zero-inflated GARCH(p, q, r)-X model)**.** Let N denote the number of assets and let $\mathcal{T}^{\text{multivariate}} = \{0, \dots, T^{\text{final}}\}$ denote the common time index set over which all assets are observed. Let $p, q, r \in \mathbb{N}$ be the lag order of the model. A multivariate zero-inflated GARCH(p, q, r)-X model for N assets is defined by

$$\begin{cases} \mathbf{r}_t = \mathbf{I}_t \odot \boldsymbol{\sigma}_t(\boldsymbol{\theta}^\star) \odot \boldsymbol{\eta}_t, \\ \sigma^2_{j,t}(\boldsymbol{\theta}^\star_j) = \alpha^\star_{j,0} + \displaystyle\sum_{\ell=1}^{p} \alpha^\star_{j,1,\ell}\, \sigma^2_{j,t^{*(\ell)}} + \sum_{m=1}^{q} \alpha^\star_{j,2,m}\, r^2_{j,t^{*(m)}} + \sum_{n=1}^{r} \alpha^\star_{j,3,n}\, I_{j,t-n}, \quad \text{for all } j \in \{1, \dots, N\}, \end{cases} \tag{4.4}$$

for all $t \in \mathcal{T}^{\text{multivariate}}$, where $\odot$ denotes the Hadamard (element-wise) product. The components of the model are defined as follows:

- $\mathbf{r}_t = (r_{1,t}, \dots, r_{N,t})^\top \in \mathbb{R}^N$ is the vector of asset returns,
- $\boldsymbol{\sigma}_t = (\sigma_{1,t}, \dots, \sigma_{N,t})^\top \in \mathbb{R}^N_+$ is the vector of conditional standard deviations,
- $\mathbf{I}_t = (I_{1,t}, \dots, I_{N,t})^\top \in \{0,1\}^N$ is the vector of binary trading indicators,
- $\boldsymbol{\eta}_t = (\eta_{1,t}, \dots, \eta_{N,t})^\top \in \mathbb{R}^N$ is the vector of standardized innovations,
- $t^{*(k)} := t^{*(k)}(j,t) := \max\{s < t^{*(k-1)} : I_{j,s} = 1\}$, with the convention $t^{*(0)} := t$; this defines the k-th most recent trading time before t for asset j, i.e. $t^{*(1)}$ is the last trading time before t, $t^{*(2)}$ the second last, and so on,
- $\boldsymbol{\theta}^\star := (\boldsymbol{\theta}^{\star\top}_1, \dots, \boldsymbol{\theta}^{\star\top}_N)^\top \in \boldsymbol{\Theta}$ is the true (unknown) parameter vector, where each asset-specific parameter vector is given by

$$\boldsymbol{\theta}^\star_j := \left(\alpha^\star_{j,0}, \alpha^\star_{j,1,1}, \dots, \alpha^\star_{j,1,p}, \alpha^\star_{j,2,1}, \dots, \alpha^\star_{j,2,q}, \alpha^\star_{j,3,1}, \dots, \alpha^\star_{j,3,r}\right)^\top \in \mathbb{R}^{1+p+q+r} \text{ for } j \in \{1, \dots, N\}.$$

The model can also be expressed in matrix form over the full time horizon similar as in Definition 4.2. Let $\mathbf{R} \in \mathbb{R}^{(T^{\text{final}}+1)\times N}$ be the matrix of returns, defined by $\mathbf{R} = \mathbf{I} \odot \boldsymbol{\Sigma}_{p,q,r}(\boldsymbol{\theta}^\star) \odot \mathbf{E}$. The only difference from Equation (4.3) is that $\boldsymbol{\Sigma}_{p,q,r}(\boldsymbol{\theta}^\star)$, while having the same dimensions as $\boldsymbol{\Sigma}(\boldsymbol{\theta}^\star)$, now specifies the conditional volatilities according to the GARCH(p, q, r)-X formulation in Equation (4.4).

Clearly, the setting $p = q = r = 1$ gives the multivariate zero-inflated GARCH(1,1)-X model introduced in Definition 4.2. It is important to note that estimation is still performed asset-wise using QMLE, and the asymptotic properties remain conceptually equivalent to those of the univariate zero-inflated GARCH(1,1)-X model. For notational simplicity, we therefore restrict our detailed analysis, estimation procedure, and asymptotic theory to the univariate case, as described in Chapter 6.

4.3 Structure of asset dependence

The conceptual framework for temporal and cross-asset dependence is illustrated in Figure 4.1. We consider by means of an example the returns of three representative assets r_j, r_k, and r_l observed over three temporal lags $t-1$, t, and $t+1$. The graph captures three forms of dependencies:

1. **Temporal dependence** (blue edges): dependency of a single asset across different time lags.
2. **Cross-dependence** (orange edges): contemporaneous dependence between distinct assets.
3. **Lagged cross-dependence** (gray edges): dependencies involving different assets across different time lags.

In this thesis, we only take into account temporal and contemporaneous cross-dependence, i.e. the blue and orange connections in Figure 4.1. The gray edges are excluded based on two practical modeling assumptions.

Assumption 4.1 (Modeling simplifications for dependency analysis)**.** *In this research, we assume the following two statements:*

- ***Single-lag memory***
  Trading behavior at time t is assumed to depend only on behavior at time $t-1$. Dependencies on earlier time points (e.g., $t-2$, $t-3$, etc.) are ignored.

- ***Contemporaneous cross-asset influence only***
  Cross-asset effects are considered only within the same time step. That is, influences such as $r_{j,t-1} \leftrightarrow r_{k,t}$ are excluded.

These assumptions are imposed to avoid the curse of dimensionality and to prevent reliance on increasingly sparse data, as accounting for all lagged inter-asset dependencies would require evaluating $2^{N_{lags} \times N_{assets}}$ possible configurations. More information about this can be found in Subsection 5.4.1.

Consequently, the analysis is limited to bivariate contemporaneous relationships between assets at the same time point. This simplification is both computationally efficient and statistically appropriate, given the illiquidity of the voluntary carbon credit market; the data is too limited to support reliable estimation of dependencies across multiple time lags. Including additional lags or modeling lagged cross-dependence would not yield meaningful improvements in the results. The cross-dependence in trading activity is further examined in Chapter 5, while the cross-dependence among residuals is discussed in Chapter 7.

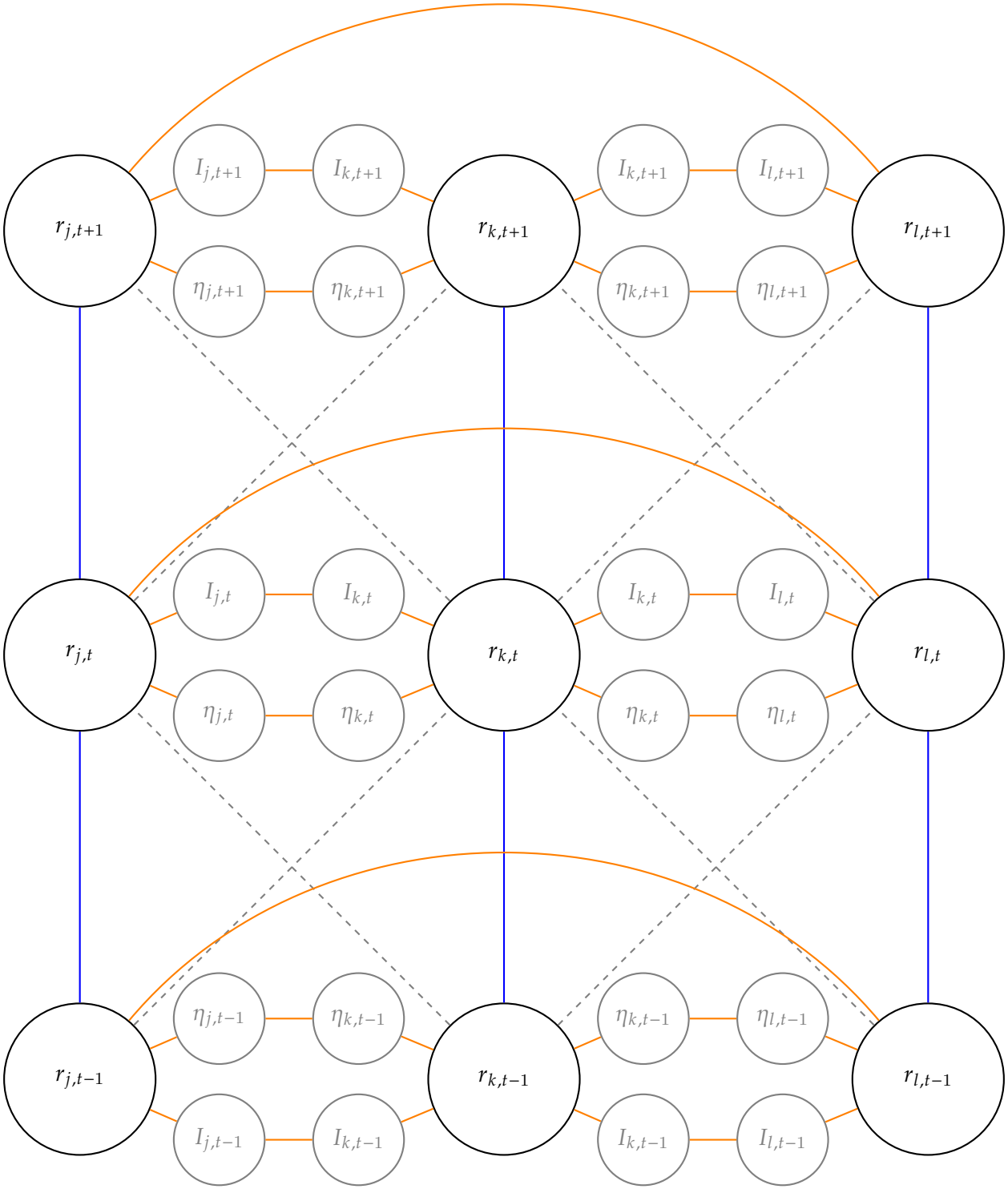


Figure 4.1: *Temporal and cross-asset dependence structure of returns, residuals $\eta_{j,t}$, and trading indicators $I_{j,t}$ for assets j, k, l over time lags $t+1, t, t-1$*

4.4 Descriptive analysis financial assets

4.4.1 Descriptive analysis Carbon Credit assets

We begin by analyzing the spot prices and corresponding daily returns for the selected voluntary carbon credit assets, organized by project category. The visualizations in Figures 4.2–4.5 present the price levels and log returns across four representative groups: Avoidance Nature, Avoidance Technology, Hybrid Nature, and Removal Nature.

Across all groups, the price data show a structural trend of declining valuations in voluntary carbon credits. Return behavior varies across assets; some assets show extreme episodic movements and others exhibit lower-magnitude but persistent fluctuations. For all assets, the returns seem to be constant around zero, indicating stationarity. This is classical in financial time series, where price series are typically non-stationary, while log returns are stationary. Moreover, in contrast to prices, returns do not depend on monetary units, which facilitates comparisons between assets [24]. It becomes also clear that the returns show volatility clustering. Interestingly, all assets exhibit more active market behavior with fewer zero returns in the earlier part of the sample period, particularly prior to 2023. This initial phase of greater price variability is followed by an extended period of price constancy, resulting in an increase in zero returns across assets. This pattern is somewhat counterintuitive, as one would typically expect market maturation over time to lead to increased liquidity and reduced incidence of zero returns.

Let us now look into each carbon asset group specifically. Figure 4.2 shows the price and return dynamics of the two assets in the Avoidance Nature group: the Nature-Based Offset (CCB-approved) and the VCS REDD+ CCB Gold Vintage 2021. Both assets have a general upward price trend until early 2022, followed by a sustained decline. This trajectory reflects a broader market revaluation, likely influenced by changing market sentiment toward nature-based credits and increased credit issuance, as discussed in Chapter 3. Return series in this group show high volatility, particularly for the Nature-Based Offset. Several extreme daily returns, exceeding ± 1.5 USD per credit, suggest the presence of illiquidity and the impact of isolated trades in a thin market.

Figure 4.3 presents the price and return dynamics of the two assets in the Avoidance Technology group: the GS Clean Cookstoves Africa V2021 and the GS Water Filtration Africa V2021. Both assets exhibit a gradual devaluation over time, but unlike the nature-based credits, the price trajectories consist of a series of downward steps. This pattern likely reflects sporadic trading and delayed price discovery in an illiquid market. The return series are relatively bounded in magnitude but show periods of clustered volatility, particularly throughout 2023 and 2024. Furthermore, the frequency of zero returns is higher than in other groups. This indicates lower trading activity and a higher degree of illiquidity.

Figure 4.4 shows the price and return dynamics for the Hybrid Nature group, represented by the ACR US IFM 2020 carbon credit. The price trajectory shows irregularities in the early trading period, with sudden jumps and sharp corrections. From mid-2023 onwards, however, the asset enters a phase of near-constant pricing, with a few observable spikes in price adjustments. Notably, a sharp price drop of nearly 3.5 USD per offset occurs in early 2024. Such a singular and abrupt adjustment is not observed in any of the other selected carbon credit assets. Among all assets analyzed, this group exhibits the highest concentration of zero returns, indicative of minimal recent trading activity and

substantial illiquidity. The transition from early price volatility to prolonged stability likely reflects a structural decline in market activity or a reduction in demand for this credit type.

Figure 4.5 presents the price and return dynamics of the Removal Nature group, consisting of the VCS CCB/GS Latam ARR 2021 and the VCS Uruguay ARR Vintage 2021 credits. Over the full sample period, both assets exhibit a general downward trend. In early to mid-2024, however, both credits experience a notable upward price adjustment, with the Latam ARR credit showing a sudden increase of nearly 3 USD per offset. Despite originating from distinct projects, the two assets display highly similar trajectories, indicating the influence of common market factors or a convergence in the valuation of ARR project types. Following these price adjustments, both series enter a phase of relative stability extending through 2025. This is reflected in the return series, which show reduced volatility and a higher concentration of zero returns, consistent with lower trading activity.

4.4.2 Descriptive analysis conventional assets

In this section, we examine the characteristics of a selected group of conventional or liquid financial assets, as introduced in Section 4.1. Our analysis focuses on their price and return dynamics to gain insight into the time series properties of more liquid markets.

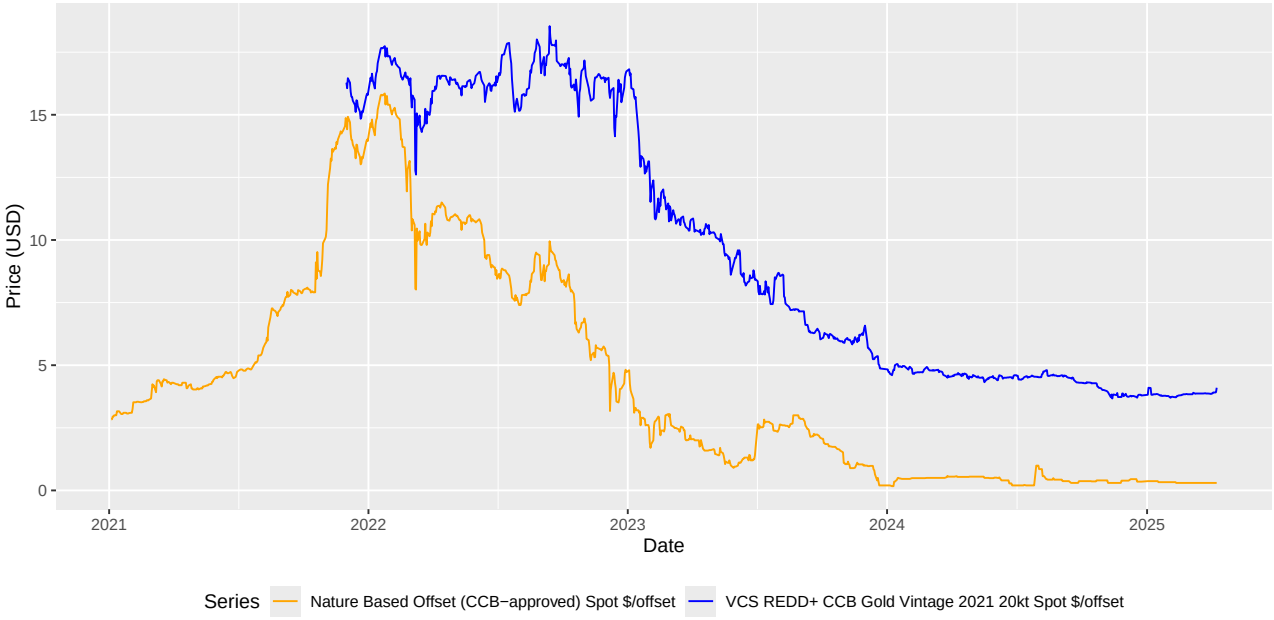
Figure 4.6 shows the price series of the conventional financial assets. For all assets except the EU ETS, the historical price series start in 2007, based on availability from Yahoo Finance. The EU ETS starts from mid-2020, as it was sourced from Quantum Commodity Intelligence. To ensure a consistent comparison across all assets, we align the time series by using a common starting point. Formally, we define the common starting point as $T_{\text{start}} = \max\{T_j : j \in \{1, \ldots, N_{\text{asset}}\}\}$, where T_j denotes the first available date for asset j, and N_{asset} is the total number of assets considered. In other words, the common starting point is the latest date on which all selected assets have data available.

The NASDAQ index shows the highest overall price levels among the selected assets, with a generally upward trend over the full time horizon. A notable drawdown occurs between 2021 and 2023, followed by a gradual recovery. This decline reflects global inflation and tighter monetary policy. After the COVID-19 pandemic began in 2020, inflation escalated in mid-2021 due to supply chain disruptions, strong demand, and economic support measures. In early 2022, the war in Ukraine pushed up prices for, among other things, energy and food. Central banks raised interest rates sharply in response. Higher rates especially affected tech stocks, contributing to the decline in the NASDAQ.

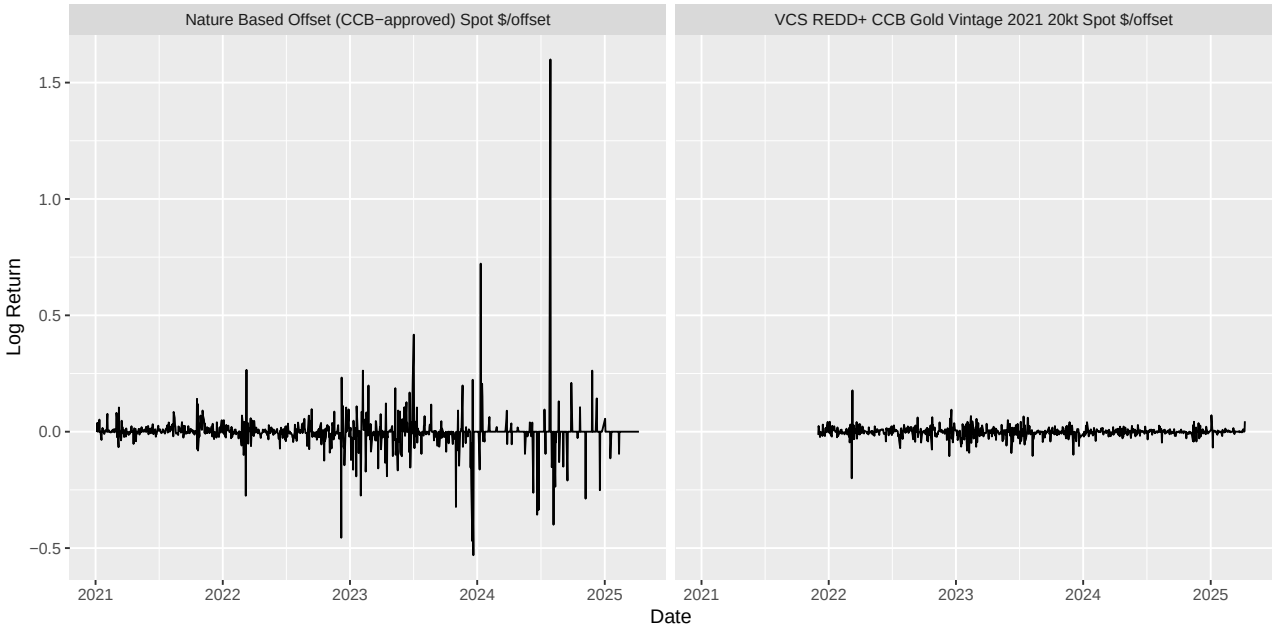
Turning to the EU ETS, the EU ETS price series begins in mid-2020 and shows a rapid increase followed by a more volatile regime. This coincides with the tightening of the EU's emissions trading rules under the European Green Deal and Fit-for-55 legislative packages, along with energy market disruptions following the onset of the war in Ukraine.

To analyze the statistical properties such as stationarity and volatility clustering of the conventional liquid financial assets, we examine their log return series, as presented in Figure 4.7. First, the return series for EU ETS shows moderate variability compared to more volatile assets such as Natural Gas and the VIX. A few sharp return spikes are visible, likely reflecting short-term market reactions to regulatory changes, COVID-19, the Ukraine war, and the energy crisis in late 2022. Returns fluctuate around zero without any visible trend, suggesting weak stationarity in the first moment. Volatility appears relatively stable over the sample period, with some visual evidence of volatility clustering.

Figure 4.2: *Price and log return plot for Group 1: Avoidance Nature Carbon Credits*

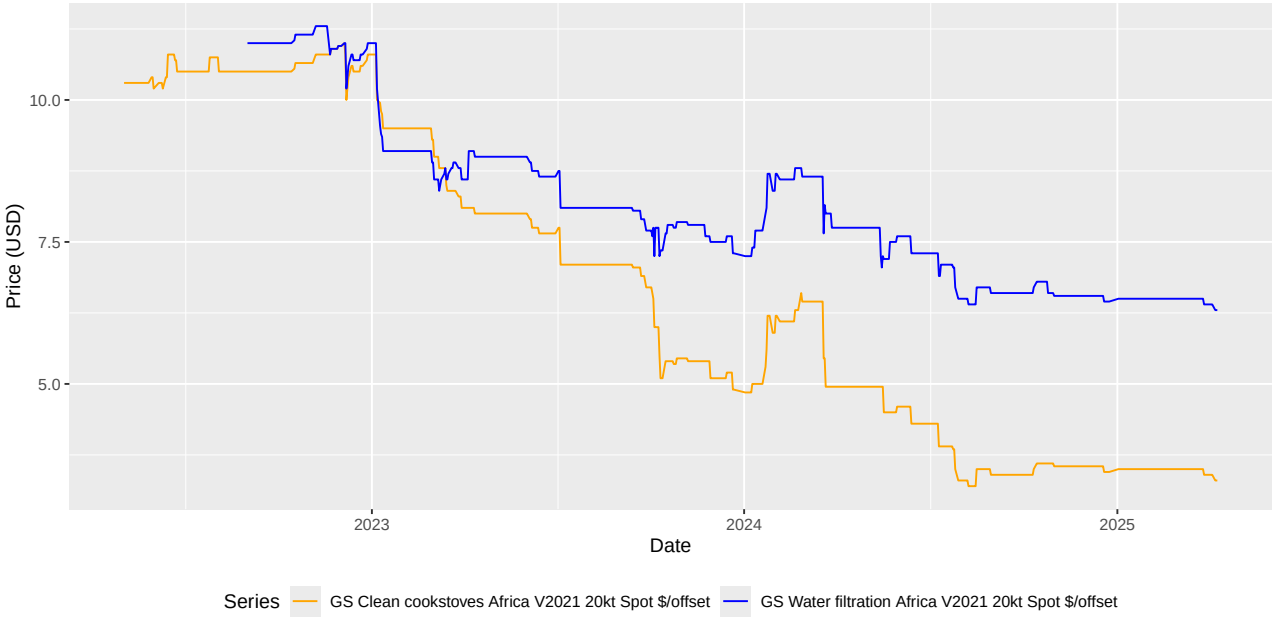


(**a**) *Plot of historical prices of two nature-based avoidance credits*

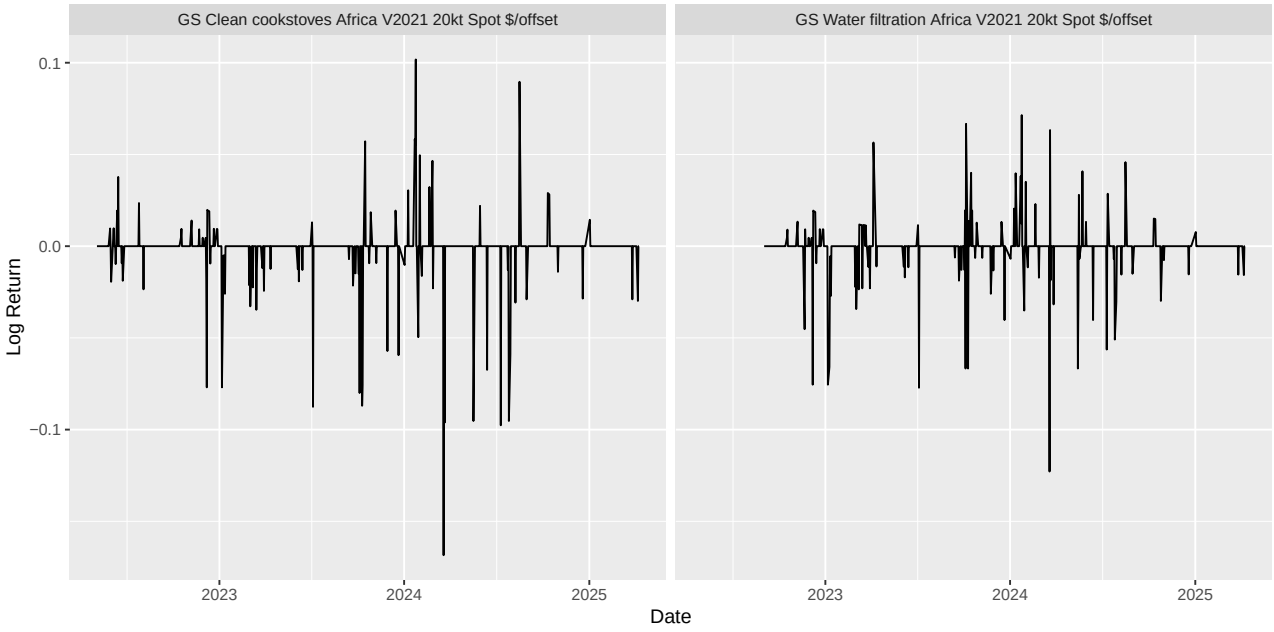


(**b**) *Plot of return values of two nature-based avoidance credits*

Figure 4.3: *Price and log return plot for Group 2: Avoidance Tech Carbon Credits*

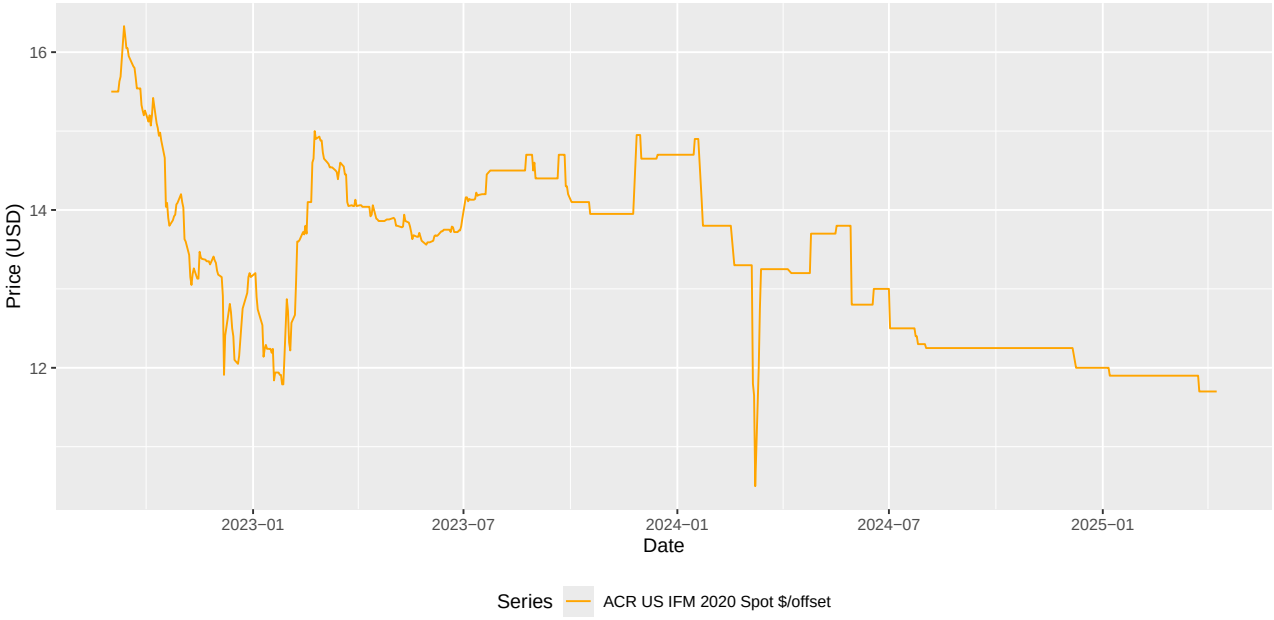


(**a**) *Plot of historical prices of two tech-based avoidance credits*

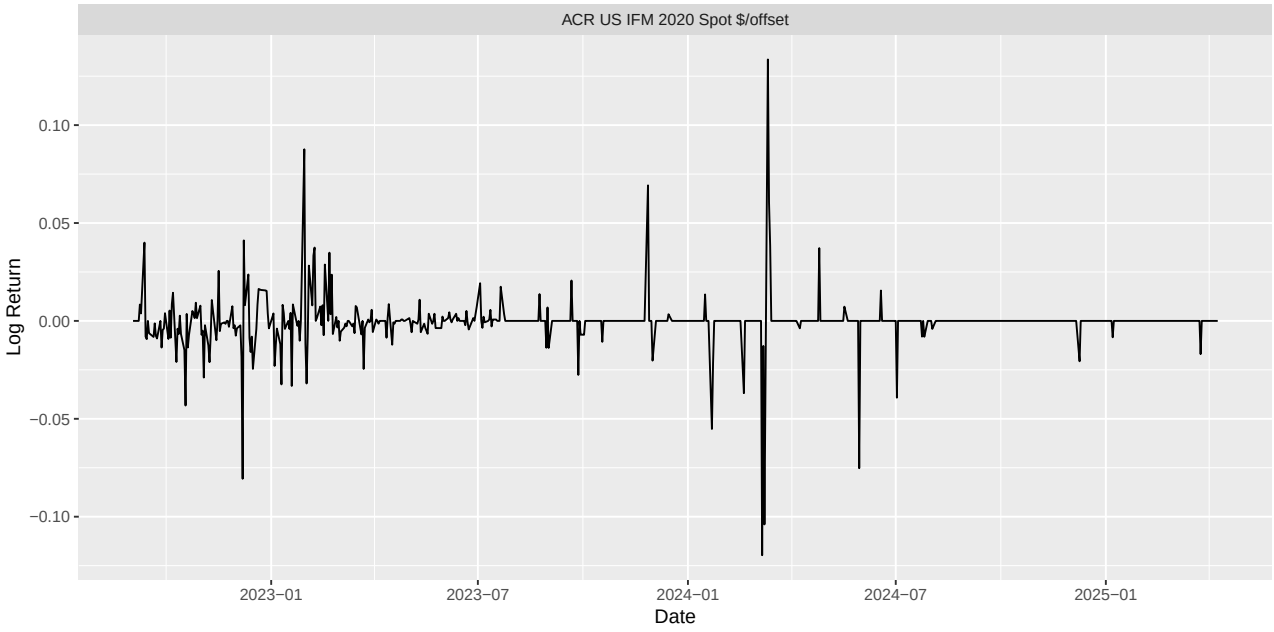


(**b**) *Plot of return values of two tech-based avoidance credits*

Figure 4.4: *Price and log return plot for Group 3: Hybrid Nature Carbon Credits*

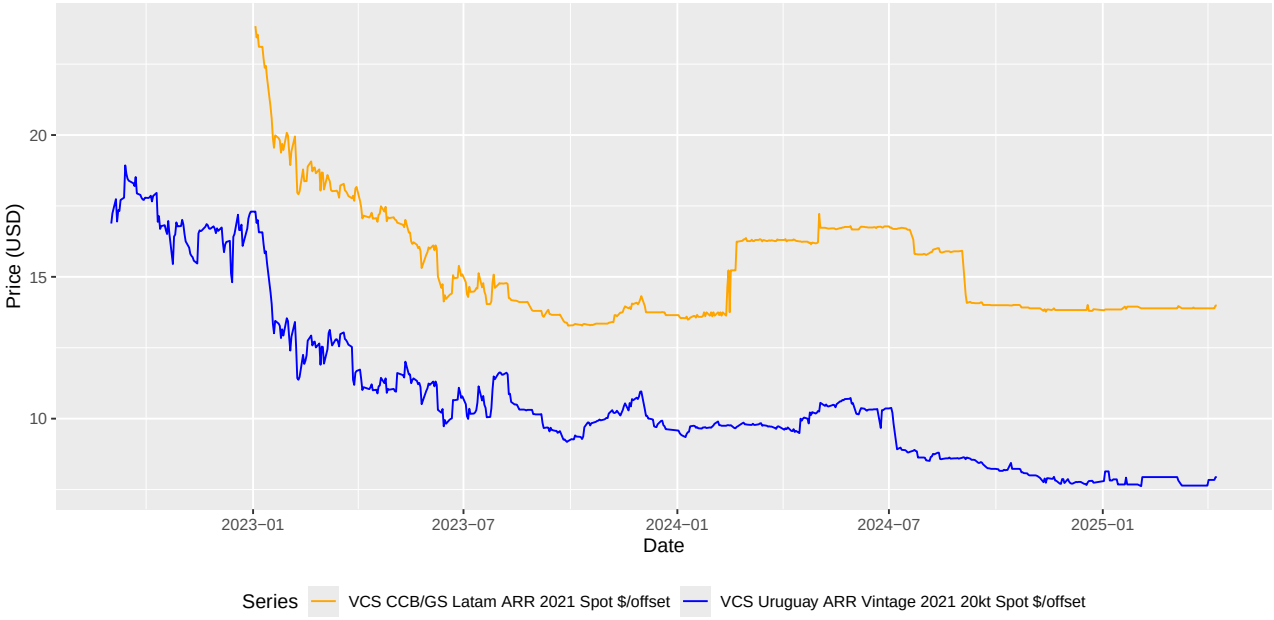


(**a**) *Plot of historical spot prices of ACR IFM Vintage 2020*

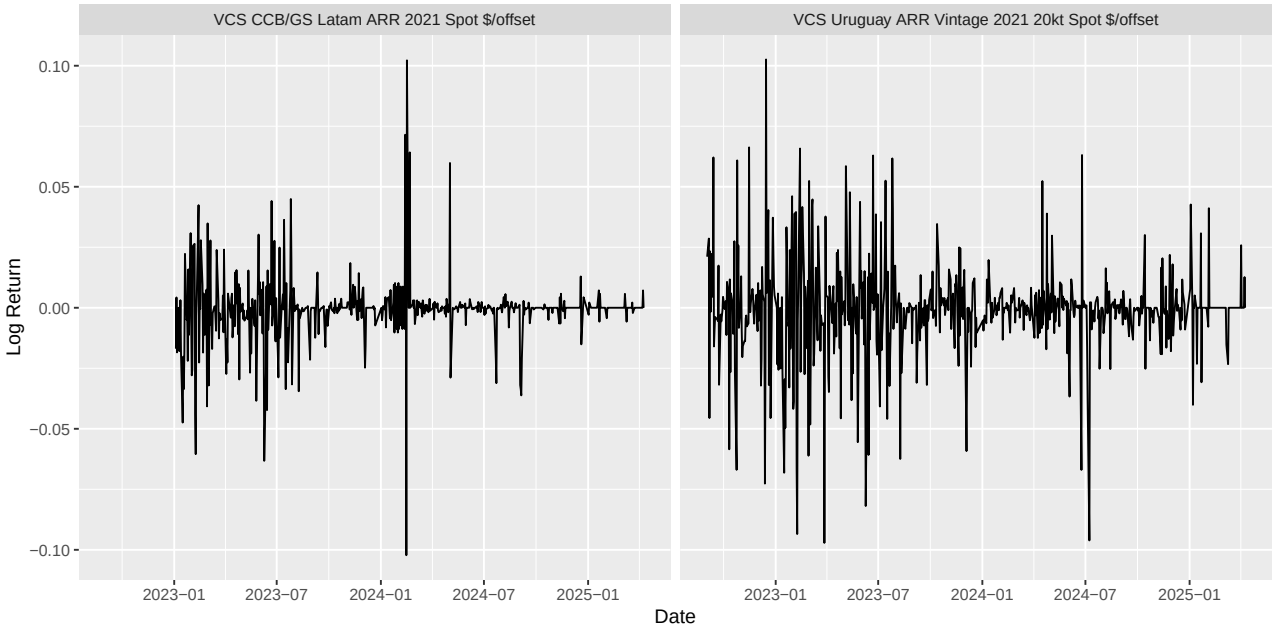


(**b**) *Plot of the return values of ACR IFM Vintage 2020*

Figure 4.5: *Price and log return plot for Group 4: Removal Nature Carbon Credits*



(**a**) *Plot of historical spot prices of two nature-based removal credits*



(**b**) *Plot of return values of two nature-based removal credits*

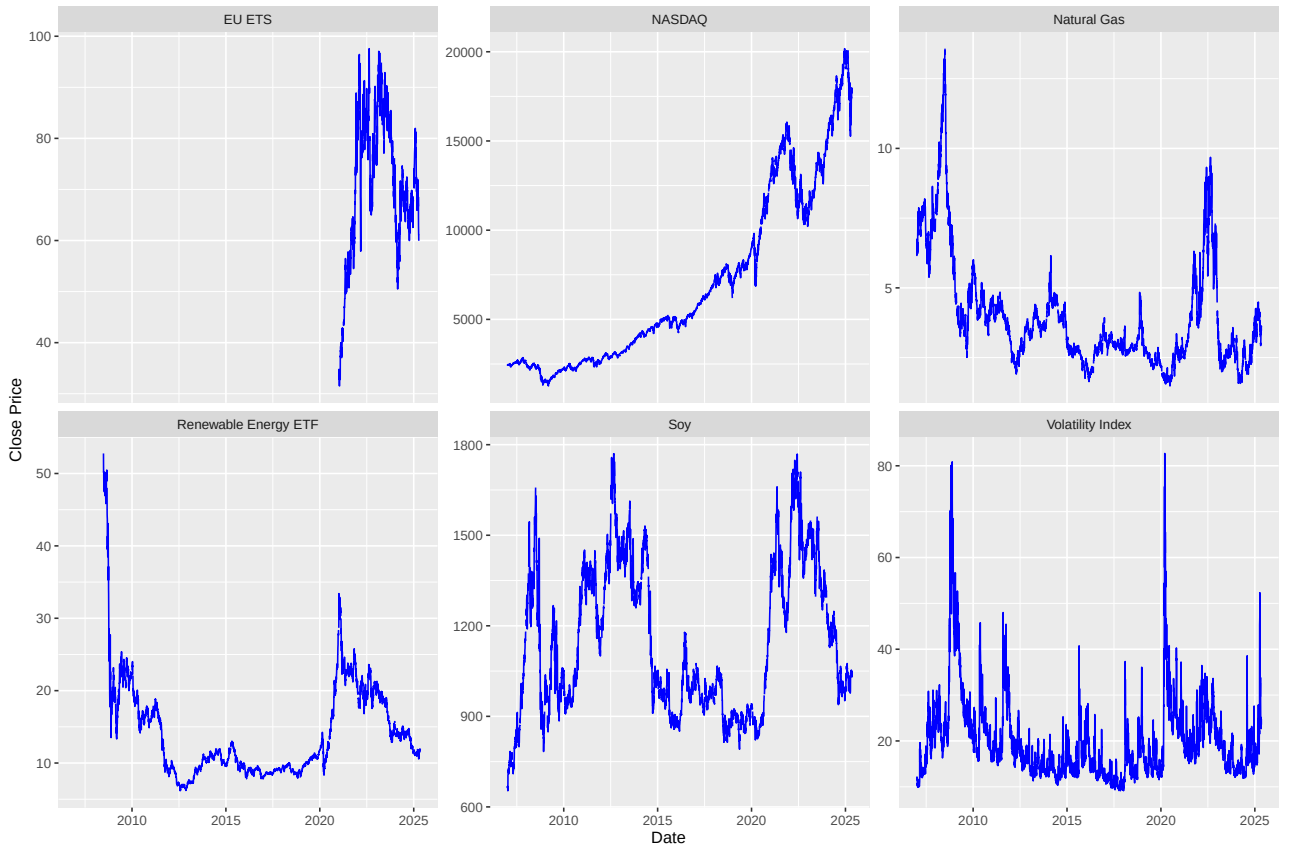


Figure 4.6: *Plots of the price series of the liquid financial assets. EU ETS is measured in EUR/tCO_2eq, and all other asset's prices are in USD.*

The NASDAQ return series shows moderate volatility relative to the other assets. Returns are tightly clustered around zero, with occasional spikes. These may be driven by macroeconomic events such as the financial crisis in 2008, the COVID-19 pandemic, and the Russia–Ukraine war. The series appears stationary in both mean and variance, and there is clear evidence of volatility clustering.

The return series for Natural Gas exhibits relatively high variability compared to the other assets. Several pronounced spikes are observed, particularly in 2022–2023, likely corresponding to geopolitical shocks such as the Russia–Ukraine war and the corresponding natural gas crisis. The series appears stationary in mean, and it seems that the variance changes over time, suggesting the presence of time-varying volatility. The occurrence of consecutive large return deviations indicates volatility clustering and supports the use of conditional heteroskedastic models.

The Renewable Energy ETF shows moderate return variability, higher than that of NASDAQ but lower than that of Natural Gas and the VIX. Spikes are observed particularly in the early part of the sample and during the 2020–2021 period. These may reflect investor sentiment driven by clean energy policies, the 2008 financial crisis, or COVID-19-related uncertainty. The series appears stationary in mean, and the temporal clustering of high return values suggests the presence of volatility clustering. Moreover, the volatility seems to be time-varying as well, which supports the use of conditional heteroskedastic models.

The Soy futures return series exhibits lower overall variability, with occasional extreme spikes, particularly in the earlier part of the sample. These may be linked to agricultural supply shocks, such as adverse weather conditions or trade policy changes, as well as during periods of financial stress, such

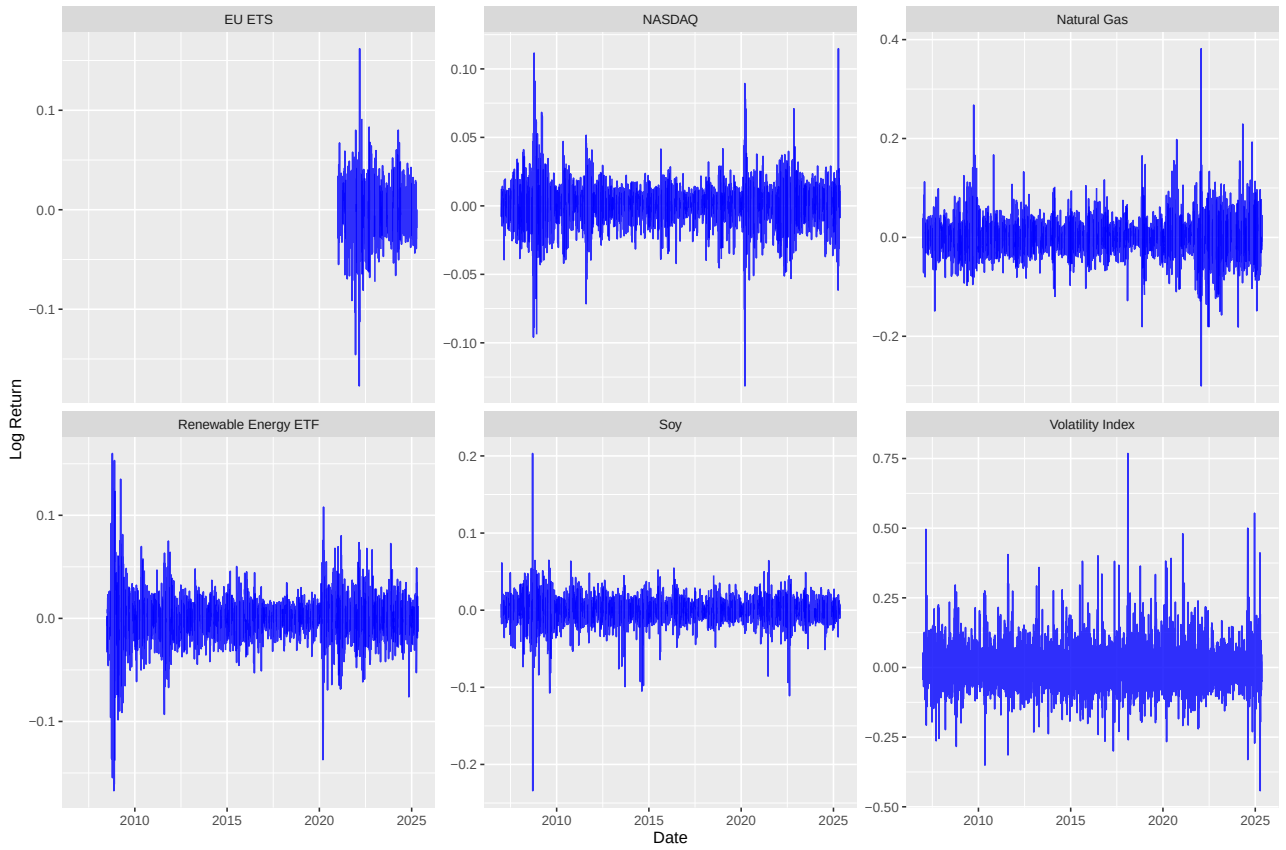


Figure 4.7: *Plot of the return values of the liquid financial assets*

as the 2008 crisis and the 2022–2023 economic slowdown. The series appears stationary. Volatility is less persistent than in the case of Natural Gas or the Renewable Energy ETF.

The VIX return series exhibits the most extreme fluctuations among the assets studied. These events correspond to periods of market stress, including the 2008 financial crisis and the COVID-19 shock in early 2020. The series fluctuates around zero, and high levels of volatility persist throughout the entire sample. Unlike other assets, the VIX does not show clear alternation between high- and low-volatility periods. Instead, volatility remains consistently elevated, indicating less pronounced volatility clustering and time-varying volatility. This behavior aligns with the nature of the VIX, which reflects implied market volatility and is inherently sensitive to ongoing uncertainty.

5

Illiquidity and multivariate analysis of trading activity

The presence of zero returns in the data creates challenges for statistical analysis, as it reflects a non-continuous time series. It is therefore essential to evaluate the degree of illiquidity among the selected voluntary carbon credit assets. In this chapter, we focus on the illiquidity features of these assets. As introduced in the previous chapter, we use a binary trading indicator $I_{j,t}$ to explicitly capture zero returns resulting from an absence of trading activity. We adopt a multivariate perspective on trading activity, distinguishing between temporal dependence (within each asset over time) and cross-sectional dependence (across different assets). We begin by formally explaining the concept of zero-inflated data and how it manifests in our setting. This is followed by a descriptive analysis of the binary indicators. To examine temporal dependence, we apply Fisher's exact test to assess whether past trading activity significantly influences current trading. We then analyze cross-asset dependence, starting with pairwise co-trading patterns and extending to a higher-dimensional perspective using UpSet plots to visualize overlapping trade activity across multiple assets. Finally, we introduce Markov Networks to formally model the multivariate dependence structure in trade/no-trade behavior.

5.1 Zero-returns analysis

Zero-inflation indicates that a dataset contains an excessive number of zeros. In finance, zero-inflated data typically manifests as periods of constant prices, resulting in sequences of zero returns. This is particularly common in illiquid markets, where irregular trading leads to extended periods of constant prices. In the literature, the fraction of zero-return days is widely used as a proxy for market illiquidity; a higher proportion of zero returns generally indicates lower trading activity [35, 44]. From a statistical perspective, these data are modeled as a mixture distribution: a discrete mass at zero and a continuous distribution over non-zero values. This violates the assumptions of standard continuous time series models and, therefore, we need alternative modeling approaches.

In our study of voluntary carbon credit assets, we observe similar zero-inflated return patterns, confirming the illiquid nature of carbon credit assets. To explicitly model the presence or absence of trading, we define a binary trading indicator, as introduced in Chapter 4. This allows us to analyze

illiquidity separately from the magnitude of returns. Recall that for each asset $j \in \{1, \ldots, N\}$ and time point $t \in \mathcal{T}_j$, the indicator is defined as:

$$I_{j,t} = \begin{cases} 1, & \text{if asset } j \text{ is traded at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

To make sure that zero returns are solely interpreted as the result of no trading activity, and not due to other market mechanisms, we introduce the following assumption:

Assumption 5.1 (Zero-inflated financial data)**.** *Let r_t denote the return of an asset at time t. Then,*

$$r_t = 0 \iff \textit{no trading activity occurred at time } t,$$
$$r_t \neq 0 \iff \textit{at least one trade occurred at time } t.$$

Under this assumption, a zero return directly indicates an absence of trading. This interpretation is natural, as it is highly unlikely that trades occur in such a way that their effects perfectly cancel out, resulting in an exact zero return. This understanding was further supported by expert discussions with professionals at Rabobank.

Since the frequency of zero returns serves as a proxy for illiquidity, we begin by examining the financial data. Table 5.1 highlights variation in trading activity across different project types. Projects such as Clean Cookstoves and Water Filtration show extremely high illiquidity, with over 85% zero returns, indicating that trading occurs only rarely. In contrast, credits like REDD+ exhibit a lower proportion of zero returns, suggesting comparatively higher liquidity and more frequent trading. Overall, it becomes clear that all carbon credit types in the dataset exhibit an excessive number of zero returns, ranging from approximately 12% to 85%. For comparison, liquid financial markets typically show less than 5% zero returns on daily data.

Table 5.1: *Number of zero and non-zero returns among the selected voluntary carbon credit types*

Asset (Spot $/offset)	**Non-zero returns**	**Total returns**	**Zero returns (%)**
Nature Based Offset Index	642	1074	40.2
REDD+ Vintage 2021	734	841	12.7
Clean cookstoves Africa Vintage 2021	89	738	87.9
Water filtration Africa Vintage 2021	97	654	85.2
US IFM Vintage 2020	215	654	67.1
Uruguay ARR Vintage 2021	533	654	18.5
Latam ARR Vintage 2021	363	569	36.2

Table 5.2 presents the number of zero and non-zero returns across these conventional assets. The conventional financial assets show far fewer zero-return observations than the carbon credit assets, which confirms that they are more liquid. The highest proportion of zero returns among them is seen in the Renewable Energy ETF (ICLN), at about 3%. These low percentages are typical for liquid markets, where prices tend to change daily due to regular trading activity. These occasional zero returns are not a sign of illiquidity but are most likely due to non-trading days such as weekends or public holidays.

Table 5.2: *Number of zero and non-zero returns for selected conventional liquid assets*

Asset (Ticker)	Non-zero returns	Total returns	Zero returns (%)
Soybean Futures (ZS.F)	4409	4455	1.03
Volatility Index (VIX)	4768	4786	0.38
Renewable Energy ETF (ICLN)	3963	4088	3.06
Natural Gas Futures (NG.F)	4432	4456	0.54
NASDAQ (IXIC)	4446	4447	0.02
EU ETS	1051	1052	0.10

5.2 Kurtosis

Kurtosis measures the degree of tail risk in a distribution. It indicates the likelihood of extreme outcomes relative to a normal distribution, which has a kurtosis of 3. Higher values suggest fat-tailed, leptokurtic behavior, a common feature in financial return series where extreme values occur more frequently than standard models would predict. Table 5.3 reports the kurtosis of each carbon asset's return distribution both with and without zero returns. Clearly, all assets exhibit high degrees of kurtosis. This suggests that extreme price movements are a persistent feature of return dynamics in the voluntary carbon market. Excluding zero returns reduces kurtosis among all assets. This result reinforces looking into the dependency of the trade/no-trade data in the subsequent sections.

Table 5.3: *Kurtosis of the returns of all voluntary carbon assets with and without zero returns*

Asset (Spot $/offset)	Kurtosis (All Returns)	Kurtosis (No Zeros)
Nature Based Index	154.40	92.66
REDD+	18.62	16.25
Cookstoves Africa	38.48	4.44
Water Filtration Africa	28.90	4.08
US IFM	44.89	14.82
Uruguay ARR	8.90	7.25
Latam ARR	19.80	12.72

For instance, the Nature Based Offset asset retains a high kurtosis of 92.66 even after filtering out zeros. This means an extreme deviation from normality and a high degree of tail risk. Similarly, assets like Cookstoves, IFM, and Water Filtration continue to exhibit strong leptokurtic behavior. Assets such as Water Filtration and Cookstoves, which have the highest share of zero returns (around 85%), show a notable drop in kurtosis after excluding those zeros. While still slightly elevated, their distributions become much closer to normal. This suggests a reduction in extreme outcomes once inactive days are removed. Overall, however, all assets remain clearly non-normally distributed, highlighting persistent tail risk in voluntary carbon markets and the need for models that account for non-Gaussian behavior.

In addition, we examine the kurtosis of the selected conventional assets, as shown in Table 5.4. All assets exhibit high kurtosis values, confirming the heavy tails typical of financial return distributions. Notably, these values are lower than most kurtosis values observed for the carbon credit assets when zero returns are included, as shown in Table 5.3. Only Uruguay ARR has a comparable kurtosis value

to those of the conventional assets. When zero returns are excluded from the carbon credit assets, the kurtosis values of most assets move closer to those of the conventional assets. However, the Nature-Based Index still exhibits an exceptionally high kurtosis of 92.66.

Table 5.4: *Kurtosis of daily returns for selected liquid assets*

Asset (Ticker)	**Kurtosis (All Returns)**
Soy (ZS.F)	23.93
Volatility Index (VIX)	9.98
Renewable Energy ETF (ICLN)	13.59
Natural Gas (NG.F)	10.39
NASDAQ Composite (IXIC)	10.95
EU ETS	7.97

5.3 Temporal dependence for the trading activity

Understanding how trading activity evolves over time assets is essential in the context of illiquid markets. Having established the presence of zero-inflated returns, we now examine the transitions between trade (T) and no-trade (NT) states to assess whether these shifts influence return behavior. Specifically, we compare return distributions conditional on trading activity—i.e., before and after trade versus no-trade days. This analysis helps determine whether trading activity is independently distributed or follows a more structured pattern within assets.

First, we examine the frequency of trades, no-trades, and particularly the transitions between them, as reported in the first four columns of Table 5.5. The distribution of trade and no-trade transitions is clearly not uniform, which indicates temporal dependence in trading activity. If trading and non-trading days were independent events, we would expect the transition counts to be more evenly spread across the four possible states. However, the observed data show clear asymmetries. For instance, in the Nature-Based Index, there are 503 consecutive trading day transitions (T →T), compared to only 139 transitions from a trade to a no-trade day (T→NT). This indicates that once trading occurs, it is likely to continue. A similar pattern is seen in the Latam ARR project, where T→T transitions (294) also exceed T→ NT transitions (69). This strengthens the hypothesis of temporal dependence in the trading activity.

Table 5.5: *Trade (T) / No-Trade (NT) transition counts for each carbon credit asset combined with Fisher Test results*

Asset (Spot $/offset)	**T → T**	**T → NT**	**NT → T**	**NT → NT**	**p-value**	**Odds Ratio**
Nature Based Index	503	139	139	292	< 2.2e-16	7.58
REDD+	658	75	75	32	$4.34e^{-7}$	3.74
Cookstoves Africa	23	66	66	582	$1.01e^{-4}$	3.07
Water Filtration Africa	30	67	67	489	$8.86e^{-6}$	3.26
US IFM	161	54	54	384	< 2.2e-16	21.06
Uruguay ARR	476	57	56	64	< 2.2e-16	9.49
Latam ARR	294	69	68	137	< 2.2e-16	8.54

To formally test the hypothesis of temporal dependence in trading activity among carbon credit assets, we apply Fisher's exact test which is introduced in Section 2.8. The resulting p-values and odds ratio are reported in the final columns of Table 5.5.

To apply the Fisher's Exact test, we constructed for each asset a binary time series reflecting the presence (1) or absence (0) of trading activity on a given day. From this, we derived 2×2 contingency tables that denote the frequency of daily state transitions, specifically: trade-to-trade, trade-to-no trade, no trade-to-trade, and no trade-to-no trade. The results in Table 5.5 reveal statistically significant dependence structures across all voluntary carbon credit assets. All assets reject the null hypothesis of independence at the 5% significance level. These findings strongly suggest that trading behavior is not memoryless, i.e. the probability of an asset trading on a given day is conditioned on whether it traded the previous day. Moreover, the odds ratios reflect the strength and direction of this dependence. All assets report odds ratios greater than one, indicating a positive temporal association. This implies that a trade on day t increases the likelihood of observing a trade on day $t+1$. The effect is particularly strong for assets such as US IFM Vintage 2020 (odds ratio around 21), Uruguay ARR (odds ratio around 9.5) and Latam ARR (odds ratio around 8.5), where past trading boosts the likelihood of continued trading activity.

This path dependence may be driven by market interest, coordinated trades, or data reporting practices, and has important implications for modeling volatility and return processes, particularly when applying conditional variance models like GARCH. To further investigate the implications of this dependence, we analyze the distribution of daily returns conditional on the previous day's trading state.

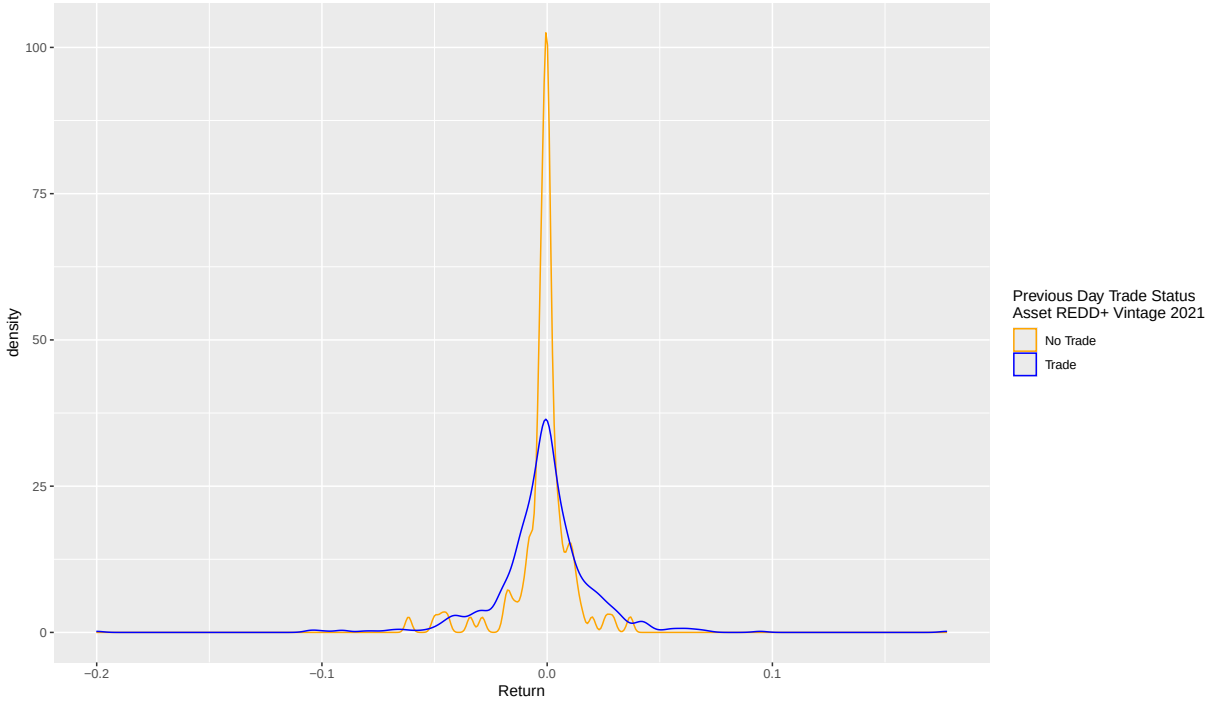


Figure 5.1: *Conditional return distribution by trading activity for a REDD+ Vintage 2021 carbon credit.*

Figure 5.1 and 5.2 show the conditional return distributions for the REDD+ and Cookstoves assets, respectively. These two assets are shown as representative examples, though similar patterns are observed across the other carbon credit types. As expected in financial applications, both distributions are centered around zero. However, the volatility differs depending on whether the previous day involved a trade or not.

Returns conditioned on a trade exhibit higher dispersion, while those following a no-trade day are

more tightly concentrated around zero. This observation supports the concept of volatility clustering: periods of trading are associated with increased return variability. Importantly, similar patterns are observed across all assets in the dataset, suggesting that volatility clustering persists even in illiquid markets such as the voluntary carbon market. On the other hand, it is important to note that the observed difference in dispersion may not capture the full dynamics of volatility. In particular, periods of latent volatility, where market uncertainty is present but not reflected in observed returns due to the absence of trading, can distort the interpretation of return-based volatility measures.

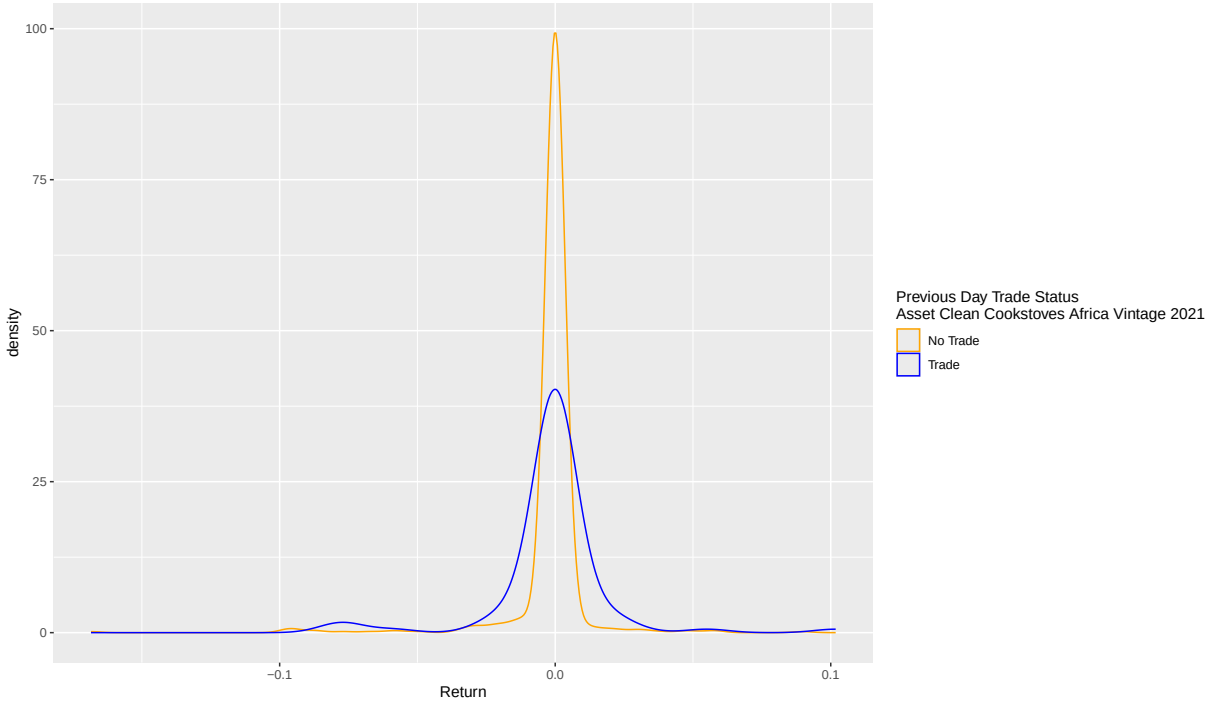


Figure 5.2: *Conditional return distribution by trading activity for a Cookstoves Africa Vintage 2021 carbon credit.*

5.4 Cross-asset dependence for the trading activity

Having discussed temporal dependence within individual assets, we now turn to the contemporaneous cross-dependence among the voluntary carbon credit assets. These relationships are shown by the orange edges in Figure 4.1. To statistically validate cross-dependence, we use Fisher's Exact Test between binary trade/no-trade indicators $I_{j,t}$ and $I_{k,t}$, for each asset pair (j, k). Table 5.6 presents the resulting p-values for the combinations of all carbon credit assets. Most asset pairs show strong statistical cross-dependence, with p-values below the 0.05 threshold. This indicates that the trading activity of one asset is statistically related to that of another; when one asset trades, the other is also more likely to trade. This suggests interlinked trading behavior across voluntary carbon credits.

There are, however, a few exceptions where independence cannot be rejected. These include: Cookstoves - REDD+, Cookstoves - Uruguay ARR, Cookstoves - Latam ARR, and Water Filtration - Latam ARR. For these combinations, p-values exceed the 0.05 level, indicated in bold in Table 5.6. One possible explanation is that they differ in project type, geographic focus, or buyer demand. Another possible reason is that Cookstoves Africa and Water Filtration are the most illiquid assets in the sample, with very few trading days, as shown in Table 5.1. The lack of statistical significance is therefore likely due to limited data, rather than the absence of actual cross-dependence. Thus, while most

assets exhibit interconnected trade behavior, a few operate differently.

Table 5.6: *Fisher Test p-values for independence between $I_{j,t}$ (trade/no-trade behavior) and $I_{k,t}$ (trade/no-trade behavior) for every pair (j,k) of voluntary carbon credit assets. Non-significant p-values are shown in bold.*

	Nature Index	REDD+	Cookstoves Africa	Water Filtration	US IFM	Uruguay ARR	Latam ARR
Nature Index	NA	1.16e-12	1.08e-04	5.55e-03	< 2.2e-16	< 2.2e-16	4.65e-10
REDD+	1.16e-12	NA	**7.56e-02**	**1.43e-02**	1.51e-08	< 2.2e-16	< 2.2e-16
Cookstoves Africa	1.08e-04	**7.56e-02**	NA	< 2.2e-16	5.18e-04	**9.00e-02**	**2.72e-01**
Water Filtration	5.55e-03	**1.43e-02**	< 2.2e-16	NA	6.79e-03	1.04e-02	1.07e-01
US IFM	< 2.2e-16	1.51e-08	5.18e-04	6.79e-03	NA	8.66e-15	8.11e-12
Uruguay ARR	< 2.2e-16	< 2.2e-16	**9.00e-02**	1.04e-02	8.66e-15	NA	< 2.2e-16
Latam ARR	4.65e-10	< 2.2e-16	**2.72e-01**	1.07e-01	8.11e-12	< 2.2e-16	NA

Table 5.7: *Odds ratios from Fisher's Test for independence of trade/no-trade behavior between $I_{j,t}$ and $I_{k,t}$ for each pair (j,k) of voluntary carbon credit assets.*

	Nature Index	REDD+	Cookstoves Africa	Water Filtration	US IFM	Uruguay ARR	Latam ARR
Nature Index	NA	4.84	2.60	2.08	22.69	8.63	3.29
REDD+	4.84	NA	2.13	2.75	5.41	14.05	11.20
Cookstoves Africa	2.60	2.13	NA	607.63	2.32	1.88	1.39
Water Filtration	2.08	2.75	607.63	NA	2.14	2.64	1.54
US IFM	22.69	5.41	2.32	2.14	NA	10.52	5.21
Uruguay ARR	8.63	14.05	1.88	2.64	10.52	NA	16.97
Latam ARR	3.29	11.20	1.39	1.54	5.21	16.97	NA

Table 5.7 presents the odds ratios from Fisher's Exact Test for each pair of voluntary carbon credit assets. Most asset pairs exhibit odds ratios well above one, indicating positive dependence in trade activity. The strongest associations are observed between US IFM and Nature Index (22.69), Uruguay ARR and Latam ARR (16.97), and Uruguay ARR and REDD+ (14.05), highlighting the tight connections among nature-based credits. In contrast, Cookstoves Africa and Water Filtration Africa show only weak associations with most other assets, with odds ratios closer to one. Yet, they are strongly associated with each other (607.63), likely reflecting a few rare instances of joint trading. The weaker association with the other carbon credits may be explained by their limited liquidity. Overall, the results confirm broad cross-asset trade activity dependence.

To explore contemporaneous cross-dependence between multiple voluntary carbon assets beyond bivariate relationships, we use an UpSet plot [15, 45]. This is implemented in `R` through the `UpSetR` package [15]. An UpSet plot is a visualization method designed to represent intersections of multiple sets in order to offer an alternative to Venn diagrams when analyzing more than three groups. In this context, each set corresponds to a specific carbon asset, and intersections indicate instances where those assets are simultaneously traded. The binary trade/no-trade data used in the plot are based on time series aligned to the first date on which the last asset began trading, i.e. the earliest point

from which all assets show trading activity. We restrict the analysis to this common starting point to ensure a fair comparison across assets. This results in a total of 569 observations.

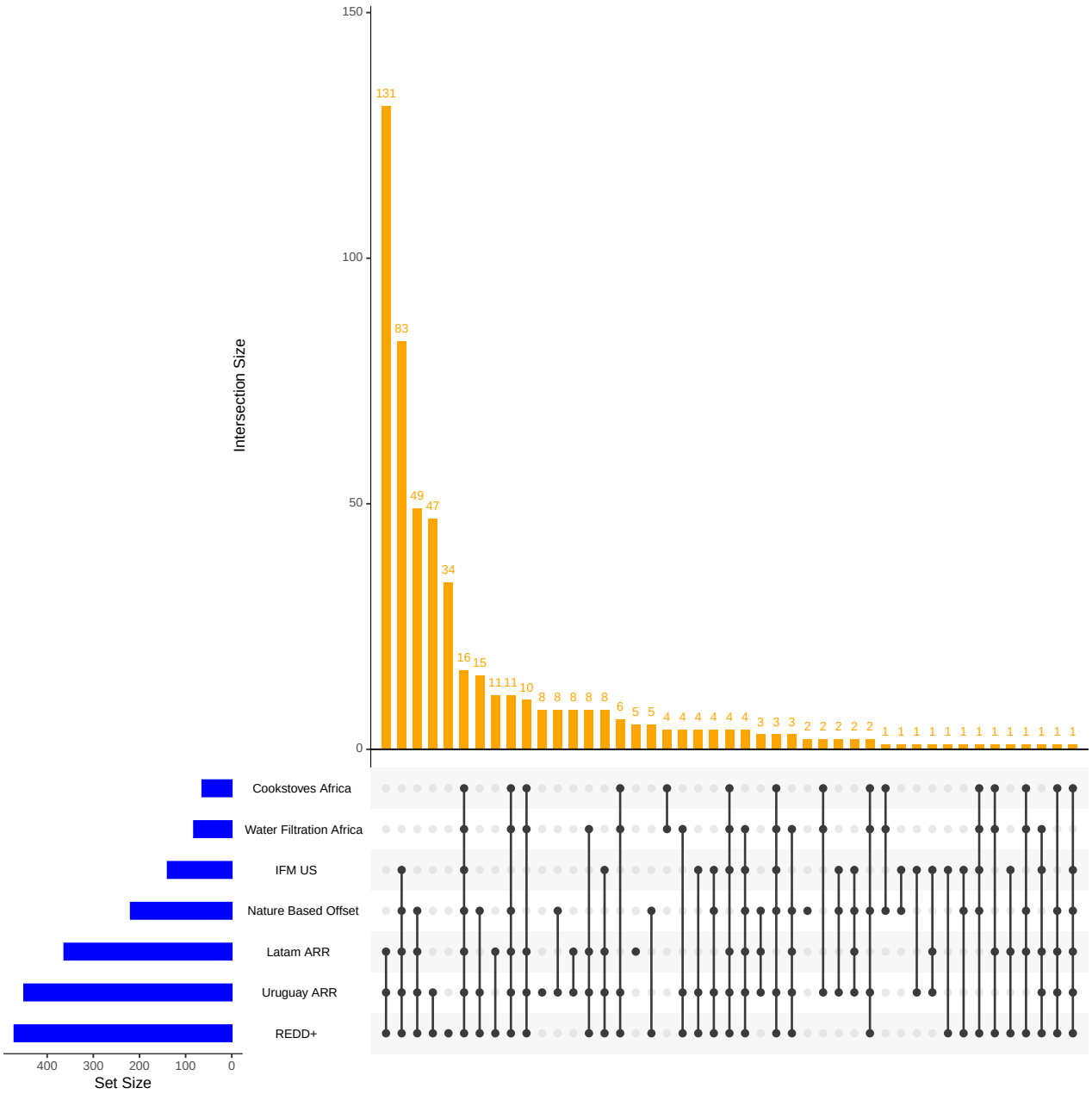


Figure 5.3: *UpSet plot of the trade/no-trade data for all voluntary carbon credits explaining joint trading activity*

The UpSet plot in Figure 5.3 presents the most frequent combinations of simultaneous trading among voluntary carbon assets. The most common pattern, observed on 131 occasions, involves the joint trading of three assets: REDD+, Uruguay ARR, and Latam ARR. The second most frequent combination, occurring 83 times, extends this set to include Nature-Based Offset and IFM US, while Cookstoves and Water filtration are still not traded. Notably, the top five intersections all include REDD+ and consist solely of these five nature-based carbon credit assets. This consistent co-occurrence suggests a degree of cross-dependence in trading behavior and points to a structural interconnectedness among nature-based credits. In contrast, other asset combinations appear very rarely, usually fewer than 20 times. This low frequency is still meaningful: it shows that joint trading between assets

outside the nature-based group is rare, which may indicate weaker links or less overlap in market activity among them.

The results of the UpSet plot align well with the findings from the Fisher tests. Nature-based carbon credits such as REDD+, Nature Index, Uruguay ARR, and Latam ARR frequently trade together. This is confirmed by the low p-values in the Fisher tests and high odds ratios between these pairs, such as REDD+–Uruguay ARR (14.05) and Nature Index–US IFM (22.69). These frequent co-occurrences are also visible in the top intersections of the UpSet plot, which are dominated by combinations of nature-based credits. In contrast, assets like Cookstoves Africa and Water Filtration Africa appear less frequently in joint trading combinations. This matches their low odds ratios and non-significant p-values with most other assets. Interestingly, although they show a very strong odds ratio with each other (607.63), they do not appear together often in the UpSet plot. This likely reflects their high illiquidity.

5.4.1 Cross-asset trading dependencies via Markov networks

Recall that for each asset j, we observe a binary trading indicator $I_{j,t} \in \{0,1\}$ over discrete time periods $t = T_j, \ldots, T_{\text{final}}$, where $I_{j,t} = 1$ denotes the presence of a trade and $I_{j,t} = 0$ indicates the absence of trading activity. Subsections 5.3 and 5.4 provide statistical evidence of significant co-trading behavior, both temporal dependence over time lags and across the assets at the same time. We focus on bivariate dependence only due to curse of dimensionality. To formally quantify the cross-dependence in trading activity, we consider a bivariate indicator vector $(I_{j,t}, I_{k,t}) \in \{0,1\}^2$ for a given pair of assets (j,k). At each time t, this pair can take one of four possible values. We define the population joint probabilities

$$
\begin{aligned}
p_{00}^{(j,k)} &= \mathbb{P}(I_{j,t} = 0, I_{k,t} = 0),\\
p_{10}^{(j,k)} &= \mathbb{P}(I_{j,t} = 1, I_{k,t} = 0),\\
p_{01}^{(j,k)} &= \mathbb{P}(I_{j,t} = 0, I_{k,t} = 1),\\
p_{11}^{(j,k)} &= \mathbb{P}(I_{j,t} = 1, I_{k,t} = 1),
\end{aligned}
$$

where $p_{00}^{(j,k)} + p_{10}^{(j,k)} + p_{01}^{(j,k)} + p_{11}^{(j,k)} = 1$. These are treated as parameters of the true but unknown data-generating process of the joint distribution of the binary indicators for assets j and k. We denote the full parameter vector as $\mathbf{p}^{(j,k)} := (p_{00}^{(j,k)}, p_{10}^{(j,k)}, p_{01}^{(j,k)}, p_{11}^{(j,k)})^\top \in \mathcal{P}^{(j,k)} \subseteq \mathbb{R}^4$, where $\mathcal{P}_{j,k}$ is the parameter space and the true parameter vector is denoted as $\mathbf{p}_{j,k}^\star = (p_{00}^{(j,k)\star}, p_{10}^{(j,k)\star}, p_{01}^{(j,k)\star}, p_{11}^{(j,k)\star})^\top \in \mathcal{P}^{(j,k)}$. This can also be denoted in a contingency table:

	$I_{k,t} = 0$	$I_{k,t} = 1$
$I_{j,t} = 0$	$p_{00}^{(j,k)}$	$p_{01}^{(j,k)}$
$I_{j,t} = 1$	$p_{10}^{(j,k)}$	$p_{11}^{(j,k)}$

Table 5.8: *Contingency table of joint trading activity between assets j and k*

To model cross-dependence among these indicators, we use Markov Networks, or Markov Random Fields. As described in Section 2.7, an MRF is defined by an undirected graph $G = (U, E)$, where each vertex $j \in U$ corresponds to a random variable. In our setting each vertex corresponds to $I_{j,t}$. Each undirected edge $(j,k) \in E$ represents a pairwise dependency between $I_{j,t}$ and $I_{k,t}$. The PC

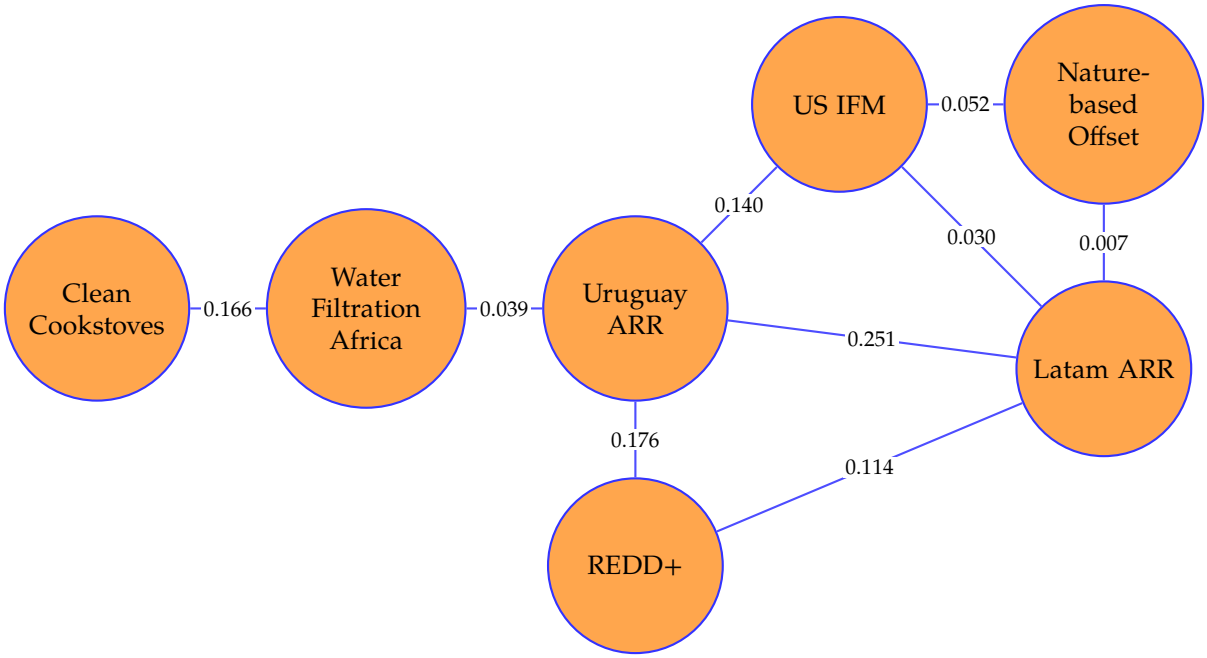


Figure 5.4: *Markov Network of binary trading indicators $I_{j,t}$ for all voluntary carbon credits, with estimated Mutual Information values provided on the edges*

algorithm, implemented in the function `pc.stable` of the `R` package `bnlearn`, is used to estimate G by testing conditional independence between variables. This is a non-parametric procedure that tests conditional independence between variables using contingency tables such as the Fisher's test as described in Section 2.8. In Figure 5.4, the estimated cross-dependence structure among the trading indicators for voluntary carbon credits are visualized. Overall, the figure suggests that nature-based credits are more interconnected with each other than with technology-based credits like Cookstoves and Water Filtration.

Once the graph structure is estimated, the strength of each edge is quantified by computing mutual information. For each connected pair (j, k), we construct the empirical joint distribution

$$\hat{p}_{ab}^{(j,k)} = \frac{1}{T} \sum_{t=T_j}^{T_{\text{final}}} \mathbf{1}\{I_{j,t} = a,\, I_{k,t} = b\}, \quad \text{for } a, b \in \{0, 1\},$$

which represents the relative frequency of joint occurrences of trading activity and non-activity over time. The corresponding empirical marginal distributions are given by

$$\hat{p}_{a\cdot}^{(j,k)} = \sum_{b=0}^{1} \hat{p}_{ab}^{(j,k)}, \qquad \hat{p}_{\cdot b}^{(j,k)} = \sum_{a=0}^{1} \hat{p}_{ab}^{(j,k)}. \tag{5.1}$$

Note that (5.1) are empirical estimates of the true marginal probabilities $\mathbb{P}(I_{j,t} = a)$ and $\mathbb{P}(I_{k,t} = b)$, respectively. The estimated mutual information is then defined as

$$\text{MI}(I_{j,t}; I_{k,t}) = \sum_{a=0}^{1} \sum_{b=0}^{1} \hat{p}_{ab}^{(j,k)} \log \left(\frac{\hat{p}_{ab}^{(j,k)}}{\hat{p}_{a\cdot}^{(j,k)} \hat{p}_{\cdot b}^{(j,k)}} \right).$$

Mutual information is always non-negative and equals zero if and only $I_{j,t}$ and $I_{k,t}$ are independent. In our setting, mutual information is a descriptive measure of marginal association strength between trading indicators and is used to assign interpretable weights to the edges in the estimated Markov network. The empirical joint and marginal distributions, as well as the mutual information values between each pair of trading indicators, are computed using the `infotheo` package in `R`.

Figure 5.4 shows estimated mutual information values on the edges, representing the strength of dependence in trading activity between pairs of binary trading indicators. Higher values indicate stronger co-movement. We observe relatively strong dependencies between Clean Cookstoves and Water Filtration Africa, REDD+ and Uruguay ARR, and Uruguay ARR and US IFM. The strongest connection appears between Uruguay ARR and Latam ARR, likely reflecting their similarity in both credit type and geographic region. In contrast, Nature-Based Offset shows relatively weak associations with its connected credits.

Once the Markov network structure is estimated and the mutual information is used to quantify marginal association strength between connected nodes, we turn to building a probabilistic model of the joint behavior. The key insight is that the global joint distribution of the trading indicators factorizes according to the structure of the graph: under the Markov property, each indicator variable $I_{j,t}$ is conditionally independent of all other nodes given its immediate neighbors $\mathcal{N}(j)$. This conditional independence structure allows us to express the high-dimensional joint distribution using only low-dimensional conditional distributions of the form

$$p(I_{j,t} \mid I_{\mathcal{N}(j),t}).$$

These local conditional distributions can be directly estimated from the binary trading indicator data. For each asset j, and for each observed configuration of its neighbors at time t, we compute the empirical frequency with which $I_{j,t} = 1$ occurs, conditional on that configuration. Formally, for a given neighbor state $\mathbf{i} \in \{0,1\}^{|\mathcal{N}(j)|}$, we compute

$$\hat{p}(I_{j,t} = 1 \mid I_{\mathcal{N}(j),t} = \mathbf{i}) = \frac{\sum_{t=T_j}^{T_{\text{final}}} \mathbf{1}\{I_{j,t} = 1,\ I_{\mathcal{N}(j),t} = \mathbf{i}\}}{\sum_{t=T_j}^{T_{\text{final}}} \mathbf{1}\{I_{\mathcal{N}(j),t} = \mathbf{i}\}}. \tag{5.2}$$

Example 2. *Suppose we consider three assets: asset j, and its two neighbors k and ℓ, with a binary trading indicator $I_{\alpha,t} \in \{0,1\}$ for each asset $\alpha \in \{j,k,\ell\}$ at time t. Assume we observe data over 10 time periods, and the neighborhood of asset j is given by $\mathcal{N}(j) = \{k,\ell\}$. That is, we want to estimate how likely asset j trades, given the trade/no-trade configuration of assets k and ℓ on the same day. We record the observed values of the indicators for these three assets in the following table:*

t	$I_{j,t}$	$I_{k,t}$	$I_{\ell,t}$
1	1	1	0
2	0	0	1
3	1	1	1
4	0	1	0
5	0	0	0
6	1	1	1
7	0	1	0
8	1	0	1
9	0	0	0
10	1	1	0

We now estimate the empirical conditional probability $\hat{p}(I_{j,t} = 1 \mid I_{\mathcal{N}(j),t} = \mathbf{i})$ *for each possible configuration* $\mathbf{i} = (I_{k,t}, I_{\ell,t}) \in \{0,1\}^2$*. There are four such configurations:*

- ***Configuration (1, 0):***
  This occurs at $t = 1, 4, 7, 10$*, with corresponding values* $I_{j,t} = 1, 0, 0, 1$*. Thus, two out of four times asset* j *trades. Hence,* $\hat{p}(I_{j,t} = 1 \mid I_{k,t} = 1, I_{\ell,t} = 0) = \frac{2}{4} = 0.5$*.*

- ***Configuration (0, 1):***
  This occurs at $t = 2, 8$*, with* $I_{j,t} = 0, 1$*. One out of two times asset* j *trades. Hence,* $\hat{p}(I_{j,t} = 1 \mid I_{k,t} = 0, I_{\ell,t} = 1) = \frac{1}{2} = 0.5$*.*

- ***Configuration (1, 1):***
  This occurs at $t = 3, 6$*, with* $I_{j,t} = 1, 1$*. In both cases asset* j *trades. Hence,* $\hat{p}(I_{j,t} = 1 \mid I_{k,t} = 1, I_{\ell,t} = 1) = \frac{2}{2} = 1.0$*.*

- ***Configuration (0, 0):***
  This occurs at $t = 5, 9$*, with* $I_{j,t} = 0, 0$*. Asset* j *never trades. Hence,* $\hat{p}(I_{j,t} = 1 \mid I_{k,t} = 0, I_{\ell,t} = 0) = \frac{0}{2} = 0.0$*.*

Putting these results together, we obtain the full local conditional distribution for $I_{j,t}$*, conditional on its neighbors' configuration:*

$(I_{k,t}, I_{\ell,t})$	$\hat{p}(I_{j,t} = 1 \mid I_{k,t}, I_{\ell,t})$
$(1, 0)$	0.5
$(0, 1)$	0.5
$(1, 1)$	1.0
$(0, 0)$	0.0

While mutual information tells us whether an edge exists and how strong the marginal dependency is, it does not specify a generative process. For this, we require the local conditional probabilities, which allow us to simulate new configurations of the full indicator vector $\{I_{j,t}\}_{j \in U}$ using Gibbs sampling. For Gibbs sampling, at each step, a node $I_{j,t}$ is updated by sampling from a Bernoulli distribution with success probability given by Equation 5.2. Here, S denotes the fixed number of simulated

observations generated using the Gibbs sampling algorithm. In this way, the estimated local distributions fully determine the multivariate dependence structure, consistent with the graphical model. From these simulated draws, the empirical joint probabilities

$$\hat{p}_{ab}^{(j,k)} = \frac{1}{S}\sum_{s=1}^{S} \mathbf{1}\{I_{j,t}^{(s)} = a, I_{k,t}^{(s)} = b\}, \quad a, b \in \{0, 1\},$$

are obtained, which approximate the true parameters $p_{ab}^{(j,k)\star}$. In this way, the estimated Markov Network, combined with Gibbs sampling, provides both a conditional dependence structure and consistent estimates of the bivariate joint probabilities that describe co-trading behavior.

Lastly, we have restricted attention to bivariate interactions due to the curse of dimensionality. The number of joint probabilities required to fully describe the distribution of trading indicators grows exponentially with the number of involved variables. If we consider N_{assets} assets and include N_{lags} time lags, then the number of distinct joint binary configurations is $2^{N_{\text{assets}} \times N_{\text{lags}}}$. As this grows quickly our analysis focuses on bivariate contemporaneous interactions between asset pairs and on univariate temporal dependence structures up to lag one.

6

Univariate model set-up, estimation, and asymptotic properties

In this chapter, we present the univariate formulation of the zero-inflated GARCH-X model for illiquid financial markets. For clarity and simplicity, we focus on the zero-inflated GARCH(1,1)-X specification, although the approach can be extended to general zero-inflated GARCH(p,q)-X models without conceptual difficulty. In Section 6.1, we conduct a descriptive analysis of both the carbon credit data and the selected liquid financial assets. We examine key time series properties, including stationarity, autocorrelation, and partial autocorrelation. Section 6.2 introduces the univariate model setup and details the estimation procedure, which is carried out using quasi-maximum likelihood estimation. Finally, we establish the consistency and asymptotic normality of the resulting estimator.

6.1 Descriptive analysis: ACF and PACF

Before presenting the univariate zero-inflated GARCH-X model, it is helpful to first examine some fundamental time series properties of the financial data introduced in Chapter 4. In particular, understanding features such as stationarity, autocorrelation, and partial autocorrelation provides valuable insight into the temporal structure of each return series. As a first step, we analyze the serial dependence of individual assets, as indicated by the blue edges in Figure 4.1. This helps to determine whether the data exhibit stylized facts commonly observed in financial time series, such as volatility clustering, which motivate the use of conditional heteroskedasticity models like GARCH. To assess temporal dependence, we compute the autocorrelation function (ACF) up to 30 lags for each asset. Figure 6.1 shows the autocorrelation of raw returns, squared returns, and their zero-filtered variants across the selected voluntary carbon credit assets. As defined in Section 2.1, the ACF measures linear dependence between lagged observations. In the ACF plots, autocorrelation at a non-zero lag is statistically significant if it exceeds the 95% confidence bounds, shown as dotted orange lines in Figures 6.1 - 6.3. Mathematically, the 95% confidence interval for the sample autocorrelation at each lag is computed by $1.96/\sqrt{n}$, where n is the sample size. This expression is derived under the null hypothesis that the underlying time series follows a white noise process, i.e. independently and identically distributed with zero autocorrelation; see also Definition 2.3. Significant values indicate

that past observations contain information relevant for future dynamics. For raw returns, this implies return predictability. In contrast, squared returns capture dependence in return magnitude. Significant autocorrelation in squared returns reflects volatility clustering.

In general, Figure 6.1 reveals consistent features that align with the stylized facts of financial time series. This supports the use of conditional heteroskedasticity models, such as GARCH or its extensions, as a natural next step in capturing the dynamics of carbon credit price.

To examine these features in more detail, we now turn to each row of Figure 6.1. Starting with the top row, the ACF of raw returns has minimal significant autocorrelation beyond lag zero for all carbon credit assets. This is expected in financial markets where asset returns are often modeled as serially uncorrelated but not independent. The absence of significant linear autocorrelation suggests that past returns provide limited predictive power for future returns in a linear sense. If significant autocorrelation had been present, this would have implied potential predictability and may have created opportunities for statistical arbitrage.

The second row of Figure 6.1, which shows the autocorrelation of returns excluding zero values, further confirms these stylized facts. In most assets, the autocorrelation remains close to zero and within the confidence bounds, although a few minor spikes are observed. These spikes are likely due to short-term illiquidity, data irregularities, or temporary distortions caused by the exclusion of zero returns, rather than reflecting genuine or persistent market autocorrelation.

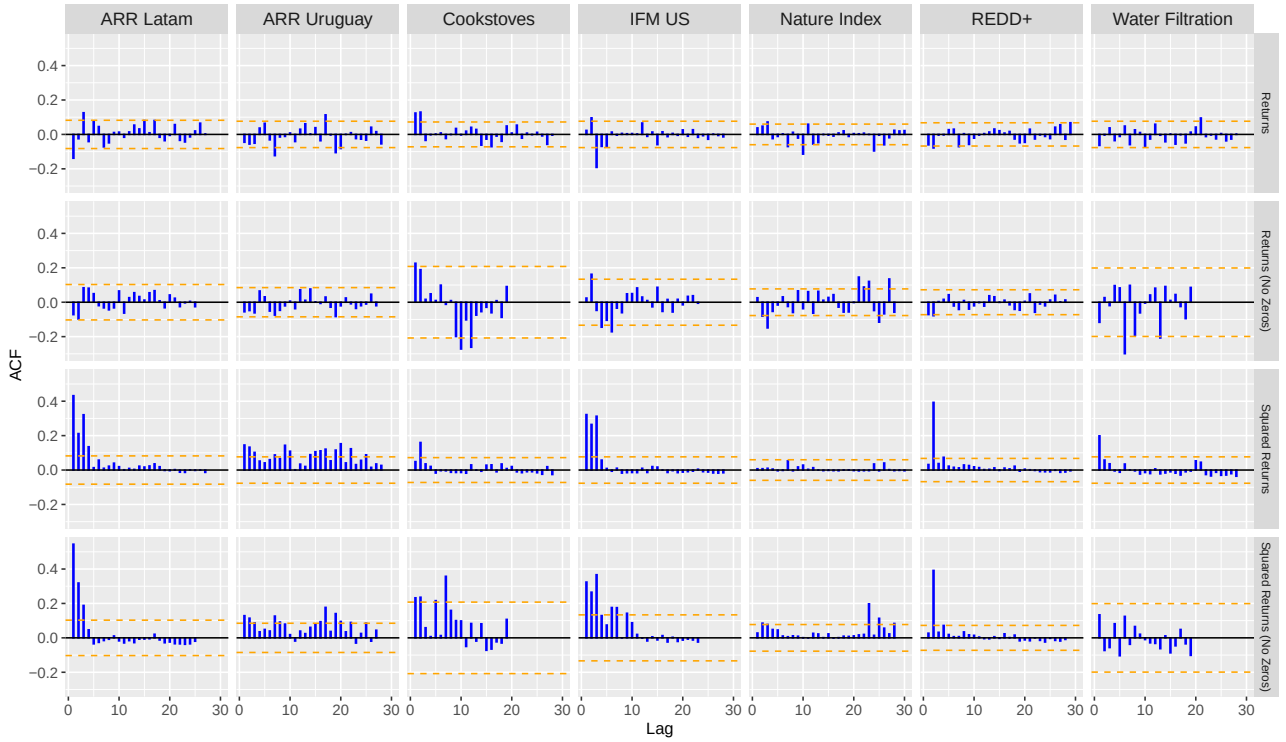


Figure 6.1: *Autocorrelation of the returns and the squared returns with and without zeros for all carbon credit assets*

The third and fourth rows, showing the ACF of squared returns (with and without zero returns), reveal statistically significant and gradually declining autocorrelation across several lags. This is particularly pronounced in ARR Latam and IFM, and is indicative of volatility clustering. The effect becomes slightly more apparent when zero returns are excluded. This suggests that non-trading days, which appear as zero returns, can weaken the visible pattern of volatility over time. These

zero returns interrupt sequences of high or low return magnitudes, which makes it harder to detect how volatility persists. In contrast, Water Filtration, REDD+, and ARR Uruguay show weaker autocorrelation in squared returns, though small positive values persist at short lags. Cookstoves and the Nature-Based Index show almost no autocorrelation, pointing to only mild volatility persistence. While less pronounced, these patterns still justify the application of GARCH-type models to capture time-varying risk.

Figure 6.2 shows the autocorrelation functions of both raw and squared returns for the selected conventional assets. Across all assets, the raw return series exhibit patterns close to white noise, with most autocorrelations lying within the 95% confidence bounds. Again, this is consistent with the stylized facts of financial return series.

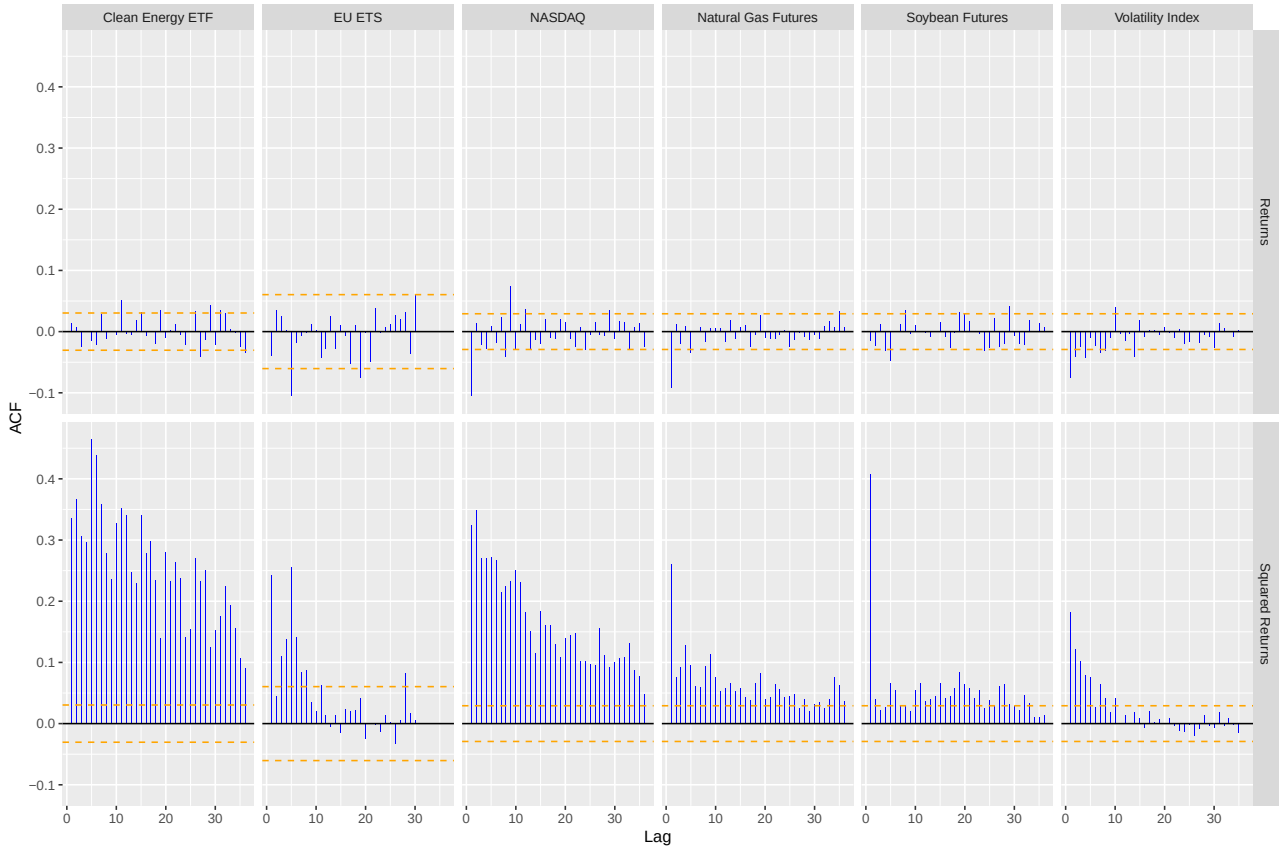


Figure 6.2: *Autocorrelation of the returns and the squared returns for the selected conventional assets*

In contrast, the ACFs of the squared returns show significant autocorrelation, often with a gradual decay pattern. This is indicative of time-varying volatility and supports the presence of conditional heteroskedasticity. Notably, the Soybean Futures series deviates from this pattern: it exhibits significant autocorrelation at the first lag and a few isolated lags later, but the smooth decay structure is not observed. Overall, the absence of serial correlation in the raw returns and the stationarity of higher-order moments suggest that these return series are weakly stationary.

Lastly, we turn to the partial autocorrelation function (PACF). As mentioned in Section 2.1, the PACF isolates the immediate contribution of each lag. This contrasts with the autocorrelation function, which reflects both direct and indirect correlations. In the context of financial returns, the PACF is particularly useful for identifying the appropriate lag structure in autoregressive models and for identifying both linear dependencies and latent structures in volatility dynamics. Figure 6.3 presents

the PACF plots for the selected voluntary carbon credit assets in a similar setup as in Figure 6.1.

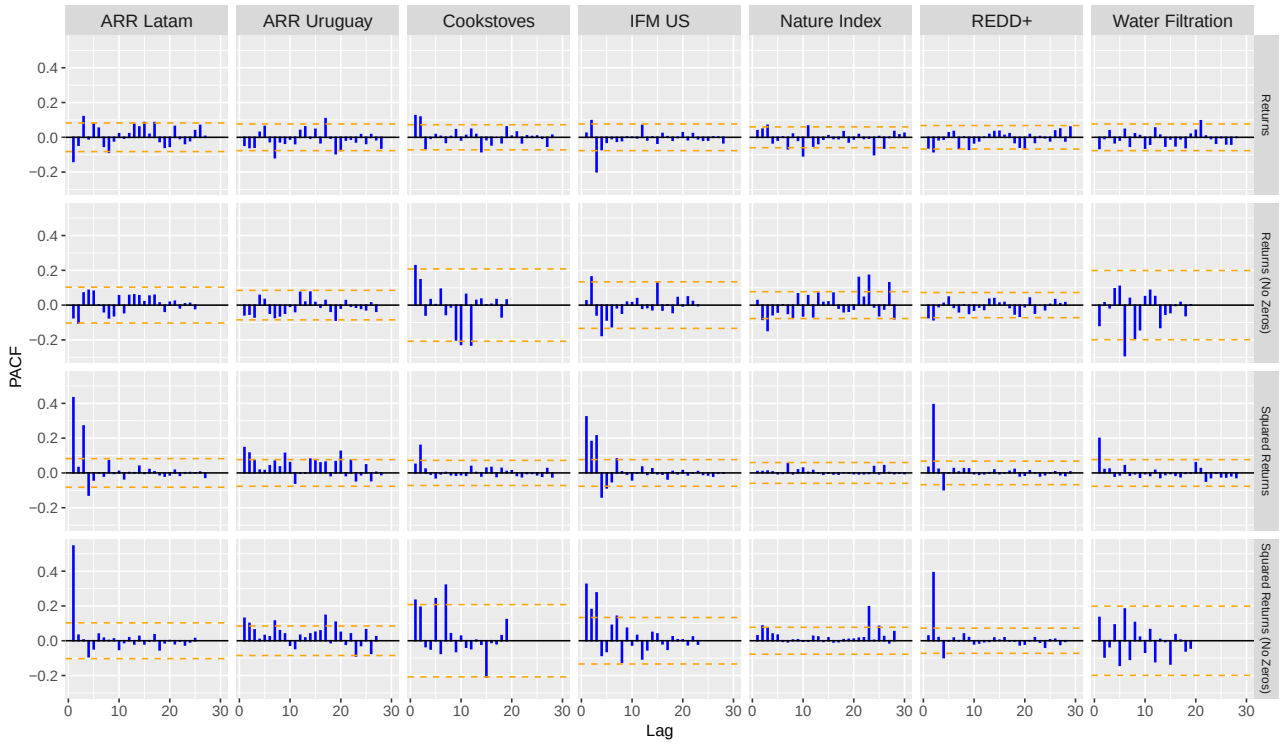


Figure 6.3: *Partial autocorrelation of the returns and the squared returns with and without zeros for all carbon credit assets*

According to Figure 6.3, the raw returns (top row) show that most lags lie within the confidence bounds. This suggests little to no direct partial autocorrelation in the majority of return series. However, some exceptions are present. In assets such as ARR Uruguay, Water Filtration, and Nature Index, the first lag is insignificant, but unexpected spikes occur at later lags. These delayed correlations are atypical and may point to irregular trading patterns or delayed reactions to market information.

After removing zero returns (second row), the PACF remains overall weak across most assets. Some series exhibit slightly stronger partial autocorrelation, while others become weaker, generally still within bounds of statistical insignificance. For instance, Nature Index and Water Filtration show marginally stronger effects, whereas ARR Latam and ARR Uruguay show weaker partial autocorrelation.

Looking at the squared returns in the third row, we see several assets show clear and significant PACF values at the first lag. This indicates the presence of volatility clustering. However, this effect is not visible across all assets. For assets such as Cookstoves, Nature Index and REDD+, significant lags start later. For the Nature Index all lags even stay within the confidence bounds, suggesting a more stable volatility profile or a lack of clear temporal structure in their variance.

Lastly, the squared returns after zero-removal in the last row show that all assets except of Water Filtration exhibit partial autocorrelation across multiple lags. Assets such as IFM US, and REDD+, retain similar PACF patterns before and after zero-return exclusion. The similarity in PACF structure before and after zero-return removal suggests that return dynamics are not distorted by non-trading days, which is typically a feature of more liquid markets. For some assets, such as Cookstoves, Nature Index, and Water Filtration, unexpected spikes appear at certain lags.

6.2 Univariate model setup and estimation

In this section, we derive the univariate zero-inflated GARCH(1,1)-X model as a special case of the general multivariate framework introduced in Chapter 4. Let $j \in \{1, \dots, N_{\text{asset}}\}$ denote an asset, and let $t \in \mathcal{T}_j := \{T_j, \dots, T_{\text{final}}\}$ be a discrete time step index, where T_j and T_{final} denote the first and the last time step at which the j-th asset price is observed respectively. We aim to model the return dynamics and conditional volatility of asset j using a zero-inflated GARCH(1,1)-X specification, defined as follows:

Definition 6.1 (Univariate zero-inflated GARCH(1,1)-X Model)**.** The returns and the conditional volatility of an asset $j \in \{1, \dots, N_{\text{asset}}\}$ for $t \in \mathcal{T}_j$ following a zero-inflated GARCH(1,1)-X model, can be written as

$$\begin{cases} r_{j,t} = I_{j,t} \cdot \sigma_{j,t}(\boldsymbol{\theta}_j^\star) \cdot \eta_{j,t}, \\ \sigma_{j,t}^2(\boldsymbol{\theta}_j^\star) = \alpha_{j,0}^\star + \alpha_{j,1}^\star \sigma_{j,t^*}^2(\boldsymbol{\theta}_j^\star) + \alpha_{j,2}^\star r_{j,t^*}^2 + \alpha_{j,3}^\star I_{j,t-1}, \end{cases} \tag{6.1}$$

where

- $r_{j,t} \in \mathbb{R}$ is the observed return of asset j at time t,
- $I_{j,t} \in \{0, 1\}$ is a binary indicator equal to one if there was any trading activity for asset j at time t, and zero otherwise.
- $\sigma_{j,t} \in \mathbb{R}^+$ is the conditional volatility of $r_{j,t}$,
- $\eta_{j,t}$ is the innovation for asset j at time t which is i.i.d and takes values in domain $(-\infty, +\infty)$,
- $t^* \in \mathcal{T}_j$ is the most recent time before $t \in \mathcal{T}_j$ such that $I_{j,t^*} = 1$. Formally, it is defined as $t^* := \max\{s < t : I_s = 1\}$,
- $\boldsymbol{\theta}_j^\star := (\alpha_{j,0}^\star, \alpha_{j,1}^\star, \alpha_{j,2}^\star, \alpha_{j,3}^\star)^\top$ denotes the true (unknown) parameter vector, belonging to the parameter space $\boldsymbol{\Theta}_j \subseteq (0, \infty) \times \mathbb{R}^3$.

Recall that three key modifications were introduced to the classical GARCH(1,1)-X specification in Definition 2.3 to better capture the characteristics of irregular trading activity and illiquid financial markets. These extensions result in the zero-inflated GARCH(1,1)-X model presented in Definition 6.1. First, we have introduced a binary trading indicator $I_{j,t}$ for each asset j, which equals one if trading occurs at time t, and zero otherwise. Note that the actual sample size now becomes a random variable, defined as

$$N_j = \sum_{t=T_j}^{T_{\text{final}}} I_{j,t}. \tag{6.2}$$

Next, we ensure that the volatility dynamics evolve only when relevant information becomes available. This is especially important in illiquid markets where trading is infrequent but informative. To achieve this, we make two other adjustments: we introduce an exogenous binary indicator $I_{j,t-1}$, which enters the volatility dynamics as GARCH-X component; and we define $t^* < t$ as the most recent time before t at which trading occurred, i.e. $t^* := \max\{s < t : I_{j,s} = 1\}$. Although returns are

observed at regular time intervals, this construction allows the model to respond only when actual trading occurs, making full use of the limited but informative signals available in illiquid markets.

Let us now examine the parameters involved in the conditional volatility dynamics. The terms $\alpha^\star_{j,1}\sigma^2_{j,t^*}$ and $\alpha^\star_{j,2}r^2_{j,t^*}$ capture the classical GARCH(1,1) dynamics, where current volatility depends on past volatility and past squared returns. Persistence in volatility is modeled via $\alpha^\star_{j,1}$, while $\alpha^\star_{j,2}$ accounts for the impact of large past returns. The third term, $\alpha^\star_{j,3}I_{j,t-1}$, is the exogenous component, which captures the effect of recent trading activity regardless of return magnitude. Intuitively, this allows the model to distinguish between periods of market activity and inactivity. For instance, if a trade occurred yesterday, today's volatility will increase through this additional term. To ensure non-negativity of the conditional variance and to reflect that trading generally introduces additional uncertainty, it is common to assume $\alpha^\star_{j,3} \geq 0$. Note that the trading indicator function $I_{j,t}$ influences both the conditional variance dynamics and the sample of returns used for estimation.

To estimate the parameters of the zero-inflated GARCH(1,1)-X model, we use the quasi-maximum likelihood estimation method, which is the usual estimation method for GARCH-type models [26]. Unlike classical maximum likelihood estimation, quasi-maximum likelihood does not require specifying the true distribution of the innovation sequence $(\eta_{j,t})_{t\geq 1}$. Instead, it proceeds by maximizing a pseudo log-likelihood function, which is in this case constructed as if the innovations were Gaussian – without necessarily assuming that this is the case. For more information on quasi-maximum likelihood estimation, the reader is referred to Section 2.2.

As seen in Section 2.2 on quasi-maximum likelihood estimation, it is standard to initialize a zero-inflated GARCH(1,1)-X model with

$$r^2_{j,\text{start}} = \sigma^2_{j,\text{start}} = \frac{\alpha_{j,0}}{1-\alpha_{j,1}-\alpha_{j,2}}.$$

Note that we do not initialize the exogenous covariate, since this part is assumed to be known.

Since the conditional variance depends on both past returns and on the exogenous trading indicators, the quasi-likelihood must also be constructed conditionally on the realized sequence of indicators $(I_{j,t})_{t\geq 1}$. This conditioning is important because $I_{j,t-1}$ directly affects the value of $\sigma^2_{j,t}$. The Gaussian likelihood function for asset j, conditional on the return series $(r_{j,t})_{t\geq 1}$ and the indicator process $(I_{j,t})_{t\geq 1}$, is given by:

$$\mathcal{L}_{j,T_{\text{final}}}\left(\boldsymbol{\theta}_j \,\middle|\, (r_{j,t}),(I_{j,t})\right) = \prod_{t=T_j}^{T_{\text{final}}}\left(\frac{1}{\sqrt{2\pi\sigma^2_{j,t}(\boldsymbol{\theta}_j)}}\exp\left(-\frac{r^2_{j,t}}{2\sigma^2_{j,t}(\boldsymbol{\theta}_j)}\right)\right)^{I_{j,t}}. \tag{6.3}$$

This means that the Gaussian log-likelihood function becomes:

$$\mathcal{L}_{\log,j,T_{\text{final}}}\left(\boldsymbol{\theta}_j \,\middle|\, (r_{j,t}),(I_{j,t})\right) = \sum_{t=T_j}^{T_{\text{final}}} I_{j,t}\left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2_{j,t}(\boldsymbol{\theta}_j)) - \frac{r^2_{j,t}}{2\sigma^2_{j,t}(\boldsymbol{\theta}_j)}\right] \tag{6.4}$$

$$= -\frac{1}{2}\sum_{t=T_j}^{T_{\text{final}}} I_{j,t}\left[\log(2\pi) + \log(\sigma^2_{j,t}(\boldsymbol{\theta}_j)) + \frac{r^2_{j,t}}{\sigma^2_{j,t}(\boldsymbol{\theta}_j)}\right] \tag{6.5}$$

Then, the quasi-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_j$ is defined as any measurable solution to the optimization problem:

$$\widehat{\boldsymbol{\theta}}_j \in \arg\max_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j} \mathcal{L}_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j). \tag{6.6}$$

For the estimation of the parameter $\boldsymbol{\theta}_j$, we use the `garchx` package in R, see also [59]. The estimation procedure implemented in the `garchx` package uses quasi-maximum likelihood estimation, as discussed in Section 2.2. Specifically, it minimizes the negative (average) Gaussian log-likelihood function as defined in Equation (6.5). The conditional variance is recursively defined and depends on exogenous variables which are in our case the indicator functions $I_{j,t-1}$. This objective function is evaluated by `garchxObjective()` and relies on the recursively computed variance via `garchxRecursion()`. Optimization is done iteratively through the `nlminb()` function, a gradient-based optimizer, that minimizes the negative Gaussian quasi-log-likelihood, subject to constraints such as the non-negativity of variance parameters. After the optimal parameters are found, the package uses `optimHess()` to compute the hessian matrix of second derivatives numerically. This hessian is then used to calculate standard errors and confidence intervals.

However, two issues arise during estimation using the `garchx` package. First, the dataset contains a large number of zero return values that reflect periods of illiquidity where no trades occur, as discussed in Chapter 5. Second, several non-defined values, indicated by `NA`, are present due to asynchronous start dates across assets and further illiquidity-driven gaps. The `garchx` package, in its current implementation, does not accommodate time series containing zeros or missing values; estimation fails when such entries are present. To ensure the model can be estimated, we pre-process the data by first removing `NA` observations. Subsequently, we omit the zero return values which is consistent with the log-likelihood as defined in Equation (6.5). Moreover, this approach aligns with the model specification in Equation (6.1), where the conditional volatility recursion is only updated at the most recent trading time t^*, whenever a trade occurs. This mechanism is formalized in the following proposition.

Proposition 6.1 (Time-filtered GARCH-X volatility dynamics)**.** *Let $(r_t)_{t\geq 1}$ be a time series following a zero-inflated GARCH(1,1)-X model. Let $I_t := \mathbf{1}_{\{r_t \neq 0\}}$ for $t \geq 1$. Let $t_1 = \min\{t \geq 1 : I_t = 1\}$, and recursively, $t_i := \min\{t > t_{i-1} : I_t = 1\}$ for $i > 1$. For $i \geq 1$, define $\tilde{r}_i := r_{t_i}$.*

Then, the process $(\tilde{r}_i)_{i\geq 1}$ follows a (classical) GARCH(1, 1)-X model with exogenous covariate $(I_{t_i-1})_{i\geq 1}$ in the sense that

$$\begin{cases} \tilde{r}_i = \tilde{\eta}_i \tilde{\sigma}_i, \\ \tilde{\sigma}_i^2 = \alpha_0 + \alpha_1 \tilde{\sigma}_{i-1}^2 + \alpha_2 \tilde{r}_{i-1}^2 + \alpha_3 I_{t_i-1}, \end{cases} \tag{6.7}$$

where $\tilde{\eta}_i := \eta_{t_i}$ is (still) i.i.d. and $\tilde{\sigma}_i := \sigma_{t_i}$.

Proof. We fix $i \geq 1$ and remark that $I_{t_i} = 1$ by construction. Since the sequence (t_i) indexes only the times when trades occur, the most recent observed time prior to t_i is

$$t^* := \max\{s < t_i : I_s = 1\} = t_{i-1}.$$

By Equation (6.1), replacing t by t_i and t^* by t_{i-1}, we know that

$$\begin{cases} r_{t_i} = I_{t_i} \sigma_{t_i} \eta_{t_i}, \\ \sigma_{t_i}^2 = \alpha_0 + \alpha_1 \sigma_{t_{i-1}}^2 + \alpha_2 r_{t_{i-1}}^2 + \alpha_3 I_{t_i-1}, \end{cases}$$

Because $I_{t_i} = 1$, we obtain $r_{t_i} = \sigma_{t_i}\eta_{t_i}$, which completes the proof. □

6.3 Consistency of the QMLE for zero-inflated GARCH-X models

This section establishes the consistency of the quasi-maximum likelihood estimator for zero-inflated GARCH(1,1)-X models. We begin by outlining the assumptions required for consistency, followed by stating the consistency theorem and its proof. Essentially, the proof consists of two steps, each corresponding to a distinct lemma. The first step applies Theorem 2.1 from Kristensen and Rahbek to establish that the estimator $\widehat{\boldsymbol{\theta}}_j^*$ of a stationary and ergodic process that is asymptotically equivalent to the process of interest, converges in probability to the true parameter $\boldsymbol{\theta}_j^\star$. In the second step, we apply Theorem 2.8 from Kristensen and Shin to show that the estimator $\widehat{\boldsymbol{\theta}}_j^*$ is asymptotically equivalent to the QMLE $\widehat{\boldsymbol{\theta}}_j$. This asymptotic equivalence implies that the consistency established in the first step extends to the QMLE $\widehat{\boldsymbol{\theta}}_j$.

The consistency and asymptotic normality of the QMLE have been studied for standard GARCH models and various extensions, including GARCH-X models (e.g., [27], [36], [40]). This thesis extends these results to the setting of multivariate zero-inflated GARCH-X models. Specifically, our theoretical contribution extends the proof frameworks developed by Han and Kristensen [36]. They examined the asymptotic properties of QMLE in GARCH-X models and proved results for both stationary and non-stationary covariates. Their findings show that QMLE remains consistent regardless of whether a covariate is stationary or not. However, the rate of convergence and the limiting distribution of the estimator are shown to be sensitive to the stationarity properties of the covariate [36].

The consistency result for GARCH-X models, as established by Han and Kristensen, requires adaptation in our setting due to two key differences. First, our framework is based on a modified version of the classical GARCH-X model. Second, the actual sample size is a random variable, as defined in Equation (6.2). Furthermore, we restrict attention to the case where the exogenous covariates are stationary. In particular, the binary trading indicator $I_{j,t}$ is included in the model as the exogenous covariate. The assumption of stationarity is motivated by the observed patterns in trading activity, as illustrated in Figures 4.2–4.5 in Chapter 4. The case involving non-stationary covariates within the zero-inflated GARCH-X framework remains an open question and is left for future research.

We now outline the set of assumptions required to prove consistency of the quasi-maximum likelihood estimator. The notation $\mathcal{F}_t$ refers to the natural filtration:

Assumption 6.1 (Assumptions for consistency of the QMLE)**.**

(*i*) $\{(\eta_{j,t}, I_{j,t})\}$ *is stationary and ergodic with* $\mathbb{E}[\eta_{j,t} \mid \mathcal{F}_{t-1}] = 0$ *and* $\mathbb{E}[\eta_{j,t}^2 \mid \mathcal{F}_{t-1}] = 1$.

(*ii*) $\mathbb{E}\left[\log\left(\alpha_{j,1}^\star \eta_{j,t}^2 + \alpha_{j,2}^\star\right)\right] < 0$ *and* $\mathbb{E}[I_{j,t}] > 0$.

(*iii*) *Given that* $0 < \underline{\alpha}_{j,0} \leq \overline{\alpha}_{j,0} < +\infty, \overline{\alpha}_{j,1} < +\infty, \overline{\alpha}_{j,2} < 1$, *and* $\overline{\alpha}_{j,3} < +\infty$, *we have that*

$$\boldsymbol{\Theta}_j = \left\{\boldsymbol{\theta}_j = (\alpha_{j,0}, \alpha_{j,1}, \alpha_{j,2}, \alpha_{j,3})^\top : \underline{\alpha}_{j,0} \leq \alpha_{j,0} \leq \overline{\alpha}_{j,0}, 0 \leq \alpha_{j,1} \leq \overline{\alpha}_{j,1}, 0 \leq \alpha_{j,2} \leq \overline{\alpha}_{j,2}, 0 \leq \alpha_{j,3} \leq \overline{\alpha}_{j,3}\right\}.$$

The true value $\theta_j^\star \in \Theta_j$ *with* $(\alpha_{j,0}^\star, \alpha_{j,3}^\star) \neq (0,0)$.

(*iv*) *For any* $(a,b) \neq (0,0)$: $a\eta_{j,t}^2 + bI_{j,t}^2 \mid \mathcal{F}_{t-1}$ *has a non-degenerate distribution.*

Assumption 6.1(i) is a generalization of the conditions found in [21], where asymptotic properties of the quasi-maximum likelihood estimator are derived for classical GARCH processes with martingale difference errors. The assumption is weaker than the i.i.d. assumption from [42]. In our context, we have adapted the covariate from a general covariate, for example $\mathbf{x}_{j,t}$, to the trading indicator function $I_{j,t}$.

The moment conditions in Assumption 6.1(ii) guarantee the existence of a strictly stationary solution to Equation (6.1) at the true parameter value $\boldsymbol{\theta}_j^\star$. They also ensure that this solution has finite polynomial moments. In particular, these conditions allow for integrated GARCH (IGARCH) processes, where $\alpha + \beta = 1$. Moreover, they impose only mild moment restrictions on the regressors to allow a broad class of exogenous covariates. We do exclude explosive volatility when the trading indicator $I_{j,t}$ is stationary. In this case, we expect that the arguments of [42] extend to GARCH-X models with $\mathbb{E}\big[\log(\alpha_{j,1}^\star \eta_{j,t-1}^2 + \alpha_{j,2}^\star)\big] > 0$. This would imply that the estimated parameters remain consistent, while the intercept $\widehat{\alpha}_{j,0}$ is inconsistent. This inconsistency arises because, in IGARCH-type models, there is no finite unconditional variance, which makes $\alpha_{j,0}$ unidentified in the long run. This restriction is important to prevent non-ergodic or explosive behavior and to ensure that the volatility dynamics remain meaningful in illiquid markets.

The compactness condition in Assumption 6.1(iii) could be weakened by the arguments of [42]. However, this would lead to more complicated proofs, and so we maintain the compactness assumption here for simplicity. The requirement that $(\alpha_{j,1}^\star, \alpha_{j,3}^\star) \neq (0,0)$ is needed to ensure the identification of $\alpha_{j,2}^\star$. Indeed, in the case where $(\alpha_{j,1}^\star, \alpha_{j,3}^\star) = (0,0)$, we have $\sigma_{j,t}^2 := \sigma_{j,t}^2(\boldsymbol{\theta}_j^\star) \xrightarrow{a.s.} \alpha_{j,0}^\star/(1-\alpha_{j,2}^\star)$, and so we would not be able to jointly identify $\alpha_{j,0}^\star$ and $\alpha_{j,2}^\star$.

Lastly, the non-degeneracy condition in Assumption 6.1(iv) is also needed for identification, as it rules out (dynamic) collinearity between $r_{j,t-1}^2$ and $I_{j,t}^2$. This assumption is similar to the no-collinearity restriction described in [42]. Intuitively, this assumption provides that the past returns and exogenous covariates have separate and independent information about volatility. In other words, it prevents these variables from moving perfectly together in such a way that their individual effects on the conditional variance could not be separated.

Note that the random sample size increases and remains non-degenerate as $T_{\text{final}} \to \infty$. This follows from Assumption 6.1(i), which states that the joint process $\{(\eta_{j,t}, I_{j,t})\}$ is stationary and ergodic. As a direct consequence, the marginal process $(I_{j,t})_{t\geq 1}$ is itself stationary and ergodic. By the ergodic theorem, we have

$$\frac{1}{T_{\text{final}}} \sum_{t=T_j}^{T_{\text{final}}} I_{j,t} \xrightarrow{\text{a.s.}} \mathbb{E}[I_{j,t}] > 0, \tag{6.8}$$

which implies that the number of non-zero observations N_j diverges to infinity almost surely as the total sample size $\mathcal{T}_j \to \infty$. This condition is important for proving consistency, since extremum estimators such as the quasi-maximum likelihood estimator require the objective function to be based on a growing number of observations. Without this, the Law of Large Numbers cannot be applied and convergence to the true parameter is not guaranteed.

To derive the asymptotic properties of $\widehat{\boldsymbol{\theta}}_j$, we establish some preliminary results. The first lemma states that a stationary solution to the model at the true parameter values exists:

Lemma 6.1. *Under Assumption 6.1, there exists a stationary and ergodic solution to Equation* (6.1) *at* $\boldsymbol{\theta}_j^\star$ *satisfying* $\mathbb{E}[\sigma_{j,t}^{2s}] < +\infty$ *and* $\mathbb{E}[r_{j,t}^{2s}] < +\infty$ *for some* $0 < s < 1$.

The proof of Lemma 6.1 can be found in Appendix B, where we have adapted the existing proof to our specific model setting. In what follows, we work under the implicit assumption that the stationary solution of the process has been observed. The next lemma shows that, for any parameter value of $\boldsymbol{\theta}_j$ within the parameter space, the volatility-ratio process $s_{j,t}(\boldsymbol{\theta}_j)$ is well-approximated by a stationary version. Here, we define $s_{j,t}(\boldsymbol{\theta}_j)$ as the variance-ratio process given by

$$s_{j,t}(\boldsymbol{\theta}_j) := \frac{\sigma_{j,t}^2(\boldsymbol{\theta}_j)}{\sigma_{j,t}^2},$$

where $\sigma_{j,t}^2 := \sigma_{j,t}^2(\boldsymbol{\theta}_j^\star)$. We denote $s_{j,t}^*$ as the stationary version of $s_{j,t}$. Here, $s_{j,t}^*$ is a stationary sequence which is asymptotically equivalent to $s_{j,t}(\boldsymbol{\theta}_j)$ which will be formally defined in Lemma 6.2. By definition of asymptotic equivalence, it holds that

$$\lim_{T_{\text{final}} \to +\infty} \frac{s_{j,t}^*(\boldsymbol{\theta}_j)}{s_{j,t}(\boldsymbol{\theta}_j)} = 1 \quad \text{almost surely.}$$

Lemma 6.2 shows that the expected deviation between $s_{j,t}(\boldsymbol{\theta}_j)$ and $s_{j,t}^*(\boldsymbol{\theta}_j)$, uniformly over the parameter space, is bounded above by a term that decreases exponentially in time, since $\alpha_{j,2} < 1$ by assumption.

Lemma 6.2. *Under Assumption 6.1, with* $s > 0$ *given in Lemma 6.1, there exists some* $K_s < +\infty$ *such that*

$$\mathbb{E}\left[sup_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j} \left| s_{j,t}(\boldsymbol{\theta}_j) - s_{j,t}^*(\boldsymbol{\theta}_j) \right|^s \right] \leq K_s \alpha_{j,2}^{st}.$$

where

$$s_{j,t}^*(\boldsymbol{\theta}_j) = \frac{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\sigma_{j,t}^2(\boldsymbol{\theta}_j^\star)}, \quad \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j) := \sum_{i=1}^{+\infty} \alpha_{j,2}^{i-1}(\alpha_{j,0} + \alpha_{j,1} r_{j,t}^2 + \alpha_{j,3} I_{j,t-1}^2).$$

The process $\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)$ *is stationary and ergodic with* $\mathbb{E}[\sup_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j} \tilde{\sigma}_{j,t}^{2s}(\boldsymbol{\theta}_j)] < +\infty$.

The proof of Lemma 6.2 can be found in Appendix B, where we have adapted the existing proof to our specific model setting. We can now show the first asymptotic main result, showing the consistency of the quasi-maximum likelihood estimator of a zero-inflated GARCH(1,1)-X model.

Theorem I (Consistency quasi-maximum likelihood estimator zero-inflated GARCH-X)**.** *Under Assumption 6.1, the QMLE* $\widehat{\boldsymbol{\theta}}_j$ *is consistent for all* $j \in \{1, \cdots, N_{assets}\}$, *i.e.* $\widehat{\boldsymbol{\theta}}_j \xrightarrow{p} \boldsymbol{\theta}_j^\star$.

Proof. First, let us define

$$\widehat{\boldsymbol{\theta}}_j^* = \arg\max_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j} \widetilde{\mathcal{L}}_{\log,j,T_{\text{final}}}^*(\boldsymbol{\theta}_j),$$

where

$$\widetilde{\mathcal{L}}_{\log,j,T_{\text{final}}}^*(\boldsymbol{\theta}_j) := \sum_{t=T_j}^{T_{\text{final}}} I_{j,t} \left[-\log\left(s_{j,t}^*(\boldsymbol{\theta}_j)\right) - \left(\frac{1}{s_{j,t}^*(\boldsymbol{\theta}_j)} - 1\right) \eta_{j,t}^2 \right] = \sum_{t=T_j}^{T_{\text{final}}} I_{j,t} \cdot l^*(\boldsymbol{\theta}_j),$$

Note that we have the decomposition

$$\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^\star = \underbrace{\left(\widehat{\boldsymbol{\theta}}_j - \widehat{\boldsymbol{\theta}}_j^*\right)}_{\xrightarrow{\text{a.s.}} 0} + \underbrace{\left(\widehat{\boldsymbol{\theta}}_j^* - \boldsymbol{\theta}_j^\star\right)}_{\xrightarrow{\text{P}} 0}.$$

Therefore, the proof is finished as a consequence of Lemma I and Lemma II.

□

Lemma I establishes the consistency of the auxiliary estimator $\widehat{\boldsymbol{\theta}}_j^*$ under the assumption that the underlying process is stationary and ergodic. Lemma II then establishes the asymptotic equivalence between the quasi-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_j$ and the auxiliary estimator $\widehat{\boldsymbol{\theta}}_j^*$. Taken together, these results imply the consistency of the quasi-maximum likelihood estimator $\widehat{\boldsymbol{\theta}}_j$ for the true parameter $\boldsymbol{\theta}_j^\star$. To formulate and prove Lemma I and Lemma II, we elaborate on the introduced notation and its interpretation. First of all, we consider a normalized version of the quasi-log-likelihood function, denoted by $\widetilde{\mathcal{L}}_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j)$. This expression is obtained by subtracting the quasi-log-likelihood evaluated at the true parameter value $\boldsymbol{\theta}_j^\star$. That is,

$$\begin{aligned}
\widetilde{\mathcal{L}}_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j) &:= \mathcal{L}_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j) - \mathcal{L}_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j^\star) \\
&= \left\{ \sum_{t=T_j}^{T_{\text{final}}} I_{j,t} \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_{j,t}^2(\boldsymbol{\theta}_j)) - \frac{r_{j,t}^2}{2\sigma_{j,t}^2(\boldsymbol{\theta}_j)} \right] \right\} \\
&\quad - \left\{ \sum_{t=T_j}^{T_{\text{final}}} I_{j,t} \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_{j,t}^2) - \frac{r_{j,t}^2}{2\sigma_{j,t}^2} \right] \right\} \\
&= \sum_{t=T_j}^{T_{\text{final}}} I_{j,t} \left[-\log(s_{j,t}(\boldsymbol{\theta}_j)) - \left(\frac{1}{s_{j,t}(\boldsymbol{\theta}_j)} - 1 \right) \eta_{j,t}^2 \right]
\end{aligned}$$

where $\mathcal{L}_{\log,j,T_{\text{final}}}(\cdot)$ denotes the quasi-log-likelihood function associated with asset j, and

$$\sigma_{j,t}^2(\boldsymbol{\theta}_j) = \alpha_{j,0} + \alpha_{j,1}\sigma_{j,t^*}^2(\boldsymbol{\theta}_j) + \alpha_{j,2} r_{j,t^*}^2 + \alpha_{j,3} I_{j,t-1}.$$

Remember that we have the

$$\sigma_{j,t}^2 := \sigma_{j,t}^2(\boldsymbol{\theta}_j^\star)$$

where σ_t^2 denotes the true data-generating conditional volatility process, and $\sigma_{j,t}^2(\boldsymbol{\theta}_j)$ denotes the conditional variance function evaluated at a given parameter vector $\boldsymbol{\theta}_j$. More precisely, for each time point t, the volatility model generates a value for the conditional variance based on the parameter values $\boldsymbol{\theta}_j$. Although notationally expressed as a function, $\sigma_{j,t}^2(\boldsymbol{\theta}_j)$ should be interpreted as a random variable once a specific parameter vector is fixed. Note that this normalization does not affect the QMLE, since the subtracted term $-\log\sigma_t^2 - r_t^2/\sigma_t^2$ is only depending on the "true values" $r_{j,t}$ and $\sigma_{j,t}$, and is therefore independent of the parameter $\boldsymbol{\theta}_j$ [36]. It should be noted, however, that the process $s_{j,t}(\boldsymbol{\theta}_j)$ is non-stationary process generally. This is due to the fact that $\sigma_t^2(\boldsymbol{\theta}_j)$ is initialized at a fixed value, and the covariate process $I_{j,t}$ may itself be non-stationary [36].

We now proceed to state and prove the two lemmas using the notation introduced above.

Lemma I (Consistency of $\widehat{\boldsymbol{\theta}}_j^*$)**.** *Let $s_{j,t}^*$ be a stationary and ergodic process that is asymptotically equivalent to the process $s_{j,t}$, as defined in Lemma 6.2. Under Assumption 6.1, the QMLE $\widehat{\boldsymbol{\theta}}_j^*$ of $s_{j,t}^*$ is strongly consistent for all $j \in \{1, \ldots, N_{assets}\}$, i.e.*

$$\widehat{\boldsymbol{\theta}}_j^* \xrightarrow{a.s.} \boldsymbol{\theta}_j^\star.$$

Proof. To show the consistency of $\widehat{\boldsymbol{\theta}}_j^*$, we apply Theorem 2.1 from Kristensen and Rahbek. We need to verify the following five conditions:

(i) **Compact parameter space**
The parameter space $\boldsymbol{\Theta}_j$ is a compact Euclidean space. This holds by assumption.

(ii) **Continuity of log-likelihood**
We need to have that

$$\boldsymbol{\theta}_j \mapsto -I_{j,t}\left[-\log(s_{j,t}^*(\boldsymbol{\theta}_j)) - \left(\frac{1}{s_{j,t}^*(\boldsymbol{\theta}_j)} - 1\right)\eta_{j,t}^2\right],$$

is continuous almost surely. This follows directly from the continuity of the mapping $\boldsymbol{\theta}_j \mapsto s_{j,t}^*(\boldsymbol{\theta}_j)$, as given in Lemma 6.2. Note that continuity of each individual term in the infinite sum, defining $\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)$, alone does not imply that the sum is continuous. In general, an infinite sum of continuous functions is continuous if the convergence is uniform. Lemma 6.2 directly addresses this by providing a uniform bound uniformly over the parameter space. In particular, the lemma states that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j} \left| s_{j,t}(\boldsymbol{\theta}_j) - s_{j,t}^*(\boldsymbol{\theta}_j) \right|^s \right] < +\infty,$$

and that the infinite sum decays geometrically, while also ensuring that

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j} \tilde{\sigma}_{j,t}^{2s}(\boldsymbol{\theta}_j)\right] < +\infty.$$

This implies that the sequence of partial sums converges uniformly to $\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)$. Consequently, Lemma 6.2 is sufficient to guarantee the continuity of $s_{j,t}^*(\boldsymbol{\theta}_j)$, and therefore the continuity of the log-likelihood function.

(iii) **Existence and finiteness of limit**
We will show that

$$\lim_{T_{\text{final}} \to +\infty} \frac{\widetilde{\mathcal{L}}_{\log,j,T_{\text{final}}}^*(\boldsymbol{\theta}_j)}{T_{\text{final}}} \xrightarrow{p} \widetilde{\mathcal{L}}_{\log,j,T_{\text{final}}}^*(\boldsymbol{\theta}_j),$$

where the limit exists, for all $\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j$. This follows by the Law of Large Numbers for stationary and ergodic sequences if the limit exists. Indeed, the limit is well-defined, since

$$-\log(s_{j,t}^*(\boldsymbol{\theta}_j)) - \left(\frac{1}{s_{j,t}^*(\boldsymbol{\theta}_j)} - 1\right)\eta_{j,t}^2 < \log(\alpha_{j,0}/\alpha_{j,0}^\star)$$

such that,

$$\mathbb{E}\Big[I_{j,t}\Big] \cdot \mathbb{E}\left[\left(-\log(s^*_{j,t}(\boldsymbol{\theta}_j)) - \left(\frac{1}{s^*_{j,t}(\boldsymbol{\theta}_j))} - 1\right)\eta^2_{j,t}\right)^+ \,\middle|\, I_{j,t} = 1\right] < +\infty.$$

This follows from the fact that $s^*_{j,t}(\boldsymbol{\theta}_j) \geq \alpha_{j,0}/\alpha^\star_{j,0}$, and therefore $-\log\left(s^*_{j,t}(\boldsymbol{\theta}_j)\right) \leq -\log(\alpha_{j,0}/\alpha^\star_{j,0})$.

(iv) **Identification (unique maximizer)**
We want to show that $\widetilde{\mathcal{L}}^*_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j) < \widetilde{\mathcal{L}}^*_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}^\star_j) = 0$ for all $\boldsymbol{\theta}_j \neq \boldsymbol{\theta}^\star_j$, i.e. the expected log-likelihood is uniquely maximized at the true parameter. This result is independent of the choice of the Gaussian surrogate likelihood for quasi-maximum likelihood.

Observe that by construction $s^*_{j,t}(\boldsymbol{\theta}^\star_j) = 1$, which implies

$$-\log(s_{j,t}(\boldsymbol{\theta}^\star_j)) - \left(\frac{1}{s_{j,t}(\boldsymbol{\theta}^\star_j))} - 1\right)\eta^2_{j,t} = 0,$$

resulting in the log-likelihood $\widetilde{\mathcal{L}}^*_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}^\star_j) = 0$. Moreover, it holds that

$$\alpha^\star_{j,0} \leq \log\left(\tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}^\star_j)\right) \quad \text{such that} \quad \mathbb{E}\left[\left(\log\left(\tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}^\star_j)\right)\right)^-\right] < +\infty.$$

By Jensen's inequality and Lemma 6.2, we then obtain

$$\mathbb{E}\left[\left(\log\left(\tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}^\star_j)\right)\right)^+\right] \leq \left(\log\left(\mathbb{E}\left[\tilde{\sigma}^{2s}_{j,t}(\boldsymbol{\theta}^\star_j)\right]\right)\right)^+ / s < +\infty.$$

Therefore, $\mathbb{E}\left[\left|-\log(s^*_{j,t}(\boldsymbol{\theta}^\star_j)) - \left(\frac{1}{s^*_{j,t}(\boldsymbol{\theta}^\star_j))} - 1\right)\eta^2_{j,t}\right|\right] < +\infty$ is well-defined. We distinguish two cases, having $\boldsymbol{\theta}_j \neq \boldsymbol{\theta}^\star_j$:

(a) $\widetilde{\mathcal{L}}^*_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j) = -\infty$ (the log-likelihood diverges). Then, clearly

$$\widetilde{\mathcal{L}}^*_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}^\star_j) > -\infty = \widetilde{\mathcal{L}}^*_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j).$$

(b) $\widetilde{\mathcal{L}}^*_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j) \in (-\infty, +\infty)$ (the log-likelihood is finite). If this holds, we can write the following:

$$\begin{aligned}\widetilde{\mathcal{L}}^*_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j) &= \mathbb{E}\Big[I_{j,t}\Big] \cdot \mathbb{E}\left[-\log(s^*_{j,t}(\boldsymbol{\theta}_j)) - \left(\frac{1}{s^*_{j,t}(\boldsymbol{\theta}_j))} - 1\right)\eta^2_{j,t} \,\middle|\, I_{j,t} = 1\right] \\ &= -\mathbb{E}\Big[I_{j,t}\Big] \cdot \mathbb{E}\left[\log(s^*_{j,t}(\boldsymbol{\theta}_j)) + \left(\frac{1}{s^*_{j,t}(\boldsymbol{\theta}_j))} - 1\right) \,\middle|\, I_{j,t} = 1\right],\end{aligned}$$

where we have used that $\mathbb{E}[\eta^2_{j,t}|\mathcal{F}_{t-1}] = 1$ by assumption. Thus, $\widetilde{\mathcal{L}}^*_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j) \leq 0 = \widetilde{\mathcal{L}}^*_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}^\star_j)$, with equality if and only if $s^2_{j,t}(\boldsymbol{\theta}_j) = 1$ almost surely under the assumption $\mathbb{E}[I_{j,t}] > 0$.

Now suppose that $s^2_{j,t}(\boldsymbol{\theta}_j) = 1$ almost surely $\iff \tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j) = \tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}^\star_j)$ almost surely. With

$c_i(\boldsymbol{\theta}_j) := (\alpha_{j,1}\alpha_{j,2}^{i-1}, \alpha_{j,3}\alpha_{j,2}^{i-1})'$, we claim that $\alpha_{j,0} = \alpha_{j,0}^\star$ and $c_i(\boldsymbol{\theta}_j) = c_i(\boldsymbol{\theta}_j^\star)$ for all $i \geq 1$. This in turn implies the aimed result $\boldsymbol{\theta}_j = \boldsymbol{\theta}_j^\star$. We show this by contradiction.

Let $m > 0$ be the smallest integer for which $c_i(\boldsymbol{\theta}_j^\star) \neq c_i(\boldsymbol{\theta}_j)$ for all $i \geq 1$. Then, we obtain:

$$a_{j,0}r_{j,t-m}^2 + b_{j,0}I_{t-m}^2 = \alpha_{j,0} - \alpha_{j,0}^\star + \sum_{i=1}^{+\infty} a_{j,i}r_{j,t-m-i}^2 + \sum_{i=1}^{+\infty} b_{j,i}I_{j,t-m-i}^2 \tag{6.9}$$

where $a_{j,i} := \alpha_{j,1}^\star(\alpha_{j,2}^\star)^{i-1} - \alpha_{j,1}\alpha_{j,2}^{i-1}$ and $b_{j,i} := \alpha_{j,3}^\star(\alpha_{j,2}^\star)^{i-1} - \alpha_{j,3}\alpha_{j,2}^{i-1}$. The right hand side of Equation (6.9) belongs to $\mathcal{F}_{t-m-1}$. Therefore, $a_{j,0}r_{j,t-m}^2 + b_{j,0}I_{t-m}^2 \mid \mathcal{F}_{t-m-1}$ is constant. This is ruled out by Assumption 6.1(iv). Hence, we have a contradiction.

So, the only way we have $s_{j,t}^*(\boldsymbol{\theta}_j) = 1$ almost surely, is when $\boldsymbol{\theta}_j = \boldsymbol{\theta}_j^\star$. Hence, the expected log-likelihood is uniquely maximized at $\boldsymbol{\theta}_j^\star$, and identification holds.

(v) **Uniform integrability**

Lastly, we need to verify that $\mathbb{E}[I_{j,t}] \cdot \mathbb{E}[\sup_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j} -\log(s_{j,t}^*(\boldsymbol{\theta}_j)) - \left(\frac{1}{s_{j,t}^*(\boldsymbol{\theta}_j))} - 1\right)\eta_{j,t}^2 \mid I_{j,t} = 1] < +\infty$. This follows from:

$$\sup_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j} \left\{ -\log(s_{j,t}^*(\boldsymbol{\theta}_j)) - \left(\frac{1}{s_{j,t}^*(\boldsymbol{\theta}_j))} - 1\right)\eta_{j,t}^2 \right\} \leq - \sup_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j} \log(\alpha_{j,0}) \leq \log(\underline{\alpha_{j,0}}) < +\infty.$$

Because the parameter space $\boldsymbol{\Theta}_j$ is compact and $s_{j,t}^*(\boldsymbol{\theta}_j)$ is continuous and bounded away from zero and infinity by assumption, the supremum over $\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j$ is well-defined and finite almost surely. Using Lemma 6.2, we obtain

$$\sup_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j} \left| \mathcal{L}_{\log,j,T_{\text{final}}}(\boldsymbol{\theta}_j) - \widetilde{\mathcal{L}}_{\log,j,T_{\text{final}}}^*(\boldsymbol{\theta}_j) \right| \leq \frac{K}{\underline{\alpha_{j,0}}^2} \sum_{t=T_j}^{T_{\text{final}}} \overline{\alpha_{j,2}}^t r_{j,t-1}^2 + \frac{K}{\underline{\alpha_{j,0}}^2} \sum_{t=T_j}^{T_{\text{final}}} \overline{\alpha_{j,2}}^t < +\infty. \tag{6.10}$$

The finiteness in (6.10) comes from Lemma 6.1 and $\overline{\alpha_{j,2}} < 1$ from Assumption 6.1.

Therefore, $\widehat{\boldsymbol{\theta}}_j^*$ is a consistent estimator of $\boldsymbol{\theta}_j^\star$.

□

Lemma II (Asymptotic equivalence between $\widehat{\boldsymbol{\theta}}_j^*$ and $\widehat{\boldsymbol{\theta}}_j$)**.** *Let us assume that it holds that*

$$\frac{1}{T_{final}} \left| \widetilde{\mathcal{L}}_{\log,T_{final},j}^*(\boldsymbol{\theta}_j) - \mathcal{L}_{\log,T_{final},j}(\boldsymbol{\theta}_j) \right| = o_p\left(\frac{1}{\sqrt{T_{final}}}\right).$$

Under Assumption 6.1, it then holds that

$$\left\| \widehat{\boldsymbol{\theta}}_j^* - \widehat{\boldsymbol{\theta}}_j \right\| = o_p\left(\frac{1}{\sqrt{T_{final}}}\right). \tag{6.11}$$

Proof. To show the asymptotic equivalence between $\widehat{\boldsymbol{\theta}}_j^*$ and $\widehat{\boldsymbol{\theta}}_j$, we apply Theorem 2.8 from Kristensen and Shin. We need to verify the following three conditions:

(i) **Condition 1 of Theorem 2.8**
By compactness of the parameter space $\boldsymbol{\Theta}_j$, clearly Condition 1 is satisfied.

(ii) **Condition 2 of Theorem 2.8**
Condition 2(a) is satisfied because the log-likelihood is smooth for our model. Hence, the log-likelihood is three times differentiable. Condition 2(b) is satisfied by the Law of Large Numbers for ergodic and stationary processes.

(iii) **Condition 3 of Theorem 2.8**
By Equation (6.10), we get

$$\sup_{\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j} \left| \widetilde{\mathcal{L}}^*_{\log, T_{\text{final}} j} - \mathcal{L}_{\log, T_{\text{final}} j} \right| = O_p(1) = o_p\left(\sqrt{T_{\text{final}}}\right).$$

Therefore, it holds that,

$$\frac{1}{T_{\text{final}}} \left| \widetilde{\mathcal{L}}^*_{\log, T_{\text{final}}, j}(\boldsymbol{\theta}_j) - \mathcal{L}_{\log, T_{\text{final}}, j}(\boldsymbol{\theta}_j) \right| = o_p\left(\frac{1}{\sqrt{T_{\text{final}}}}\right)$$

Then, it follows by Theorem 2.8 that the difference between the two estimators satisfies

$$\left\| \widehat{\boldsymbol{\theta}}_j - \widehat{\boldsymbol{\theta}}_j^* \right\| = o_p\left(\frac{1}{\sqrt{T_{\text{final}}}}\right).$$

□

6.4 Asymptotic normality for zero-inflated GARCH-X models

Next, we will prove the asymptotic normality of the quasi-maximum likelihood estimator. We will first give a sketch of the proof, followed by the formulation of the theorem and its proof.

6.4.1 Sketch of the proof

To derive the asymptotic distribution of $\boldsymbol{\theta}_j$, we proceed to analyze the score and hessian of the quasi-log-likelihood. We denote the score vector by $S_{j,n}(\boldsymbol{\theta}_j) \in \mathbb{R}^4$, and the hessian matrix is given by $H_{j,n}(\boldsymbol{\theta}_j) \in \mathbb{R}^{4\times 4}$. For notational convenience, we write the sample size as n, although in our specific setting this corresponds to T_{final}. We define the score vector and the hessian matrix as follows:

$$S_{j,n}(\boldsymbol{\theta}_j) = \frac{\partial \widetilde{\mathcal{L}}_{\log,j,n}(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} = \begin{bmatrix} \frac{\partial \widetilde{\mathcal{L}}_{\log,j,n}}{\partial \alpha_{j,0}} \\ \frac{\partial \widetilde{\mathcal{L}}_{\log,j,n}}{\partial \alpha_{j,1}} \\ \frac{\partial \widetilde{\mathcal{L}}_{\log,j,n}}{\partial \alpha_{j,2}} \\ \frac{\partial \widetilde{\mathcal{L}}_{\log,j,n}}{\partial \alpha_{j,3}} \end{bmatrix}, \quad H_{j,n}(\boldsymbol{\theta}_j) = \frac{\partial^2 \widetilde{\mathcal{L}}_{\log,j,n}(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} = \begin{bmatrix} \frac{\partial^2 \widetilde{\mathcal{L}}_{\log,j,n}}{\partial \alpha_{j,0}^2} & \cdots & \cdots & \frac{\partial^2 \widetilde{\mathcal{L}}_{\log,j,n}}{\partial \alpha_{j,0} \partial \alpha_{j,3}} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial^2 \widetilde{\mathcal{L}}_{\log,j,n}}{\partial \alpha_{j,3} \partial \alpha_{j,0}} & \cdots & \cdots & \frac{\partial^2 \widetilde{\mathcal{L}}_{\log,j,n}}{\partial \alpha_{j,3}^2} \end{bmatrix}.$$

A standard first-order Taylor expansion of the score vector yields

$$S_{j,n}(\widehat{\boldsymbol{\theta}}_j) = S_{j,n}(\boldsymbol{\theta}_j^\star) + H_{j,n}(\overline{\boldsymbol{\theta}}_j)(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^\star),$$

where $\overline{\boldsymbol{\theta}}_j$ lies on the line segment connecting $\widehat{\boldsymbol{\theta}}_j$ and $\boldsymbol{\theta}_j^\star$. Assuming that $\boldsymbol{\theta}_j^\star$ lies in the interior of the parameter space, $\widehat{\boldsymbol{\theta}}_j$ must be an interior solution with probability approaching one. That is, $\mathbb{P}(\{\widehat{\boldsymbol{\theta}}_{j,n} \in \text{int}(\boldsymbol{\Theta}_j)\}) \to 1$ as $n \to \infty$. It remains to derive the limiting distribution of $S_{j,n}(\boldsymbol{\theta}_j^\star)$ and $H_{j,n}(\overline{\boldsymbol{\theta}}_j)$. We can use the Law of Large Numbers for stationary and ergodic sequences and the Central Limit Theorem for martingales to show that

$$\frac{1}{\sqrt{n}} S_{j,n}(\boldsymbol{\theta}_j^\star) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_j), \quad \text{and} \quad -\frac{1}{n} H_{j,n}(\overline{\boldsymbol{\theta}}_j) \xrightarrow{p} \boldsymbol{H}_j > 0, \tag{6.12}$$

where $\boldsymbol{\Sigma}_j, \boldsymbol{H}_j \in \mathbb{R}^{4\times 4}$ are constant, positive definite matrices. This implies that

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^\star\right) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{H}_j^{-1}\boldsymbol{\Sigma}_j\boldsymbol{H}_j^{-1}\right).$$

6.4.2 Theorem and proof

We proceed to verify that Equation (6.12) holds under the following assumption:

Assumption 6.2 (Additional assumptions for asymptotic normality)**.**

(*i*) $\kappa_4 = \mathbb{E}\left[\left(\eta_{j,t}^2 - 1\right)^2 \,\middle|\, \mathcal{F}_{t-1}\right] < \infty$ *is constant.*

(*ii*) $\boldsymbol{\theta}_j^\star$ *is in the interior of* $\boldsymbol{\Theta}_j$*, i.e.* $\boldsymbol{\theta}_j^\star \in \text{int}(\boldsymbol{\Theta}_j)$*.*

Assumption 6.2(i) is used to ensure that the variance of the score exists. While it could be weakened to allow $\mathbb{E}\left[(\eta^2 - 1)^2 \mid \mathcal{F}_{t-1}\right)\Big]$ to be time-varying, as in [21], we retain Assumption 6.2(i) for simplicity. Assumption 6.2(ii) is required to guarantee that $S_n(\widehat{\boldsymbol{\theta}}) = 0$ holds with probability approaching one.

As a first step toward Equation (6.12), the following lemma is useful. Essentially, it shows that the derivatives of the volatility-ratio process $s_t^*(\boldsymbol{\theta})$ are stationary under some conditions.

Lemma 6.4.1. *Under Assumptions 6.1-6.2, the first and second derivatives of the volatility-ratio process* $s_{j,t}^*(\boldsymbol{\theta})$*, given by* $\frac{\partial s_{j,t}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ *and* $\frac{\partial^2 s_{j,t}^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}$*, are stationary and ergodic for all* $\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j$*. Moreover, there exist stationary and ergodic sequences* $B_{k,t} \in \mathcal{F}_{t-1}$*, for* $k = 0, 1, 2$*, which are independent of* $\boldsymbol{\theta}_j$*, such that the following bounds hold for all* $\boldsymbol{\theta}_j$ *in a neighborhood of* $\boldsymbol{\theta}_j^\star$*:*

$$\frac{1}{s_{j,t}^*(\boldsymbol{\theta}_j)} \le B_{0,t}, \quad \left\|\frac{\partial s_{j,t}^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}\right\| \le B_{1,t} \cdot s_{j,t}^*(\boldsymbol{\theta}_j), \quad \left\|\frac{\partial^2 s_{j,t}^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'}\right\| \le B_{2,t} \cdot s_{j,t}^*(\boldsymbol{\theta}_j).$$

where it holds that

$$\mathbb{E}\left[B_{1,t} + B_{2,t}^2\right] < +\infty, \quad \textit{and} \quad \mathbb{E}\left[B_{0,t} \cdot \left(B_{1,t} + B_{2,t}^2\right)\right] < +\infty.$$

The proof of Lemma 6.4.1 can be found in Appendix B, where we have adapted the existing proof to our specific model setting. Lemma 6.4.1 is used to construct suitable bounds for the score and hessian that allow us to use the Central Limit Theorem for martingales and the Law of Large Numbers for stationary and ergodic sequences. We now proceed to establish the second main asymptotic result.

Theorem II (Asymptotic normality quasi-maximum likelihood estimator zero-inflated GARCH-X)**.** *Under Assumptions 6.1 - 6.2, the QMLE* $\widehat{\boldsymbol{\theta}}_j$ *satisfies*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^\star) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{H}_j^{-1}\boldsymbol{\Sigma}_j\boldsymbol{H}_j^{-1}\right),$$

where $\boldsymbol{\Sigma}_j = \kappa_4 \boldsymbol{H}_j$, *and* $\boldsymbol{H}_j = \mathbb{E}\left[\frac{\partial s_{j,t}^*(\boldsymbol{\theta}_j^\star)}{\partial \boldsymbol{\theta}_j}\frac{\partial s_{j,t}^*(\boldsymbol{\theta}_j^\star)}{\partial \boldsymbol{\theta}_j'}\right]$.

Proof. Recall from the result on consistency in Theorem I, that it holds that $\|\widehat{\boldsymbol{\theta}}_j^* - \widehat{\boldsymbol{\theta}}_j\| = o_p\left(\frac{1}{\sqrt{N_j}}\right)$. Therefore, it suffices to analyze $\widehat{\boldsymbol{\theta}}_j^*$. Without loss of generality, we denote the time index by $\mathcal{T}_j = \{1, \ldots, n\}$. First, we compute the score vector:

$$\begin{aligned}
\boldsymbol{S}_{j,n}^*(\boldsymbol{\theta}_j) &= \frac{\partial \widetilde{\mathcal{L}}_{\log,j,n}^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} \\
&= \sum_{t=1}^n I_{j,t} \cdot \frac{\partial \ell_{j,t}^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} \\
&= \sum_{t=1}^n I_{j,t} \cdot \left\{ -\frac{1}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}\frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} - \left(-\frac{1}{\tilde{\sigma}_{j,t}^4(\boldsymbol{\theta}_j)}\right) r_{j,t}^2 \frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} \right\} \\
&= \sum_{t=1}^n I_{j,t} \cdot \left\{ \left(-\frac{1}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)} + \frac{r_{j,t}^2}{\tilde{\sigma}_{j,t}^4(\boldsymbol{\theta}_j)}\right) \frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} \right\} \\
&= \sum_{t=1}^n I_{j,t} \cdot \frac{1}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}\frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}\left(\frac{r_{j,t}^2}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)} - 1\right),
\end{aligned}$$

where the derivatives with respect to $\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)$ can be found in the proof of Lemma 6.4.1. Next, we compute the hessian matrix:

$$\begin{aligned}
\boldsymbol{H}_{j,n}^*(\boldsymbol{\theta}_j) = \frac{\partial^2 \widetilde{\mathcal{L}}_{\log,j,n}^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} &= \sum_{t=1}^n I_{j,t} \cdot \left\{ \frac{\partial}{\partial \boldsymbol{\theta}_j}\left[\left(-\frac{1}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)} + \frac{r_{j,t}^2}{\tilde{\sigma}_{j,t}^4(\boldsymbol{\theta}_j)}\right)\frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}\right]\right\} \\
&= \sum_{t=1}^n I_{j,t} \cdot \left\{ \left(\frac{2}{\tilde{\sigma}_{j,t}^3(\boldsymbol{\theta}_j)} - \frac{4r_{j,t}^2}{\tilde{\sigma}_{j,t}^5(\boldsymbol{\theta}_j)}\right)\frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}\left(\frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}\right)' + \left(-\frac{1}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)} + \frac{r_{j,t}^2}{\tilde{\sigma}_{j,t}^4(\boldsymbol{\theta}_j)}\right)\frac{\partial^2 \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} \right\} \\
&= \sum_{t=1}^n I_{j,t} \cdot \left\{ \left(\frac{1}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}\frac{\partial^2 \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} - \frac{1}{\tilde{\sigma}_{j,t}^4(\boldsymbol{\theta}_j)}\frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}\frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j'}\right)\left(\frac{r_{j,t}^2}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)} - 1\right) \right. \\
&\qquad \left. - \frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}\frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j'}\frac{r_{j,t}^2}{\tilde{\sigma}_{j,t}^6(\boldsymbol{\theta}_j)} \right\} \\
&= \sum_{t=1}^n I_{j,t} \cdot h_{j,t}^*(\boldsymbol{\theta}_j),
\end{aligned}$$

where we applied the product rule, and we denote $h^*(\boldsymbol{\theta}_j)$ as follows:

$$h^*_{j,t}(\boldsymbol{\theta}_j) = \left(\frac{1}{\tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j)} \frac{\partial^2 \tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} - \frac{1}{\tilde{\sigma}^4_{j,t}(\boldsymbol{\theta}_j)} \frac{\partial \tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} \frac{\partial \tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j'} \right) \left(\frac{r^2_{j,t}}{\tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j)} - 1 \right) - \frac{\partial \tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} \frac{\partial \tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j'} \frac{r^2_{j,t}}{\tilde{\sigma}^6_{j,t}(\boldsymbol{\theta}_j)}.$$

To derive the asymptotic distribution of $\widehat{\boldsymbol{\theta}}^*_j$, we expand the score around the true parameter using Taylor expansion:

$$0 = \boldsymbol{S}^*_{j,n}(\widehat{\boldsymbol{\theta}}^*_j) = \boldsymbol{S}^*_{j,n}(\boldsymbol{\theta}^\star_j) + \boldsymbol{H}^*_{j,n}(\bar{\boldsymbol{\theta}}_j)(\widehat{\boldsymbol{\theta}}^*_j - \boldsymbol{\theta}^\star_j),$$

where $\bar{\boldsymbol{\theta}}_j$ lies on the line segment between $\widehat{\boldsymbol{\theta}}^*_j$ and $\boldsymbol{\theta}^\star_j$. Rewriting yields:

$$\begin{aligned} \sqrt{n}\left(\widehat{\boldsymbol{\theta}}^*_j - \boldsymbol{\theta}^\star_j\right) &= \sqrt{n} \cdot \left(- \boldsymbol{H}^*_{j,n}(\bar{\boldsymbol{\theta}}_j)^{-1} \cdot \boldsymbol{S}^*_{j,n}(\boldsymbol{\theta}^\star_j)\right) \\ &= - \left[\frac{1}{n}\boldsymbol{H}^*_{j,n}(\bar{\boldsymbol{\theta}}_j)\right]^{-1} \cdot \frac{1}{\sqrt{n}} \boldsymbol{S}^*_{j,n}(\boldsymbol{\theta}^\star_j). \end{aligned}$$

We will verify the two related convergence results stated in Equation (6.12). We begin by applying the Central Limit Theorem for martingale differences, as stated in Theorem 2.13 and originally denoted in the work of Brown [11], to show that the normalized score converges in distribution to a multivariate normal distribution. To apply the Martingale Central Limit Theorem, we must verify three conditions. Let us first define the sequence

$$M_{j,t} := \frac{\partial s^*_{j,t}(\boldsymbol{\theta}^\star_j)}{\partial \boldsymbol{\theta}_j} (\eta^2_{j,t} - 1). \tag{6.13}$$

By construction, $\boldsymbol{S}^*_{j,n}(\boldsymbol{\theta}^\star_j) = \sum_{t=1}^n M_{j,t}$ which follows from $s^*_{j,t}(\boldsymbol{\theta}^\star_j) = 1$. Clearly, $M_{j,t}$ is a martingale difference by Assumption 6.1(i). Now, it remains to prove the other two conditions according to Theorem 2.13:

(i) **Finite quadratic variation**
The quadratic variation of $\boldsymbol{S}^*_{j,n}(\boldsymbol{\theta}_j)/\sqrt{n}$ is defined as follows:

$$\begin{aligned} \left\langle \boldsymbol{S}^*_{j,n}(\boldsymbol{\theta}^\star_j)/\sqrt{n} \right\rangle &:= \frac{1}{n} \sum_{t=1}^n \mathbb{E}\left[M_{j,t} M_{j,t}' \,\middle|\, \mathcal{F}_{t-1} \right] && (6.14) \\ &= \frac{1}{n} \sum_{t=1}^n \left\{ \frac{\partial s^*_{j,t}(\boldsymbol{\theta}^\star_j)}{\partial \boldsymbol{\theta}_j} \frac{\partial s^*_{j,t}(\boldsymbol{\theta}^\star_j)}{\partial \boldsymbol{\theta}_j'} \right\} \mathbb{E}\left[(\eta^2_{j,t} - 1)^2 \,\middle|\, \mathcal{F}_{t-1} \right] && (6.15) \\ &= \kappa_4 \frac{1}{n} \sum_{t=1}^n \frac{\partial s^*_{j,t}(\boldsymbol{\theta}^\star_j)}{\partial \boldsymbol{\theta}_j} \frac{\partial s^*_{j,t}(\boldsymbol{\theta}^\star_j)}{\partial \boldsymbol{\theta}_j'} && (6.16) \\ &\xrightarrow{p} \kappa_4 \mathbb{E}\left[\frac{\partial s^*_{j,t}(\boldsymbol{\theta}^\star_j)}{\partial \boldsymbol{\theta}_j} \frac{\partial s^*_{j,t}(\boldsymbol{\theta}^\star_j)}{\partial \boldsymbol{\theta}_j'} \right] && (6.17) \\ &= \kappa^4 \boldsymbol{H}_j < +\infty && (6.18) \end{aligned}$$

where we have used $\kappa_4 = \mathbb{E}[(\eta_{j,t}^2 - 1)^2 \mid \mathcal{F}_{t-1}]$ from Assumption 6.2, and Lemma 6.4.1. Lemma 6.4.1 ensures that the derivatives $\frac{\partial s_{j,t}^*}{\partial \boldsymbol{\theta}_j}$ and $\frac{\partial^2 s_{j,t}^*}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'}$ of the volatility-ratio process $s_{j,t}^*(\boldsymbol{\theta}_j)$ are stationary and ergodic for all $\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j$. Moreover, the lemma provides uniform bounds on these derivatives in terms of stationary and ergodic sequences $B_{1,t}, B_{2,t} \in \mathcal{F}_{t-1}$, such that

$$\left\| \frac{\partial s_{j,t}^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} \right\| \leq B_{1,t} \cdot s_{j,t}^*(\boldsymbol{\theta}_j), \quad \left\| \frac{\partial^2 s_{j,t}^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} \right\| \leq B_{2,t} \cdot s_{j,t}^*(\boldsymbol{\theta}_j).$$

The corresponding moment condition $\mathbb{E}[B_{1,t} + B_{2,t}^2] < \infty$ ensures that these derivatives are square-integrable. Consequently, we can apply the ergodic law of large numbers to show the convergence of sample averages to their expectations in probability. This convergence step is written in Equation (6.17).

(ii) **Lindeberg condition**

We need to show that:

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[\|M_{j,t}\|^2 \mathbf{1}_{\{\|M_t\| > c\sqrt{n}\}} \right] \to 0, \quad \text{for all } c > 0.$$

By assumption of stationarity of the process $M_{j,t}$, this condition reduces to showing that:

$$\mathbb{E}\left[\|M_{j,t}\|^2 \mathbf{1}_{\{\|M_{j,t}\| > c\sqrt{n}\}} \right] \to 0 \text{ as } n \to +\infty \tag{6.19}$$

To prove this formally, we apply the Dominated Convergence Theorem. Let $X_n := \|M_{j,t}\|^2 \cdot \mathbf{1}_{\{\|M_{j,t}\| > c\sqrt{n}\}}$. As $c\sqrt{n} \to \infty$ and $\|M_{j,t}\|$ is almost surely finite, the indicator function eventually vanishes. Hence, we have $X_n \to 0$ almost surely. Furthermore, by the square-integrability of $M_{j,t}$, which follows from Assumption 6.1(i) and Lemma 6.4.1, the function $\|M_{j,t}\|^2$ is integrable and thus serves as a valid dominating function. Therefore, the conditions of the Dominated Convergence Theorem are satisfied, and we conclude that

$$\lim_{n \to +\infty} \mathbb{E}\left[\|M_{j,t}\|^2 \cdot \mathbf{1}_{\left\{\|M_{j,t}\| > c\sqrt{n}\right\}} \right] = 0.$$

This verifies that the Lindeberg condition holds.

Therefore, we can conclude that:

$$\frac{1}{\sqrt{n}} \boldsymbol{S}_n^*(\boldsymbol{\theta}_j^\star) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_j), \text{ with } \boldsymbol{\Sigma}_j = \kappa_4 \boldsymbol{H}_j.$$

Next, we turn to the second convergence statement in Equation (6.12), i.e. $-\frac{1}{n} H_n(\overline{\boldsymbol{\theta}}_j) \xrightarrow{p} \boldsymbol{H}_j$. We need to show that the hessian converges and that it is non-singular.

(a) **Uniform convergence**

Let us start by pointing out that the following bound holds for all $\boldsymbol{\theta}_j$ in a neighborhood of $\boldsymbol{\theta}_j^\star$,

and that the right-hand side has finite first moment using Lemma 6.4.1:

$$\|h_t^*(\boldsymbol{\theta}_j)\| \leq \left\{B_{2,t} + B_{1,t}^2\right\}\left\{1 + B_{0,t}\cdot\eta_{k,t}^2\right\} + B_{1,t}^2 B_{0,t}\cdot\eta_{k,t}^2 < +\infty. \tag{6.20}$$

By standard results on uniform convergence for averages of stationary sequences (see Kristensen and Rahbek [42]), it then follows that:

$$\sup_{\|\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^\star\| < \delta} \left\|\boldsymbol{H}_n^*(\boldsymbol{\theta}_j) - \boldsymbol{H}^{\text{stat}}(\boldsymbol{\theta}_j)\right\| \xrightarrow{p} 0,$$

for some $\delta > 0$, where

$$\boldsymbol{H}^{\text{stat}}(\boldsymbol{\theta}_j) = \mathbb{E}\left[h^*_{\boldsymbol{\theta}_{j,t}}(\boldsymbol{\theta}_j)\right].$$

Moreover, the mapping $\boldsymbol{\theta}_j \mapsto H^{\text{stat}}(\boldsymbol{\theta}_j)$ is continuous. Since both $\widehat{\boldsymbol{\theta}}_j^* \xrightarrow{p} \boldsymbol{\theta}_j^\star$ and $\bar{\boldsymbol{\theta}}_j \xrightarrow{p} \boldsymbol{\theta}_j^\star$, it follows that these parameters lie in any arbitrarily small neighborhood of $\boldsymbol{\theta}_j^\star$ with probability approaching one due to the convergence in probability. Hence, we have shown convergence of the hessian to a non-random limit.

(b) **Non-singularity**
Lastly, it remains to show that the hessian $\boldsymbol{H}^{\text{stat}}(\boldsymbol{\theta}_j^\star)$ is non-singular, i.e. invertible. We prove this by contradiction. First, let us define

$$\boldsymbol{\Psi}_{j,t} := \frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j^\star)}{\partial \boldsymbol{\theta}_j} \in \mathbb{R}^4.$$

This process can be written recursively as:

$$\boldsymbol{\Psi}_{j,t} = \beta \boldsymbol{\Psi}_{j,t-1} + \mathbf{W}_{j,t},$$

where $\mathbf{W}_{j,t} := [1, r_{j,t-1}, I_{j,t-1}, \tilde{\sigma}_{j,t-1}^2(\boldsymbol{\theta}_j^\star)]^\top \in \mathbb{R}^4$. Suppose that there exists a non-zero vector $\boldsymbol{\lambda} \in \mathbb{R}^4 \setminus \{\mathbf{0}\}$ and $t \geq 1$ such that $\boldsymbol{\lambda}^\top \boldsymbol{\Psi}_{j,t} = 0$ almost surely. By stationarity of $\boldsymbol{\Psi}_{j,t}$ this must hold for all t. Then, this implies that $\boldsymbol{\lambda}^\top \mathbf{W}_{j,t} = 0$ almost surely for all t. However, this contradicts Assumption 6.1(iv). Therefore, it must hold that

$$\frac{\boldsymbol{\lambda}^\top \boldsymbol{\Psi}_{j,t}}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j^\star)} = 0 \quad \text{almost surely } \forall t \iff \boldsymbol{\lambda} = \mathbf{0}.$$

This gives that the hessian matrix $\boldsymbol{H}^{\text{stat}}(\boldsymbol{\theta}_j^\star) = \mathbb{E}[\boldsymbol{\Psi}_{j,t}\boldsymbol{\Psi}_{j,t}'/\tilde{\sigma}_{j,t}^4(\boldsymbol{\theta}_j^\star)]$ is positive definite because for every nonzero vector $\boldsymbol{\lambda} \in \mathbb{R}^4 \setminus \{\mathbf{0}\}$, it holds that $\boldsymbol{\lambda}^\top \boldsymbol{H}^{\text{stat}}(\boldsymbol{\theta}_j^\star)\boldsymbol{\lambda} > 0$. Then, this implies that $\boldsymbol{H}^{\text{stat}}$ is nonsingular [29].

Therefore, we conclude that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j^\star) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{H}_j^{-1}\boldsymbol{\Sigma}_j\boldsymbol{H}_j^{-1}\right).$$

□

6.5 Real data application

Table 6.1 presents the estimated parameters for the univariate zero-inflated GARCH(1,1)-X model across the seven voluntary carbon credit assets. The results confirm several expected features of financial time series. Most notably, for all assets, the GARCH(1) parameter dominates the ARCH(1) component, indicating strong persistence in volatility, a common feature in asset return modeling. In particular, the Nature Based Offset and Uruguay ARR projects exhibit a high GARCH(1) coefficient around 0.90, suggesting that volatility shocks have long-lasting effects. The intercept terms are generally close to zero, which reflects the low unconditional variance of returns in these markets. The exogenous trading indicator, included as the Xreg term, is non-zero for most assets, albeit small in magnitude. This supports the hypothesis that past trading activity provides some information about volatility. Interestingly, the water filtration and Latam ARR projects show zero Xreg coefficients, suggesting that for these assets, volatility may not be strongly linked to recent trading activity. Overall, the estimates support the modeling approach and show the relevance of accounting for both persistence and trading behavior in capturing volatility dynamics in carbon credit markets.

Table 6.1: *Estimated GARCH(1,1)-X parameters per carbon asset*

Asset	**Intercept** $(\widehat{\alpha}_{j,0})$	**ARCH(1)** $(\widehat{\alpha}_{j,1})$	**GARCH(1)** $(\widehat{\alpha}_{j,2})$	**Xreg** $(\widehat{\alpha}_{j,3})$
Nature Based Offset	0.00000	0.02051	0.94852	0.00019
VCS REDD+ CCB Gold V2021	0.00001	0.34363	0.69034	0.00003
Clean Cookstoves Africa V2021	0.000000	0.18996	0.29075	0.00014
Water Filtration Africa V2021	0.000059	0.09739	0.52569	0.00000
ACR US IFM 2020	0.000001	0.15462	0.38702	0.00007
VCS Uruguay ARR V2021	0.000005	0.08684	0.90496	0.00000
VCS CCB/GS Latam ARR 2021	0.000000	0.52502	0.75340	0.00000

The estimation results in Table 6.2 show that, for all conventional assets, the GARCH(1) parameters dominate the ARCH(1) terms. GARCH(1) values range from 0.68 to 0.93, which indicate strong persistence in volatility and volatility clustering. The intercept terms are small, as expected for asset return series. The estimated parameter for trading activity (Xreg) has a very limited effect across most assets, with values close to zero or exactly zero in the case of Clean Energy ETF and EU ETS. These estimated values are lower than those observed for the carbon credits, except for the volatility index. This could indicate that illiquidity and irregular trading can play a more substantial role in volatility dynamics in those markets. The Volatility Index is the only conventional asset where trading intensity has a relatively larger coefficient (0.00047). Notably, the GARCH(1) coefficients for carbon credit assets were in general lower than those observed for the conventional assets, pointing to less persistent volatility in the carbon market. This difference may reflect the more irregular trading patterns typical of voluntary carbon credits, in which past volatility has less predictive power than in more liquid markets.

Table 6.2: *Estimated GARCH(1,1)-X parameters for conventional assets*

Asset	**Intercept** $(\widehat{\alpha}_{j,0})$	**ARCH(1)** $(\widehat{\alpha}_{j,1})$	**GARCH(1)** $(\widehat{\alpha}_{j,2})$	**Xreg** $(\widehat{\alpha}_{j,3})$
Soybean Futures	0.00000	0.07496	0.89907	0.00001
Volatility Index	0.00045	0.10778	0.72739	0.00047
Clean Energy ETF	0.00001	0.04999	0.92542	0.00000
Natural Gas Futures	0.00001	0.10893	0.89300	0.00001
NASDAQ	0.00003	0.09998	0.67632	0.00002
EU ETS	0.00002	0.09603	0.87539	0.00000

An important condition for the stationarity of GARCH-type models is that the sum of the ARCH(1) and GARCH(1) coefficients remains below one. Tables 6.1 and 6.2 show that this condition is satisfied for all conventional assets except Natural Gas Futures, where the sum slightly exceeds one. Among carbon credit assets, the condition is violated in two cases: VCS REDD+ CCB Gold and VCS CCB/GS Latam ARR. This may indicate potential model instability, possibly due to limited observations or irregular volatility patterns in these illiquid markets.

7

Cross-asset dependence among residuals

The zero-inflated GARCH-X model is presented in Chapter 4 in the multivariate setting and in Chapter 6 in the univariate setting. A distinguishing feature of the model is its ability to capture cross-asset dependence, both through the binary trading indicators $I_{j,t}$ and in the residuals $\eta_{j,t}$ for some asset j. In this chapter, we focus on modeling the dependence structure of the residuals using a copula-GARCH framework. The zero-inflated nature of the data introduces a specific complications in copula-GARCH: when observed returns are exactly zero, residuals are not defined. To map the residuals onto the copula scale, we apply the probability integral transform (PIT), restricting attention to time points where returns are non-zero, i.e. when $I_{j,t} = 1$. This treatment ensures that the zero-inflated structure can be properly handled in dependence modeling. In particular, we discuss two ways of applying the PIT in a bivariate setting. We demonstrate that only the PIT with pairwise conditioning on trading activity guarantees the copula estimator of interest.

7.1 Adaptation of copula techniques to illiquid markets

7.1.1 Cross-asset dependence in residuals

In this section, we study the bivariate joint behavior of the residuals $(\eta_{j,t}, \eta_{k,t})_{t\geq 0}$ for assets $j, k \in \{1, \dots, N\}$. To capture this, we adopt a copula-GARCH framework: the marginal dynamics of $(\eta_{j,t})_{t\in\mathcal{T}_j}$ and $(\eta_{k,t})_{t\in\mathcal{T}_k}$ are handled by the zero-inflated GARCH-X models as defined in (6.1), while their joint dependence structure is described through a copula, estimated in a nonparametric way.

For asset j, the standardized residuals are defined as

$$\eta_{j,t} := \frac{r_{j,t} - \mu_{j,t}}{\sigma_{j,t}} = \frac{r_{j,t}}{\sigma_{j,t}},$$

where $\mu_{j,t} = 0$ is assumed throughout this thesis. Here, the returns are observed, but the conditional volatility is unknown and therefore estimated. For details on the estimation of the conditional volatility, we refer to Chapter 6. To study residual dependence empirically, we estimate the standardized residuals from the fitted GARCH-X model. Formally, the estimated standardized residuals $\widehat{\eta}_{j,t}$ are

defined as

$$\widehat{\eta}_{j,t} := \frac{r_{j,t} - \widehat{\mu}_{j,t}}{\widehat{\sigma}_{j,t}} = \frac{r_{j,t}}{\widehat{\sigma}_{j,t}},$$

where $\widehat{\sigma}_{j,t}$ is the estimated conditional standard deviation. Recall that we used the `garchx` package in `R`, described in Chapter 6, that provides estimates of the conditional volatility $\widehat{\sigma}_{j,t}$ using quasi-maximum likelihood. This package therefore also directly provides the corresponding estimated residuals $\widehat{\eta}_{j,t}$.

To map these estimated residuals to the copula scale, we apply the probability integral transform. Recall from Section 2.3 that the probability integral transform states that if a random variable X has a continuous distribution for which the CDF F_X exists, then the random variable $Y := F_X$ has a standard uniform distribution. Recall that the presence of zero-inflated returns requires particular care. More precisely, residuals are observed only when the corresponding asset is traded, otherwise the residual is undefined. Let us first look at the univariate case, where we look at $\eta_{j,t}$ for some asset j. Then, the probability integral transform gives the pseudo-random variable

$$U_{j,t}^{\{I_{j,t}=1\}} := F_{j|\{I_{j,t}=1\}}(\widehat{\eta}_{j,t}), \tag{7.1}$$

where $F_{j|\{I_{j,t}=1\}}$ denotes the conditional marginal distribution function of the estimated standardized residual $\widehat{\eta}_{j,t}$ for asset j, given that a trade occurred. It is defined as

$$F_{j|\{I_{j,t}=1\}}(x) := \mathbb{P}\Big(\widehat{\eta}_{j,t} \leq x \,\Big|\, I_{j,t} = 1\Big).$$

Note that $U_{j,t}$ is uniformly distributed on $[0,1]$ conditional on asset j being traded at time t, i.e. conditional on $I_{j,t} = 1$. As a result, the PIT is applied only on the days when asset j was actively traded and a nonzero return was observed.

The filtered time index set containing only non-zero returns for asset j is defined as

$$\widetilde{\mathcal{T}_j} := \{t \in \mathcal{T}_j : I_{j,t} = 1\} = \{\widetilde{t}_1^{(j)}, \ldots, \widetilde{t}_{N_j}^{(j)}\}, \tag{7.2}$$

where $\widetilde{t}_1^{(j)}$ denotes the first time point at which asset j is traded and has a non-zero return, and N_j is the total number of such trading days, as defined in Equation (6.2). Next, we estimate the conditional CDF non-parametrically based on the empirical distribution function:

$$\widehat{F}_{j|\{I_{j,t}=1\}}(x) := \frac{1}{|\widetilde{\mathcal{T}_j}|} \sum_{t \in \widetilde{\mathcal{T}_j}} \mathbf{1}\{\widehat{\eta}_{j,t} \leq x\}. \tag{7.3}$$

where $|\widetilde{\mathcal{T}_j}|$ represent the cardinality of the set $\widetilde{\mathcal{T}_j}$. In this sense, by this conditioning, we mean that we restrict attention to the subset of the data for which $I_{j,t} = 1$, effectively omitting all time points where residuals are undefined due to zero returns.

We now turn our attention to the bivariate joint behavior of the residuals $(\eta_{j,t}, \eta_{k,t})_{t \geq 0}$ for assets $j, k \in \{1, \ldots, N\}$. The zero-inflated nature of the data introduces an additional complication: how should the probability integral transform be applied in this bivariate setting, given the presence of undefined values due to the non-trading of one or both assets?

In both approaches considered, the PIT is applied to each series of residuals individually. However, the distinction lies in the conditioning strategy. One option is to apply the PIT separately for each asset using only the time points when that specific asset is traded, and then merge the resulting transformed series by discarding time points where either asset is not traded. Alternatively, the PIT can be applied to the subset of observations where both assets are traded simultaneously, conditioning the transformation jointly on this co-trading activity.

Similar as in the univariate case in Equation (7.2), we introduce a filtered time index set for this bivariate case, corresponding to the time points where both assets are traded, defined as

$$\widetilde{\mathcal{T}_{ij}} := \{t \in \mathcal{T}_i \cap \mathcal{T}_j : I_{i,t} = 1,\ I_{j,t} = 1\}. \tag{7.4}$$

We distinguish the following two possible approaches for constructing the PIT in a multivariate setting:

(i) **Separate PIT approach**
The probability integral transform is constructed separately for each asset by conditioning only on its own trading activity, a procedure we refer to as marginal conditioning. Specifically, we define

$$\widehat{U}_{j,t}^{\{I_{j,t}=1\}} := \widehat{F}_{j|\{I_{j,t}=1\}}(\widehat{\eta}_{j,t}), \tag{7.5}$$

where $\widehat{F}_{j|\{I_{j,t}=1\}}$, as defined in Equation (7.3), denotes the empirical conditional distribution function of the estimated standardized residuals $\widehat{\eta}_{j,t}$, restricted to the subset $\widetilde{\mathcal{T}_j}$. Under this approach, each asset is treated separately: we consider only those time points at which the asset is traded, estimate its conditional distribution over this subset of the data, and subsequently apply the PIT to obtain pseudo-observations on the copula scale. To analyze cross-sectional dependence, the joint distribution is then approximated by selecting only those transformed PIT values $(\widehat{U}_{i,t}, \widehat{U}_{j,t})$ corresponding to time points where both assets i and j are simultaneously traded, i.e. restricted to time time set $\widetilde{\mathcal{T}_{ij}}$ as defined in Equation (7.4).

(ii) **Joint PIT approach**
The alternative approach, referred to as the pairwise conditioning method, incorporates joint trading information by conditioning on the simultaneous trading of two assets. The conditional empirical distribution of $\widehat{\eta}_{j,t}$ is then given by

$$\widehat{F}_{j|\{I_{j,t}=1,\ I_{i,t}=1\}}(x) := \frac{1}{|\widetilde{\mathcal{T}_{ij}}|} \sum_{t \in \widetilde{\mathcal{T}_{ij}}} \mathbf{1}\{\widehat{\eta}_{j,t} \leq x\},$$

and the corresponding (estimated) PIT becomes

$$\widehat{U}_{j,t}^{\{I_{j,t}=1,\ I_{i,t}=1\}} := \widehat{F}_{j|\{I_{j,t}=1,\ I_{i,t}=1\}}(\widehat{\eta}_{j,t}). \tag{7.6}$$

This approach explicitly incorporates the joint trading structure of assets i and j.

The separate approach is a natural and intuitive choice, as it conditions only on an asset's own trading activity. However, the joint (or pairwise) approach is motivated by the observation that $I_{i,t}$ may carry additional information about the distribution of $\widehat{\eta}_{j,t}$ beyond what is conveyed by $I_{j,t}$ alone.

This is particularly relevant in settings with interdependent trading decisions across assets. Since our primary interest is in modeling such cross-asset dependence, conditioning jointly on trading activity before applying the PIT may offer a more informative transformation. The presence of zero returns, giving undefined residuals, necessitates conditioning on time points where both assets are actively traded. The resulting conditional copula of interest is then mathematically expressed as

$$
\begin{aligned}
C_{ij|\{I_{i,t}=1,\, I_{j,t}=1\}}(u_i, u_j) &:= F_{ij|\{I_{i,t}=1,\, I_{j,t}=1\}}\big(F^-_{i|\{I_{i,t}=1\}}(u_i)\,,\, F^-_{j|\{I_{j,t}=1\}}(u_j)\big) \qquad (7.7)\\
&= \mathbb{P}\big(\widehat{\eta}_{i,t} \le F^-_{i|\{I_{i,t}=1\}}(u_i)\,,\, \widehat{\eta}_{j,t} \le F^-_{j|\{I_{j,t}=1\}}(u_j) \,\big|\, I_{i,t}=1,\, I_{j,t}=1\big)\\
&= \mathbb{P}\big(F_{i|\{I_{j,t}=1,\, I_{i,t}=1\}}(\widehat{\eta}_{i,t}) \le u_i\,,\, F_{j|\{I_{j,t}=1,\, I_{i,t}=1\}}(\widehat{\eta}_{j,t}) \le u_j \,\big|\, I_{i,t}=1,\, I_{j,t}=1\big)\\
&= \mathbb{P}\big(U_{i,t}^{\{I_{j,t}=1,\, I_{i,t}=1\}} \le u_i\,,\, U_{j,t}^{\{I_{j,t}=1,\, I_{i,t}=1\}} \le u_j \,\big|\, I_{i,t}=1,\, I_{j,t}=1\big), \qquad (7.8)
\end{aligned}
$$

where the second equality comes from the definition of the joint conditional cumulative distribution function

$$
F_{ij|\{I_{i,t}=1,\, I_{j,t}=1\}}(x_i, x_j) := \mathbb{P}\big(\widehat{\eta}_{i,t} \le x_i\,,\, \widehat{\eta}_{j,t} \le x_j \,\big|\, I_{i,t}=1,\, I_{j,t}=1\big),
$$

and the third equality comes from the inversion of the marginal conditional cumulative distribution functions. The last equality comes from the definition of the variables $U_{i,t}^{\{I_{j,t}=1,\, I_{i,t}=1\}}$ and $U_{j,t}^{\{I_{j,t}=1,\, I_{i,t}=1\}}$. The conditional copula of interest $C_{ij|\{I_{i,t}=1,\, I_{j,t}=1\}}$ links the marginals $F_{i|\{I_{j,t}=1,\, I_{i,t}=1\}}$ and $F_{j|\{I_{j,t}=1,\, I_{i,t}=1\}}$ to form the joint distribution $F_{ij\,|\{I_{j,t}=1,\, I_{i,t}=1\}}$. Note, by Equation (7.8), that this conditional copula is the joint distribution of $\big(U_{i,t}^{\{I_{j,t}=1,\, I_{i,t}=1\}}, U_{j,t}^{\{I_{j,t}=1,\, I_{i,t}=1\}}\big)$ given the event $\{I_{j,t}=1,\, I_{i,t}=1\}$. The construction of a consistent probability integral transform under this conditioning is formally stated and proved in the next subsection.

7.1.2 Asymptotic results of the probability integral transform

We now formally analyze the asymptotic behavior of the copula estimators corresponding to the two PIT approaches defined as in Equations (7.5) and (7.6). The aim is to determine which approach – marginal or joint conditioning – yields a consistent estimator that converges to the conditional copula of interest, as defined in Equation (7.7). We begin by stating and proving a general theorem that provides the asymptotic result of the copula estimator based on the two probability integral transform approaches in a multivariate setting. It is followed by two corollaries that adapt the general theorem to the specific bivariate setting used throughout this thesis.

Let $(\mathbf{X}_t, \mathbf{Z}_t)_{t\in\mathbb{Z}}$ be a strict stationary and ergodic multivariate time series of dimension $(d+p)$, distributed as $(\mathbf{X}, \mathbf{Z})$. Let $\widetilde{\mathcal{T}}_{\mathcal{A}}$ denote the filtered time index set consisting of all time points at which d assets trade simultaneously, defined as

$$
\widetilde{\mathcal{T}}_{\mathcal{A}} := \Big\{ t \in \{T_{\text{start}}, \ldots, T_{\text{final}}\} : \mathbf{Z}_t \in \mathcal{A} \Big\}.
$$

Recall that T_{start} is defined as in Equation (4.1), i.e. $T_{\text{start}} := \max_k T_k$. Here, $N_{\mathcal{A}}$ denotes the total number of observations for which $\mathbf{Z}_i \in \mathcal{A}$, which is defined as

$$
N_{\mathcal{A}} = \sum_{t=T_{\text{start}}}^{T_{\text{final}}} \prod_{k=1}^{d} I_{j,t}.
$$

Without loss of generality, we will work with $T_{\text{start}} = 1$. For a measurable set $\mathcal{A} \subseteq \mathbb{R}^p$, and $k \in \{1, \dots, d\}$, let $F_{k|\mathcal{A}}$ be the conditional CDF of X_k given $\mathbf{Z} \in \mathcal{A}$. For $t \in \mathbb{R}$, let

$$\widehat{F}_{k|\mathcal{A}}(t) := \frac{1}{N_{\mathcal{A}}} \sum_{i=1}^{T_{\text{final}}} \mathbf{1}\{X_{k,i} \leq t, \mathbf{Z}_i \in \mathcal{A}\},$$

For $(d+1)$ measurable sets $\mathcal{A}, \mathcal{A}_1, \dots, \mathcal{A}_d \subseteq \mathbb{R}^p$, let the nonparametric copula estimator be defined as

$$\widehat{C}_{1,\dots,d|\mathcal{A},\mathcal{A}_1,\dots,\mathcal{A}_d}(u_1, \dots, u_d) := \frac{1}{N_{\mathcal{A}}} \sum_{i=1}^{T_{\text{final}}} \mathbf{1}\{\widehat{F}_{1|\mathcal{A}_1}(X_{1,i}) \leq u_1, \dots, \widehat{F}_{d|\mathcal{A}_d}(X_{d,i}) \leq u_d, \mathbf{Z} \in \mathcal{A}\}. \tag{7.9}$$

Theorem III (Consistent conditional copula probability integral transform)**.** *For every $\mathbf{u} \in [0,1]^d$ and every measurable sets $\mathcal{A}, \mathcal{A}_1, \dots, \mathcal{A}_d$, such that all the conditional marginal CDFs $F_{1|\mathcal{A}_1}(X_1), \dots, F_{d|\mathcal{A}_d}(X_d)$ are continuous, we have*

$$\widehat{C}_{1,\dots,d|\mathcal{A},\mathcal{A}_1,\dots,\mathcal{A}_d}(u_1, \dots, u_d) \xrightarrow[T_{final} \to +\infty]{a.s.} \mathbb{P}\Big(F_{1|\mathcal{A}_1}(X_1) \leq u_1, \dots, F_{d|\mathcal{A}_d}(X_d) \leq u_d \mid \mathbf{Z} \in \mathcal{A}\Big) \tag{7.10}$$

$$= C_{\mathbf{X}|\mathbf{Z}\in\mathcal{A}}(\psi_1(u_1), \dots, \psi_d(u_d)), \tag{7.11}$$

where, for $k \in \{1, \dots, d\}$,

$$\psi_k := F_{k|\mathcal{A}} \circ F^-_{k|\mathcal{A}_k}, \tag{7.12}$$

$F^-_{k|\mathcal{A}_k}$ is the inverse of $F_{i|\mathcal{A}_i}$, and $C_{\mathbf{X}|\mathbf{Z}\in\mathcal{A}}$ is the conditional copula of $\mathbf{X}$ given that $\mathbf{Z} \in \mathcal{A}$.

Note that in general, the convergence of the empirical copula $\widehat{C}_{|\mathcal{A},\mathcal{A}_1,\dots,\mathcal{A}_d}(u_1, \dots, u_d)$ is not guaranteed. This comes from the dependence between the pseudo-observations $\widehat{U}_{k,t|\mathcal{A}} = \widehat{F}_{k|\mathcal{A}}(X_{k,t})$, even when the original observations $(\mathbf{X}_t)$ are independent and identically distributed.

Proof. Note that we have the following decomposition

$$\widehat{C}_{\mathbf{X}|\mathcal{A},\mathcal{A}_1,\dots,\mathcal{A}_d}(u_1, \dots, u_d) - \mathbb{P}\Big(U_{1|\mathcal{A}_1} \leq u_1, \dots, U_{d|\mathcal{A}_d} \leq u_d \,\Big|\, \mathbf{Z} \in \mathcal{A}\Big) = K_1 + K_2$$

where,

$$K_1 = \widehat{C}_{\mathbf{X}|\mathcal{A},\mathcal{A}_1,\dots,\mathcal{A}_d}(u_1, \dots, u_d) - \widetilde{C}_{\mathbf{X}|\mathcal{A},\mathcal{A}_1,\dots,\mathcal{A}_d}(u_1, \dots, u_d)$$
$$K_2 = \widetilde{C}_{\mathbf{X}|\mathcal{A},\mathcal{A}_1,\dots,\mathcal{A}_d}(u_1, \dots, u_d) - \mathbb{P}(U_{1|\mathcal{A}_1} \leq u_1, \dots, U_{d|\mathcal{A}_d} \leq u_d \mid \mathbf{Z} \in \mathcal{A}),$$

where $\widetilde{C}_{\mathbf{X}|\mathcal{A},\mathcal{A}_1,\dots,\mathcal{A}_d}(u_1, \dots, u_d)$ is the empirical copula based on the true pseudo-observations $U_{k|\mathcal{A}} := F_{k|\mathcal{A}_k}(\eta_{k,t})$.

Under the assumption that the process $(\mathbf{X}_t, \mathbf{Z}_t)_{t\in\mathbb{Z}}$ is stationary and ergodic, it follows that the transformed process $(\mathbf{U}_t, \mathbf{Z}_t)_{t\in\mathbb{Z}}$ is also stationary and ergodic, where $\mathbf{U}_t := (U_{1,t}, \dots, U_{d,t})$. Moreover, since that $\mathbb{E}[|\mathbf{U}_t|] < \infty$ for all $j \in \{1, \dots, d\}$ and all t, the law of large numbers for stationary and

ergodic sequences implies that

$$\begin{aligned}
\widetilde{C}_{\mathbf{X}|\mathcal{A},\mathcal{A}_1,\ldots,\mathcal{A}_d}(u_1,\ldots,u_d) &= \frac{1}{N_{\mathcal{A}}}\sum_{i=1}^{T_{\text{final}}} \mathbf{1}\{F_{1|\mathcal{A}_1}(X_{1,i}) \leq u_1,\ldots,F_{d|\mathcal{A}_d}(X_{d,i}) \leq u_d, \mathbf{Z} \in \mathcal{A}\} \\
&= \frac{T_{\text{final}}}{N_{\mathcal{A}}} \times \frac{1}{T_{\text{final}}}\sum_{i=1}^{T_{\text{final}}} \mathbf{1}\{F_{1|\mathcal{A}_1}(X_{1,i}) \leq u_1,\ldots,F_{d|\mathcal{A}_d}(X_{d,i}) \leq u_d, \mathbf{Z} \in \mathcal{A}\} \\
&\xrightarrow{a.s.} \frac{1}{\mathbb{P}(\mathbf{Z} \in \mathcal{A})} \times \mathbb{P}(U_{1|\mathcal{A}_1} \leq u_1, \ldots, U_{d|\mathcal{A}_d} \leq u_d, \mathbf{Z} \in \mathcal{A}) \\
&= \mathbb{P}(U_{1|\mathcal{A}_1} \leq u_1, \ldots, U_{d|\mathcal{A}_d} \leq u_d \mid \mathbf{Z} \in \mathcal{A}).
\end{aligned}$$

Therefore, $K_2 \xrightarrow{a.s.} 0$. Now it remains to show that K_1 tends to zero. Note that we have

$$\begin{aligned}
K_1 &= \widehat{C}_{\mathbf{X}|\mathcal{A},\mathcal{A}_1,\ldots,\mathcal{A}_d}(u_1,\ldots,u_d) - \widetilde{C}_{\mathbf{X}|\mathcal{A},\mathcal{A}_1,\ldots,\mathcal{A}_d}(u_1,\ldots,u_d) \\
&= \frac{1}{N_{\mathcal{A}}}\sum_{i=1}^{T_{\text{final}}} \mathbf{1}\{\widehat{F}_{1|\mathcal{A}_1}(X_{1,i}) \leq u_1,\ldots,\widehat{F}_{d|\mathcal{A}_d}(X_{d,i}) \leq u_d, \mathbf{Z} \in \mathcal{A}\} - \mathbf{1}\{F_{1|\mathcal{A}_1}(X_{1,i}) \leq u_1,\ldots,F_{d|\mathcal{A}_d}(X_{d,i}) \leq u_d\} \\
&= \frac{1}{N_{\mathcal{A}}}\sum_{i=1}^{T_{\text{final}}} \mathbf{1}\{\mathbf{Z}_i \in \mathcal{A}\}\,\delta_{i,T_{\text{final}}},
\end{aligned}$$

where

$$\delta_{i,T_{\text{final}}} := \mathbf{1}\{\widehat{F}_{1|\mathcal{A}_1}(X_{1,i}) \leq u_1,\ldots,\widehat{F}_{d|\mathcal{A}_d}(X_{d,i}) \leq u_d\} - \mathbf{1}\{F_{1|\mathcal{A}_1}(X_{1,i}) \leq u_1,\ldots,F_{d|\mathcal{A}_d}(X_{d,i}) \leq u_d\}.$$

By the generalization of the Glivenko-Cantelli theorem from Tucker, that we denoted in Chapter 2 as Theorem 2.11, we know that for any $k \in \{1,\ldots,d\}$,

$$\sup_{x\in(-\infty,+\infty)} \left|\widehat{F}_{k|\mathcal{A}_k}(x) - F_{k|\mathcal{A}_k}(x)\right| \to 0$$

with probability 1. Let E_1 be the event of probability 1 where all these convergences occur. Let E_2 be the event defined by

$$E_2 = \{\forall k = 1,\ldots,d,\ \forall t \geq 1,\ \forall s \neq t,\ X_{k,t} \neq X_{k,s}\}.$$

The event E_2 is the event on which none of the marginals of $\mathbf{X}$ have ties. This event has probability 1 since all marginal distributions are continuous. Let $E = E_1 \cap E_2$. Then E has still probability 1, and we fix $\omega \in E$.

Let $\varepsilon > 0$, and let $n_0 \in \mathbb{N}$ such that for all $T_{\text{final}} \geq n_0$, then for any $k \in \{1,\ldots,d\}$,

$$\sup_{x\in(-\infty,+\infty)} \left|\widehat{F}_{k|\mathcal{A}_k}(x) - F_{k|\mathcal{A}_k}(x)\right| \leq \varepsilon.$$

On this event, we have that for all $t \in \widetilde{\mathcal{T}}_{\mathcal{A}}$, for all $k \in \{1,\ldots,d\}$,

$$\left|\widehat{F}_{k|\mathcal{A}_k}(X_{k,t}) - F_{k|\mathcal{A}_k}(X_{k,t})\right| \leq \varepsilon.$$

Note that $\delta_{i,T_{\text{final}}} = 0$ if

$$\left\{k \in \{1,\dots,d\},\ \widehat{F}_{k|\mathcal{A}_k}(X_{k,t}) \le u_k\right\} = \left\{k \in \{1,\dots,d\},\ F_{k|\mathcal{A}_k}(X_{k,t}) \le u_k\right\}$$

A sufficient condition for this to hold is:

$$\forall k = 1,\dots,d, \quad \left|\widehat{F}_{k|\mathcal{A}_k}(X_{k,t}) - F_{k|\mathcal{A}_k}(X_{k,t})\right| \le \left|u_k - \widehat{F}_{k|\mathcal{A}_k}(X_{k,t})\right|.$$

Indeed, let a, b, c three numbers such that $|a-b| \le |c-a|$. Then if $a \le c$ we know that $b \le c$, since $b = a+(b-a) \le a+(c-a) = c$. And if $a > c$ we know that $b > c$ too, since $b = a+(b-a) > a+(c-a) = c$. We can apply this principle with $a = \widehat{F}_{k|\mathcal{A}_k}(X_{k,t})$, $b = F_{k|\mathcal{A}_k}(X_{k,t})$, and $c = u_k$.

Remark that the numbers $(\widehat{F}_{k|\mathcal{A}_k}(X_{k,t}))_{t\in\widetilde{\mathcal{T}}_{\mathcal{A}}}$ are ranks, i.e. they take the values $1/N_{\mathcal{A}}, 2/N_{\mathcal{A}}, \dots, N_{\mathcal{A}}/N_{\mathcal{A}}$. Here, we have used that $\omega \in E_2$, i.e. we have no ties. So, for all $k = 1,\dots,d$,

$$\begin{aligned}
Card\left(\left\{t \in \widetilde{\mathcal{T}}_{\mathcal{A}} :\ \left|u_k - \widehat{F}_{k|\mathcal{A}_k}(X_{k,t})\right| > \epsilon\right\}\right) &= Card\left(\left\{t \in \widetilde{\mathcal{T}}_{\mathcal{A}} :\ \widehat{F}_{k|\mathcal{A}_k}(X_{k,t}) \notin (u_k-\varepsilon, u_k+\varepsilon)\right\}\right) \\
&= Card\left(\left\{i = 1,\dots,N_{\mathcal{A}} :\ i/N_{\mathcal{A}} \notin (u_k-\varepsilon, u_k+\varepsilon)\right\}\right) \\
&= Card\left(\left\{i = 1,\dots,N_{\mathcal{A}} :\ i \notin (N_{\mathcal{A}}u_k - N_{\mathcal{A}}\varepsilon\ ,\ N_{\mathcal{A}}u_k + N_{\mathcal{A}}\varepsilon)\right\}\right) \\
&= N_{\mathcal{A}} - Card\left(\left\{i = 1,\dots,N_{\mathcal{A}} :\ i \in (N_{\mathcal{A}}u_k - N_{\mathcal{A}}\varepsilon\ ,\ N_{\mathcal{A}}u_k + N_{\mathcal{A}}\varepsilon)\right\}\right) \\
&> N_{\mathcal{A}} - 2\epsilon N_{\mathcal{A}} \\
&= (1-2\epsilon)N_{\mathcal{A}}.
\end{aligned}$$

Therefore, by a union bound, we get that

$$Card\left(\left\{t \in \widetilde{\mathcal{T}}_{\mathcal{A}} :\ \forall k = 1,\dots,d,\ \left|u_k - \widehat{F}_{k|\mathcal{A}_k}(X_{k,t})\right| > \epsilon\right\}\right) > (1-2d\epsilon)N_{\mathcal{A}}.$$

Hence,

$$Card\left(\left\{t \in \widetilde{\mathcal{T}}_{\mathcal{A}} :\ \delta_{i,T_{\text{final}}} = 0\right\}\right) > (1-2d\epsilon)N_{\mathcal{A}}.$$

So

$$\begin{aligned}
Card\left(\left\{t \in \widetilde{\mathcal{T}}_{\mathcal{A}} :\ |\delta_{i,T_{\text{final}}}| = 1\right\}\right) &= N_{\mathcal{A}} - Card\left(\left\{t \in \widetilde{\mathcal{T}}_{\mathcal{A}} :\ \delta_{i,T_{\text{final}}} = 0\right\}\right) \\
&< N_{\mathcal{A}} - (1-2d\epsilon)N_{\mathcal{A}} = 2d \times \epsilon \times N_{\mathcal{A}}.
\end{aligned}$$

By the triangular inequality, we thus obtain

$$|K_1| \le \frac{1}{N_{\mathcal{A}}} \sum_{i=1}^{T_{\text{final}}} |\delta_{i,T_{\text{final}}}| \le \frac{1}{N_{\mathcal{A}}} \times 2d \times \epsilon \times N_{\mathcal{A}} = \varepsilon \times 2d.$$

This can be made arbitrary small by choosing ε small enough. Therefore, $K_1 \to 0$ as $T_{\text{final}} \to +\infty$.

Since this happens on E, which is an event of probability 1, we have proved that $K_1 \xrightarrow{a.s.} 0$.

We now prove the second part of Equation (7.10). We have

$$\begin{aligned}
\lim_{n\to+\infty} \widehat{C}_{|\mathcal{A},\mathcal{A}_1,\ldots,\mathcal{A}_d}(u_1,\ldots,u_d) &= \mathbb{P}\Big(U_{1|\mathcal{A}_1} \le u_1\,,\ldots,\, U_{d|\mathcal{A}_d} \le u_d \,\Big|\, \mathbf{Z}\in\mathcal{A}\Big) \\
&= \mathbb{P}\Big(F_{1|\mathcal{A}_1}(\eta_1) \le u_i\,,\ldots,\, F_{d|\mathcal{A}_d}(\eta_d) \le u_j \,\Big|\, \mathbf{Z}\in\mathcal{A}\Big) \\
&= \mathbb{P}\Big(\eta_1 \le F^-_{1|\mathcal{A}_1}(u_1)\,,\ldots,\, \eta_d \le F^-_{d|\mathcal{A}_d}(u_d) \,\Big|\, \mathbf{Z}\in\mathcal{A}\Big) \\
&= F_{1,\ldots,d|\mathcal{A}}\Big(F^-_{1|\mathcal{A}_1}(u_1)\,,\ldots,\, F^-_{d|\mathcal{A}_d}(u_d)\Big) \\
&= C_{\mathbf{X}|\mathbf{Z}\in\mathcal{A}}(\psi_1(u_1),\ldots,\psi_d(u_d)),
\end{aligned}$$

as claimed.

□

As an immediate application of Theorem III, we apply its general form to the zero-inflated GARCH-X model introduced in Chapter 4. The resulting statement is given in the following corollary.

Corollary 7.1 (Consistency of the copula estimator for pairwise conditioning)**.** *Let $(\eta_{j,t})_{t\in\mathcal{T}_j}$ be standardized residuals with indicators $(I_{j,t})_{t\in\mathcal{T}_j}$ taking values in $\{0,1\}$. Let $\mathcal{T}_{ij}$ denote the pairwise active trading time index set as defined in (7.4). Let $\widehat{C}^{\text{pairwise}}_{ij}$ and $\widehat{C}^{\text{marginal}}_{ij}$ be the empirical copula estimators based on pseudo-observations derived from the continuous conditional marginal distributions $F_{k|\{I_{i,t}=1,I_{j,t}=1\}}$ and $F_{k|\{I_{k,t}=1\}}$ for $k\in\{i,j\}$ respectively. Then, it holds that*

$$\widehat{C}^{\text{pairwise}}_{ij}(u_i,u_j) \xrightarrow[n\to+\infty]{a.s.} C_{ij|\{I_{i,t}=1,I_{j,t}=1\}}(u_i,u_j), \tag{7.13}$$

whereas

$$\begin{aligned}
\widehat{C}^{\text{marginal}}_{ij}(u_i,u_j) \xrightarrow[n\to+\infty]{a.s.} C_{ij|\{I_{i,t}=1,\ I_{j,t}=1\}}\Big(&F_{i|\{I_{i,t}=1,\ I_{j,t}=1\}}(F^-_{i|\{I_{i,t}=1\}}(u_i))\,, \\
&F_{j|\{I_{i,t}=1,\ I_{j,t}=1\}}(F^-_{j|\{I_{j,t}=1\}}(u_j))\Big)
\end{aligned} \tag{7.14}$$

Proof. Remark we have conditioning sets $\{I_{i,t}=I_{j,t}=1\}, \{I_{i,t}=1\}, \{I_{j,t}=1\}$, which are measurable sets. By applying Theorem III, we obtain

$$\begin{aligned}
\lim_{n\to+\infty} \widehat{C}^{\text{pairwise}}_{ij}(u_i,u_j) &= \lim_{n\to+\infty} \frac{1}{|\mathcal{T}_{ij}|} \sum_{t\in\mathcal{T}_{ij}} \mathbf{1}\{\widehat{U}^{\{I_{j,t}=1,\ I_{i,t}=1\}}_{i,t} \le u_i\,,\ \widehat{U}^{\{I_{j,t}=1,\ I_{i,t}=1\}}_{j,t} \le u_j\} \\
&= C_{ij|\{I_{i,t}=1,\ I_{j,t}=1\}}(u_i,u_j)
\end{aligned}$$

Hence, the copula estimator based on pairwise conditioning converges almost surely to the true conditional copula $C_{ij|\{I_{i,t}=1,\ I_{j,t}=1\}}(u_i,u_j)$. Next, we turn to the copula estimator constructed under

marginal conditioning. By applying Theorem III, we obtain

$$\begin{aligned}
\lim_{n\to+\infty} \widehat{C}_{ij}^{\text{marginal}}(u_i,u_j) &= \lim_{n\to+\infty} \frac{1}{|\mathcal{T}_{ij}|} \sum_{t\in\mathcal{T}_{ij}} \mathbf{1}\{\widehat{U}_{i,t}^{\{I_{i,t}=1\}} \le u_i \,,\, \widehat{U}_{j,t}^{\{I_{j,t}=1\}} \le u_j\} \\
&= \mathbb{P}\big(U_{i,t}^{\{I_{i,t}=1\}} \le u_i \,,\, U_{j,t}^{\{I_{j,t}=1\}} \le u_j \,\big|\, I_{j,t}=1,\, I_{i,t}=1\big) \\
&= \mathbb{P}\big(F_{i|\{I_{i,t}=1\}}(\eta_i) \le u_i \,,\, F_{j|\{I_{j,t}=1\}}(\eta_j) \le u_j \,\big|\, I_{j,t}=1,\, I_{i,t}=1\big) \\
&= \mathbb{P}\big(\eta_i \le F^-_{i|\{I_{i,t}=1\}}(u_i) \,,\, \eta_j \le F^-_{j|\{I_{j,t}=1\}}(u_j) \,\big|\, I_{j,t}=1,\, I_{i,t}=1\big) \\
&= F_{ij|\{I_{i,t}=1,\, I_{j,t}=1\}}\big(F^-_{i|\{I_{i,t}=1\}}(u_i) \,,\, F^-_{j|\{I_{j,t}=1\}}(u_j)\big) \\
&= F_{ij|\{I_{i,t}=1,\, I_{j,t}=1\}}\Big(F^-_{i|\{I_{i,t}=1,\, I_{j,t}=1\}}(F_{i|\{I_{i,t}=1,\, I_{j,t}=1\}}(F^-_{i|\{I_{i,t}=1\}}(u_i))) \,, \\
&\qquad\qquad F^-_{j|\{I_{i,t}=1,\, I_{j,t}=1\}}(F_{j|\{I_{i,t}=1,\, I_{j,t}=1\}}(F^-_{j|\{I_{j,t}=1\}}(u_j)))\Big) \qquad (7.15) \\
&= C_{ij|\{I_{i,t}=1,\, I_{j,t}=1\}}\Big(F_{i|\{I_{i,t}=1,\, I_{j,t}=1\}}(F^-_{i|\{I_{i,t}=1\}}(u_i)) \,, \\
&\qquad\qquad F_{j|\{I_{i,t}=1,\, I_{j,t}=1\}}(F^-_{j|\{I_{j,t}=1\}}(u_j))\Big).
\end{aligned}$$

We used in Equation (7.15) the identity $x = F^-_{i|\{I_{i,t}=1,\, I_{j,t}=1\}}(F_{i|\{I_{i,t}=1,\, I_{j,t}=1\}}(x))$. □

Corollary 7.1 shows that the copula estimator for the separate conditioning, $\widehat{C}_{ij}^{\text{marginal}}(u_i,u_j)$, is in general not a consistent estimator of the conditional copula $C_{ij|\{I_{i,t}=1,\, I_{j,t}=1\}}(u_i,u_j)$. This observation motivates the question of whether the estimated copulas under separate or joint conditioning events can coincide under some conditions. Corollary 7.2 answers this question. The proof relies on Remark 7.1, which formalizes an independence property for conditional probabilities and specifies when enlarging the conditioning set leaves the conditional distribution unchanged.

Remark 7.1. *Let A, B, C be events defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose the pair (A, B) is independent of C, then $\mathbb{P}(A \mid B, C) = P(A \mid B)$. This identity follows from the definition of conditional probability and the assumption of independence:*

$$\mathbb{P}(A \mid B, C) = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)} = \frac{\mathbb{P}(A \cap B) \cdot \mathbb{P}(C)}{P(B) \cdot \mathbb{P}(C)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A \mid B),$$

where $\mathbb{P}(C) \ge 0$.

Corollary 7.2. *The probability integral transform using either a separate or a joint conditioning approach, under the same conditions and notations as in Corollary 7.1, have the same limit if and only if $F_{i|\{I_{i,t}=1,\, I_{j,t}=1\}} = F_{i|\{I_{i,t}=1\}}$ and $F_{j|\{I_{i,t}=1,\, I_{j,t}=1\}} = F_{j|\{I_{i,t}=1\}}$.*

Proof. By Remark 7.1, for any $(u_i, u_j) \in [0,1]^2$, if $F_{i|I_{i,t}=1, I_{j,t}=1} = F_{i|I_{i,t}=1}$ and $F_{j|I_{i,t}=1, I_{j,t}=1} = F_{j|I_{j,t}=1}$, then

the copula limits under marginal and joint conditioning clearly coincide:

$$\lim_{n\to+\infty} \widehat{C}_{ij}^{\text{marginal}}(u_i, u_j) = F_{ij|\{I_{i,t}=1,\ I_{j,t}=1\}}\left(F^-_{i|\{I_{i,t}=1\}}(u_i)\ ,\ F^-_{j|\{I_{j,t}=1\}}(u_j)\right) \tag{7.16}$$

$$= F_{ij|\{I_{i,t}=1,\ I_{j,t}=1\}}\left(F^-_{i|\{I_{i,t}=1,\ I_{j,t}=1\}}(u_i)\ ,\ F^-_{j|\{I_{i,t}=1,\ I_{j,t}=1\}}(u_j)\right) \tag{7.17}$$

$$= C_{ij|\{I_{i,t}=1,\ I_{j,t}=1\}}(u_i, u_j) \tag{7.18}$$

$$= C_{ij}^{\text{pairwise}}(u_i, u_j) \tag{7.19}$$

The other direction is proved by contrapositive. We assume that that the conditional CDFs are not equal, i.e. $F_{\{i|I_{i,t}=1,I_{j,t}=1\}} \neq F_{\{i|I_{i,t}=1\}}$ and $F_{\{j|I_{i,t}=1,I_{j,t}=1\}} \neq F_{\{j|I_{j,t}=1\}}$. We show that this implies that the limits of the estimated copulas are not equal. Without loss of generality, choose $x \in \mathbb{R}$ such that that $F_{\{i|I_{i,t}=1,I_{j,t}=1\}}(x) \neq F_{\{i|I_{i,t}=1\}}(x)$. Remember that we assume that $F_{i|\{I_{i,t}=1,I_{j,t}=1\}}$ and $F_{i|\{I_{i,t}=1\}}$ are continuous. Now fix

$$u_i := F_{\{i|I_i=1\}}(x) \in (0,1), \quad u_j := 1.$$

Thus, under the separate conditioning scheme, we have

$$\begin{aligned}
\lim_{n\to+\infty} \widehat{C}_{ij}^{\text{marginal}}(u_i, u_j) &= F_{ij|\{I_{i,t}=I_{j,t}=1\}}\left(F^-_{i|\{I_i=1\}}(u_i), F^-_{j|\{I_j=1\}}(u_j)\right) \\
&= F_{ij|\{I_{i,t}=I_{j,t}=1\}}\left(F^-_{i|\{I_i=1\}}\left(F_{i|\{I_i=1\}}(x)\right), F^-_{j|\{I_j=1\}}(1)\right) \\
&= F_{ij|\{I_{i,t}=I_{j,t}=1\}}\left(x, +\infty\right) \\
&= F_{i|\{I_{i,t}=I_{j,t}=1\}}(x).
\end{aligned}$$

Then, under joint conditioning scheme, we obtain

$$\begin{aligned}
\lim_{n\to+\infty} \widehat{C}_{ij}^{\text{pairwise}}(u_i, u_j) &= F_{ij|\{I_{i,t}=I_{j,t}=1\}}\left(F^-_{i|\{I_{i,t}=I_{j,t}=1\}}(u_i), F^-_{j|\{I_{i,t}=I_{j,t}=1\}}(u_j)\right) \\
&= F_{ij|\{I_{i,t}=I_{j,t}=1\}}\left(F^-_{i|\{I_{i,t}=I_{j,t}=1\}}\left(F_{i|\{I_{i,t}=1\}}(x)\right), F^-_{j|\{I_{i,t}=I_{j,t}=1\}}(1)\right) \\
&= F_{ij|\{I_{i,t}=I_{j,t}=1\}}\left(F^-_{i|\{I_{i,t}=I_{j,t}=1\}}\left(F_{i|\{I_{i,t}=1\}}(x)\right), +\infty\right) \\
&= F_{i|\{I_{i,t}=I_{j,t}=1\}}\left(F^-_{i|\{I_{i,t}=I_{j,t}=1\}}\left(F_{i|\{I_{i,t}=1\}}(x)\right)\right) \\
&= F_{i|\{I_{i,t}=1\}}(x)
\end{aligned}$$

By assumption, we have that $F_{i|\{I_{i,t}=1\}}(x) \neq F_{i|\{I_{i,t}=I_{j,t}=1\}}(x)$. Therefore, the limit of the copula estimators are not the same for the separate and joint conditioning approaches. This completes the proof. □

Corollary 7.2 states the precise condition under which the limit of the copula estimator based on separate conditioning of two indicators is equal to the limit of the estimated copula under joint conditioning: both continuity of all conditional marginals, and marginal invariance should hold. With marginal invariance, we mean $F_{i|\{I_{i,t}=1,I_{j,t}=1\}} = F_{i|\{I_{i,t}=1\}}$ and $F_{j|\{I_{i,t}=1,I_{j,t}=1\}} = F_{j|\{I_{j,t}=1\}}$. Under these assumptions the estimators for the two different conditioning approaches are consistent for the same copula of interest. If marginal invariance fails, the limits differ (a marginal selection effect), and only the joint–conditioning copula estimator remains consistent.

7.2 Real data application: unconditional dependence

7.2.1 Kendall's tau

In the unconditional case, we consider the estimated values of Kendall's tau for each pair of assets. Figure 7.1 shows these estimates, ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no correlation.

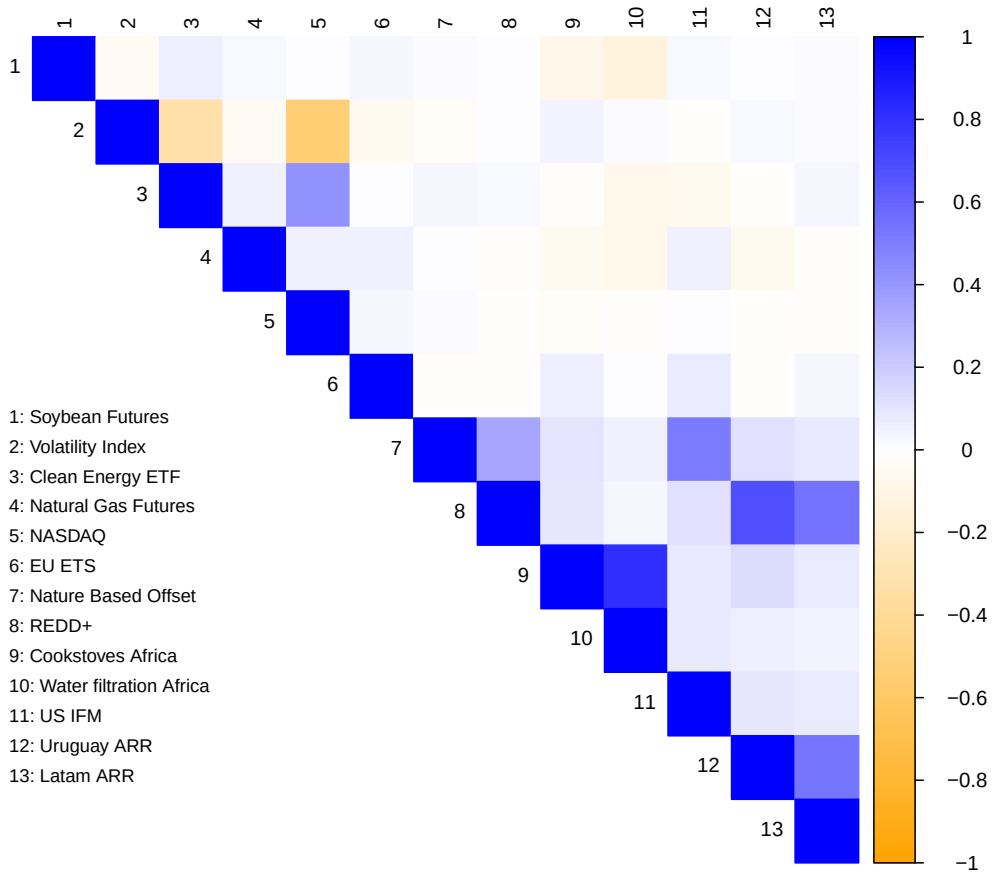


Figure 7.1: *Estimated Kendall's tau for each combination of financial assets*

The results reveal three groups of relationships. First, the group of correlations among conventional, more liquid assets (upper left). We see that the volatility index VIX is negatively correlated with both the Clean Energy ETF and the NASDAQ, while the Clean Energy ETF and the NASDAQ are positively correlated. Previous research has identified a negative correlation between the NASDAQ and the volatility index VIX [33], which is consistent with the results of our study. Second, the group of correlations between carbon assets and conventional assets (upper right). The carbon credits show no association with the conventional markets. Although no causal inference is drawn, this absence of correlation could strengthen the hypothesis that illiquid or early-stage markets tend to behave independently from established, highly liquid markets. Third, the group of correlations among carbon credits (bottom right). Several positive correlations are observed. Nature-Based Offsets are positively correlated with both REDD+ and U.S. IFM credits. REDD+ is also positively correlated with

ARR credits from Uruguay and Latin America, and these two ARR series are themselves strongly correlated, likely due to the Uruguay ARR project being included in the broader Latin America ARR dataset. This suggests that nature-based credits are closely linked. A similar pattern appears for technology-based credits: Cookstoves and Water Filtration projects from Africa display the strongest positive correlation in the entire sample. This may be explained by their common origin and the fact that both are among the most illiquid assets in the dataset, with relatively few observations, which reduces the robustness of the correlation estimates. Across all carbon credits, no negative correlation is observed, not even weak negative dependence. Overall, the results for the correlation among the carbon credits support the view within the market that their prices tend to move together, as valuations frequently draw on the prices of comparable projects.

7.2.2 Copulas

Next, we present scatterplots with overlaid contour lines of the estimated copula densities, obtained using the `kdecop` function from the `kdecopula` package in R. This package provides a nonparametric estimate of a bivariate copula density via kernel smoothing techniques. Based on the correlation plot from Figure 7.1, we focus on asset pairs that exhibit either positive or negative Kendall's tau, in order to examine their dependence structure in more detail.

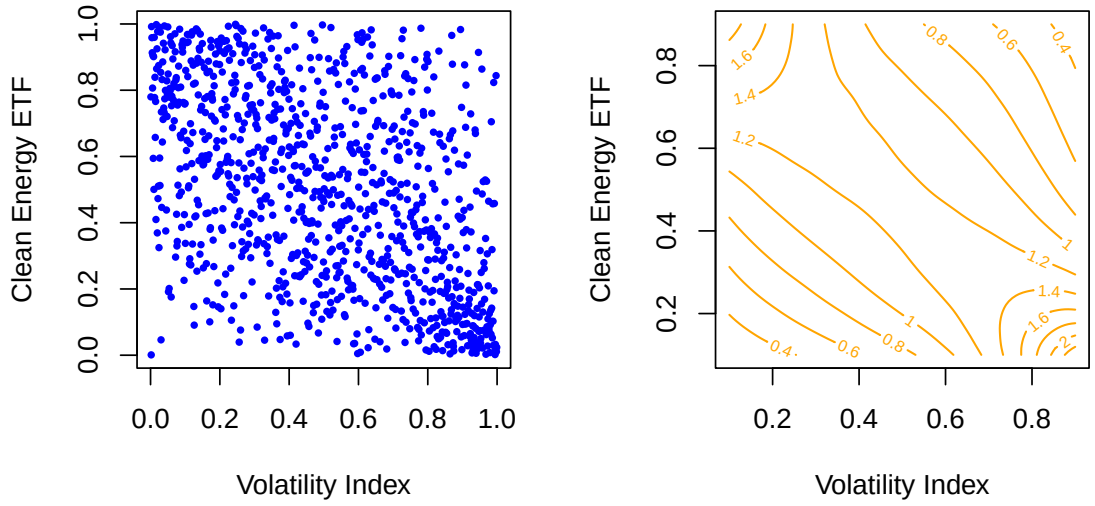
For each selected pair, we plot the scatterplot of the corresponding pseudo-observations (uniformly distributed on $[0, 1]$) alongside the kernel-estimated copula density. This approach reveals features of the dependence structure that cannot be captured by a single summary statistic such as Kendall's τ. While Kendall's tau measures the overall strength and direction of monotonic dependence, it does not provide information on the form or asymmetry of the dependence, particularly in the tails. By applying the probability integral transform, we separate the marginal distributions from the joint dependence, and directly visualize the copula.

The unconditional copula analysis is conducted for the following asset combinations: X.VIX vs ICLN (2–3), X.VIX vs X.IXIC (2–5), ICLN vs X.IXIC (3–5), Nature-Based Offset vs REDD+ CCB (7–8), Nature-Based Offset vs U.S. IFM (7–11), REDD+ CCB vs Uruguay ARR (8–12), REDD+ CCB vs Latam ARR (8–13), Cookstoves Africa vs Water Filtration Africa (9–10), and Uruguay ARR vs Latam ARR (12–13). The pseudo-observation scatterplots with copula density contours, presented in Figures 7.2 - 7.4, reveal distinct patterns of dependence for the different asset pairs.

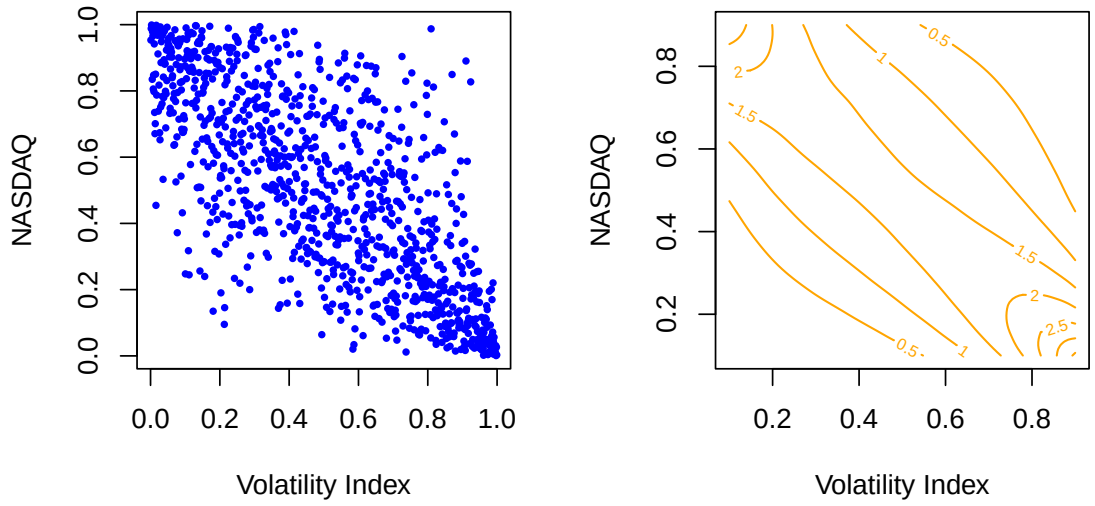
In Figure 7.2, we see that for VIX–ICLN and VIX–NASDAQ, the plots show a clear negative trend: as one variable increases, the other tends to decrease. The inverse relation between the equity index and implied volatility is consistent with the leverage effect. The contour lines are smooth and elongated, indicating a consistent monotonic relationship without strong signs of joint extreme movements, meaning that simultaneous large increases or decreases are relatively rare. In contrast, the NASDAQ–ICLN pair shows a strong positive trend, with the data points and density concentrated along the main diagonal. This suggests that the two assets often move in the same direction, reflecting their similar market drivers. However, the curvature of the contours suggests that this co-movement is stronger during normal market conditions and weaker during extreme events, when the assets may diverge more.

In Figure 7.3, the pseudo-observation scatterplots with estimated copula density contours for a selection of carbon credits show a clear positive trend: higher ranks in one series tend to coincide

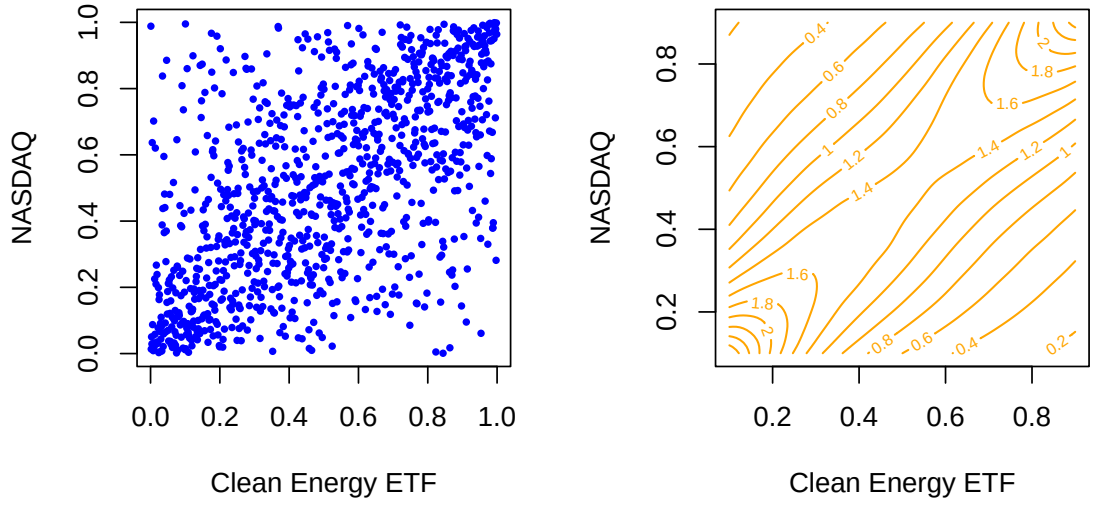
Figure 7.2: *Dependence plots for selected conventional asset pairs*



(**a**) *Scatter plot of the pseudo-observations with copula density contours of the VIX and Clean Energy ETF*

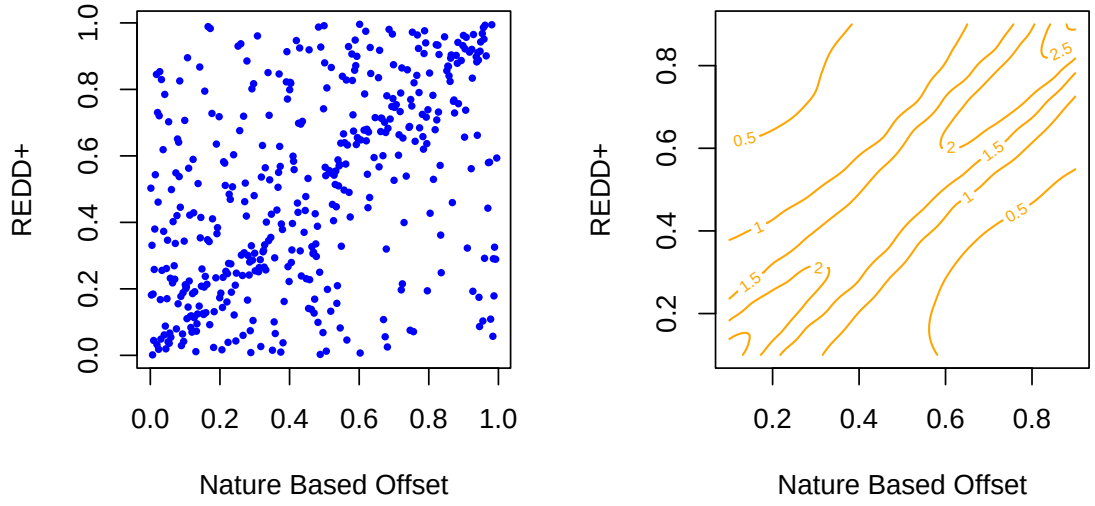


(**b**) *Scatter plot of the pseudo-observations with copula density contours of the VIX and NASDAQ*

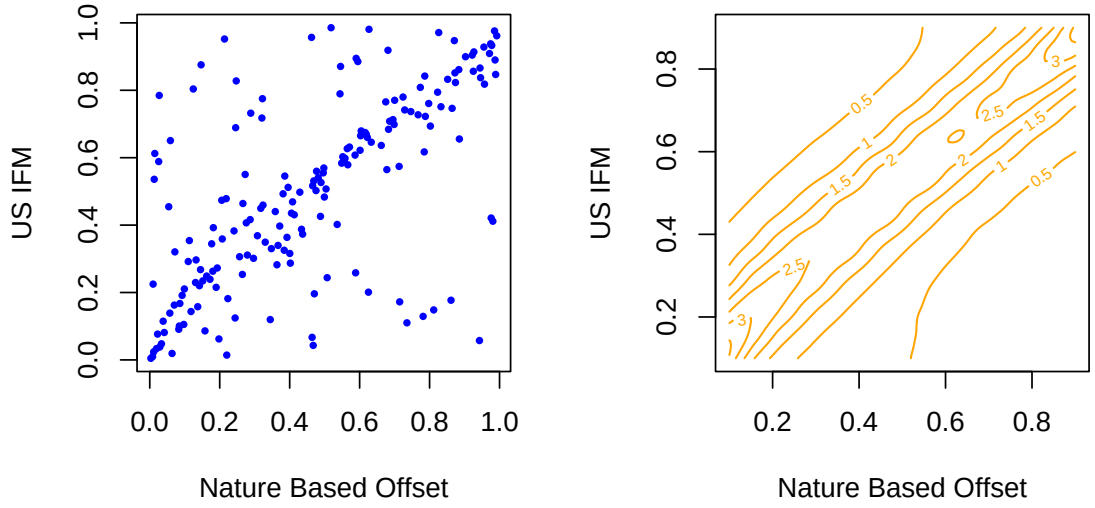


(**c**) *Scatter plot of the pseudo-observations with copula density contours of the NASDAQ and Clean Energy ETF*

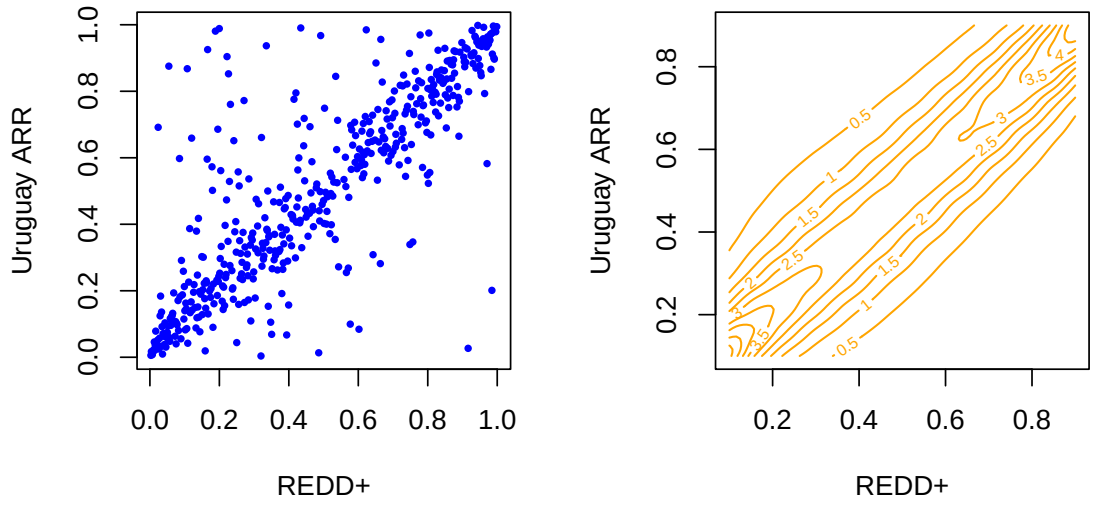
Figure 7.3: *Dependence plots for selected carbon credit asset pairs (I/II)*



(**a**) *Scatter plot of the pseudo-observations with copula density contours of the Nature Based Offset and REDD+ credits*

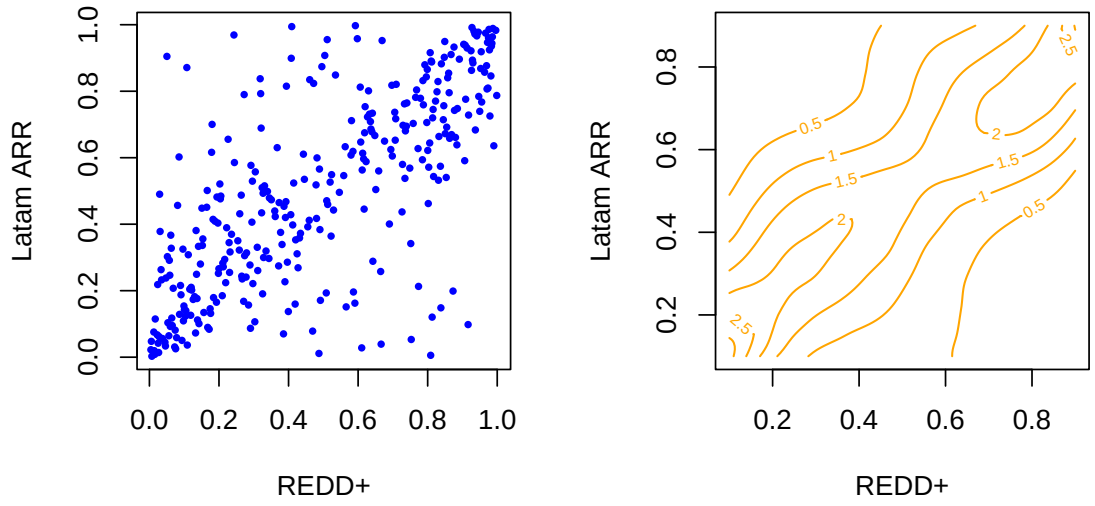


(**b**) *Scatter and contour plot of the pseudo-observations of the Nature Based Offset and US IFM credits*

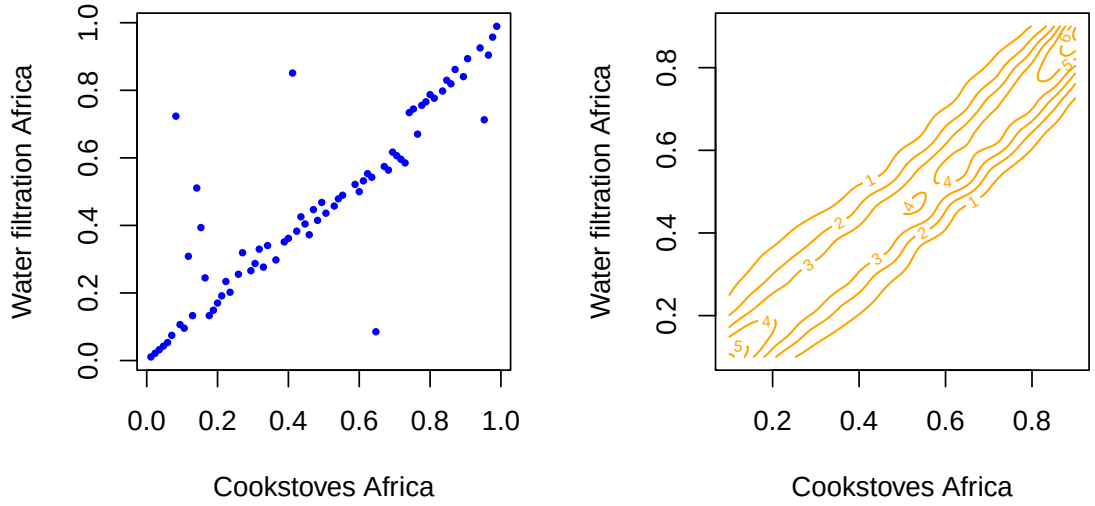


(**c**) *Scatter and contour plot of the pseudo-observations of the Uruguay ARR and REDD+ credits*

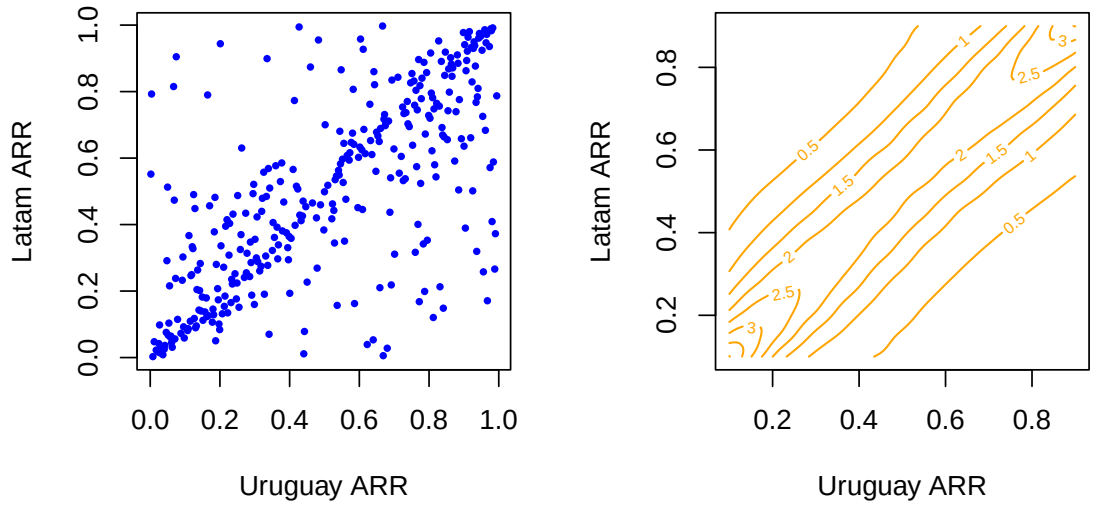
Figure 7.4: *Dependence plots for selected carbon credit asset pairs (II/II)*



(**a**) *Scatter and contour plot of the pseudo-observations of the Latam ARR and REDD+ credits*



(**b**) *Scatter and contour plot of the pseudo-observations of the Water filtration and Cookstoves Africa credits*



(**c**) *Scatter and contour plot of the pseudo-observations of the Latam ARR and Uruguay ARR credits*

with higher ranks in the other. Comparing the three, NBO–REDD+ exhibits a positive but weaker dependence than the other pairs, as seen in the more dispersed cloud of points and less concentrated contours. This indicates moderate concordance. ARR Uruguay–REDD+ shows the tightest diagonal alignment and the narrowest contours, indicating the strongest monotone dependence in this group. NBO–U.S. IFM also shows a positive trend, with contours that show a gentle S-shaped curvature. This suggests that the strength of dependence varies across ranks: stronger around mid-quantiles and weaker toward the extremes. There is no notable clustering of the contours in the corners; at most there is a slight density near the extremes, so any tail dependence appears weak in this sample. An S-shaped ridge means the dependence is state-dependent. This can occur in empirical copulas when common drivers align prices under typical conditions but idiosyncratic factors dominate in extremes. Given the low sample size in combination with the use of kernel smoothing, these features should be read cautiously. The strong linkage between ARR and REDD+ may possibly reflect ARR's emergence being anchored to the historically important REDD+ price signal while market preferences evolve; this interpretation is supported by discussions with Rabobank [55].

In Figure 7.4, the pseudo-observation scatterplots with copula density contours all show a clear positive trend. Comparing the three, the Latam ARR–REDD+ pair exhibits an uneven spread of points, with more clustering in the joint tails than in the middle ranks. This suggests that dependence is particularly strong during extreme events, while in moderate conditions the relationship is less tightly aligned. The Water Filtration Africa–Cookstoves Africa pair shows the strongest and most concentrated positive dependence of the three, with data points and contours forming a dense, narrow diagonal band. This indicates a near one-to-one movement across the rank distribution. Latam ARR–Uruguay ARR also shows tail clustering, though less pronounced than in Latam ARR–REDD+, and the contours here are straighter and less curved, suggesting a more uniform monotone dependence throughout the distribution. The strong co-movement in the tails for Latam ARR–REDD+ and Latam ARR–Uruguay ARR could indicate that prices in these credit types react more similarly during market extremes, potentially due to shared benchmarks or overlapping buyer interest when volatility is high. The extremely tight dependence between Water Filtration and Cookstoves Africa likely reflects shared origination channels and market participants, meaning their prices tend to move together very closely in all market conditions.

Compared with the carbon credit pairs in Figure 7.3, the ARR pairs here (Latam–REDD+, Latam–Uruguay) show stronger evidence of tail-driven dependence, whereas the other nature-based pairs (NBO) tend to have steadier dependence across the distribution with weaker tail effects. The African technology-based credits stand out as the most strongly linked across all carbon credit pairs examined.

7.3 Real data application: conditional dependence

7.3.1 Conditioning on one event

We selected a subset of conditional Kendall's tau plots where the estimated curve is not entirely flat. A non-flat CKT curve indicates the existence conditional dependence between the pair (X_1, X_2) given certain values of a third asset Z. For most combinations, however, the conditional Kendall's tau curve is essentially flat, suggesting that the conditioning variable has little or no effect on the dependence.

We do not analyze these further. This observation also relates to the simplifying assumption, which is discussed in more detail in the Appendix C. Moreover, with 13 assets, there are $\binom{13}{2} = 78$ distinct asset pairs, and for each we can condition on a selection of 10 assets, giving $78 \times 10 = 780$ possible combinations. Listing all of them here would be excessive, so we focus only on the cases where conditional Kendall's tau shows notable variation. Therefore, we distinguish the following four groups where the conditional correlation changes with the conditioning value.

(i) Cookstoves Africa or Water Filtration Africa as one of the assets or the conditioning variable, where non-flat conditional Kendall's tau often appears, likely due to the high proportion of zero returns and small sample size.
(ii) Conditioning on U.S. IFM, which produces several irregular patterns worth further examination.
(iii) Conditioning on EU ETS or Natural Gas, given prior research suggesting a relationship between these assets.
(iv) Pairs already correlated in the unconditional analysis, for which specific conditioning variables are chosen to highlight cases where the conditional Kendall's tau departs from flatness.

The four groups under consideration are displayed in Figures 7.5–7.7. In each case, we employ the two–sided pointwise asymptotic 95% confidence interval for the parameter of interest. More precisely, for every fixed $\mathbf{z}$, the construction satisfies

$$\forall \mathbf{z} \in \mathbf{Z}, \quad \mathbb{P}\big(\tau_{\mathbf{z}}^{\star} \in (a_n(\mathbf{z}), b_n(\mathbf{z}))\big) \longrightarrow 0.95, \quad \text{as } n \to +\infty,$$

where $\tau_{\mathbf{z}}^{\star}$ denotes the true conditional Kendall's τ between $\mathbf{X}_1$ and $\mathbf{X}_2$ given $\mathbf{Z} = \mathbf{z}$, and $(a_n(\mathbf{z}), b_n(\mathbf{z}))$ represents the 95% confidence interval $CI_n(\mathbf{z})$ constructed from a sample of size n, for some $a_n(\mathbf{z}), b_n(\mathbf{z}) \in \mathbb{R}$.

All conditional Kendall's tau estimates were computed using the `CKT.kernel` function from the `CondCopulas` package in `R`. The estimation relies on a nonparametric kernel smoothing approach, where we used the Epanechnikov kernel and a fixed bandwidth of $h = 0.1$ based on visual inspection.

In Figure 7.5, when considering Cookstoves Africa and Water Filtration Africa in the role of asset 1, asset 2, or conditioning asset, several distinct non-flat patterns in the conditional Kendall's tau are observed. In the case of Cookstoves Africa and Soybean Futures given REDD+, Kendall's tau is positive when REDD+ values are low but shifts to negative as REDD+ performance improves. For Cookstoves Africa and the Volatility Index conditioned on the Nature-Based Offset, Kendall's tau becomes increasingly positive as the Nature-Based Offset performs well. The relationship between Cookstoves Africa and Water Filtration Africa given the Clean Energy ETF is generally strongly positive in terms of Kendall's tau but exhibits a pronounced drop at intermediate quantiles of the Clean Energy ETF. Also, for the Volatility Index and NASDAQ conditioned on Cookstoves Africa, Kendall's tau is more negative for intermediate values of Cookstoves Africa and weakens when Cookstoves Africa performs poorly. Across all cases, the confidence intervals are relatively wide, which reflects a substantial sampling variability. This is likely due to the limited number of observations within each conditioning quantile, which reduces the precision of the estimated Kendall's tau and results in greater uncertainty around the point estimates.

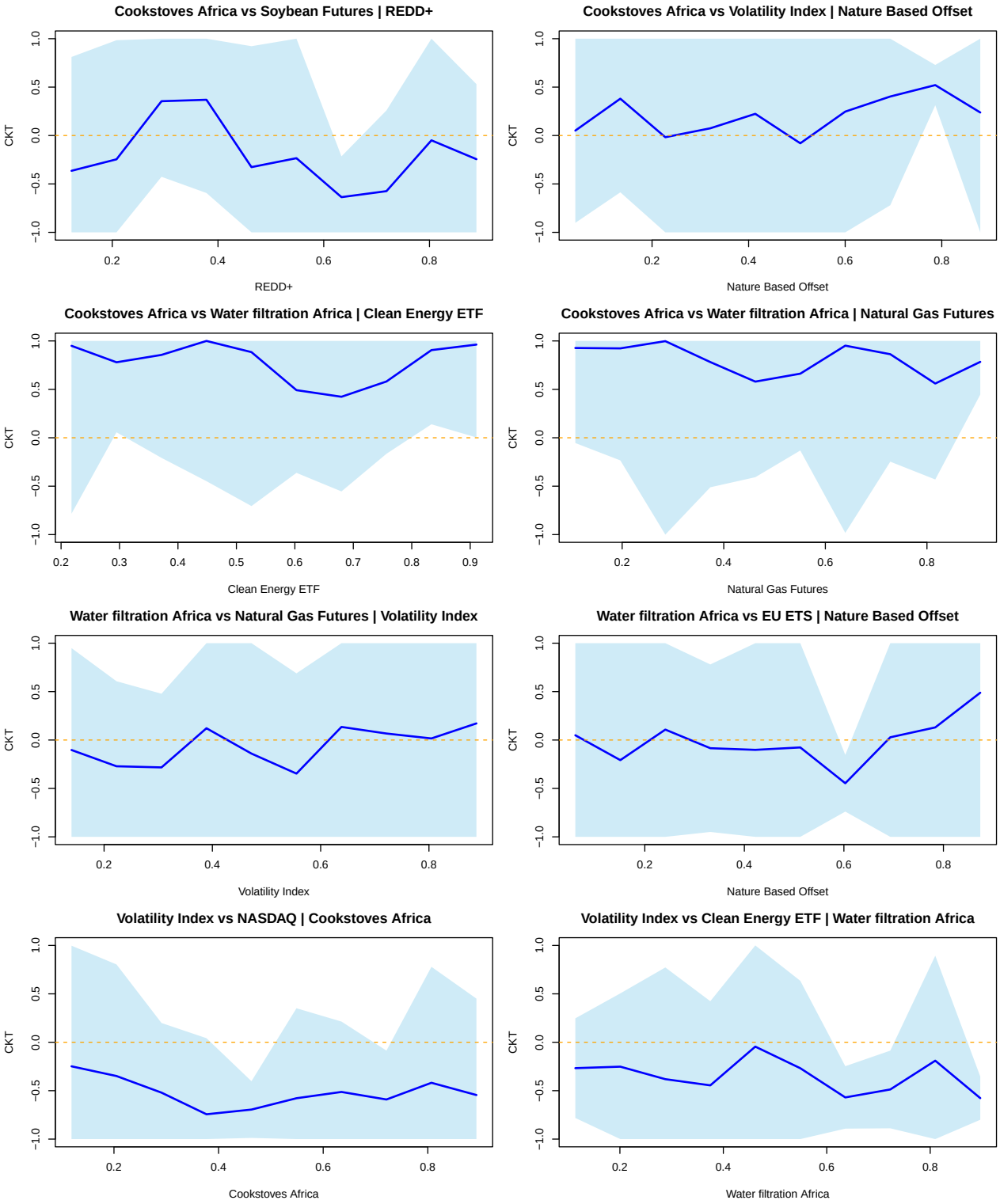


Figure 7.5: *Conditional Kendall's tau estimates for various asset pairs of Cookstoves Africa and Water Filtration Africa, with 95% confidence intervals shown in blue*

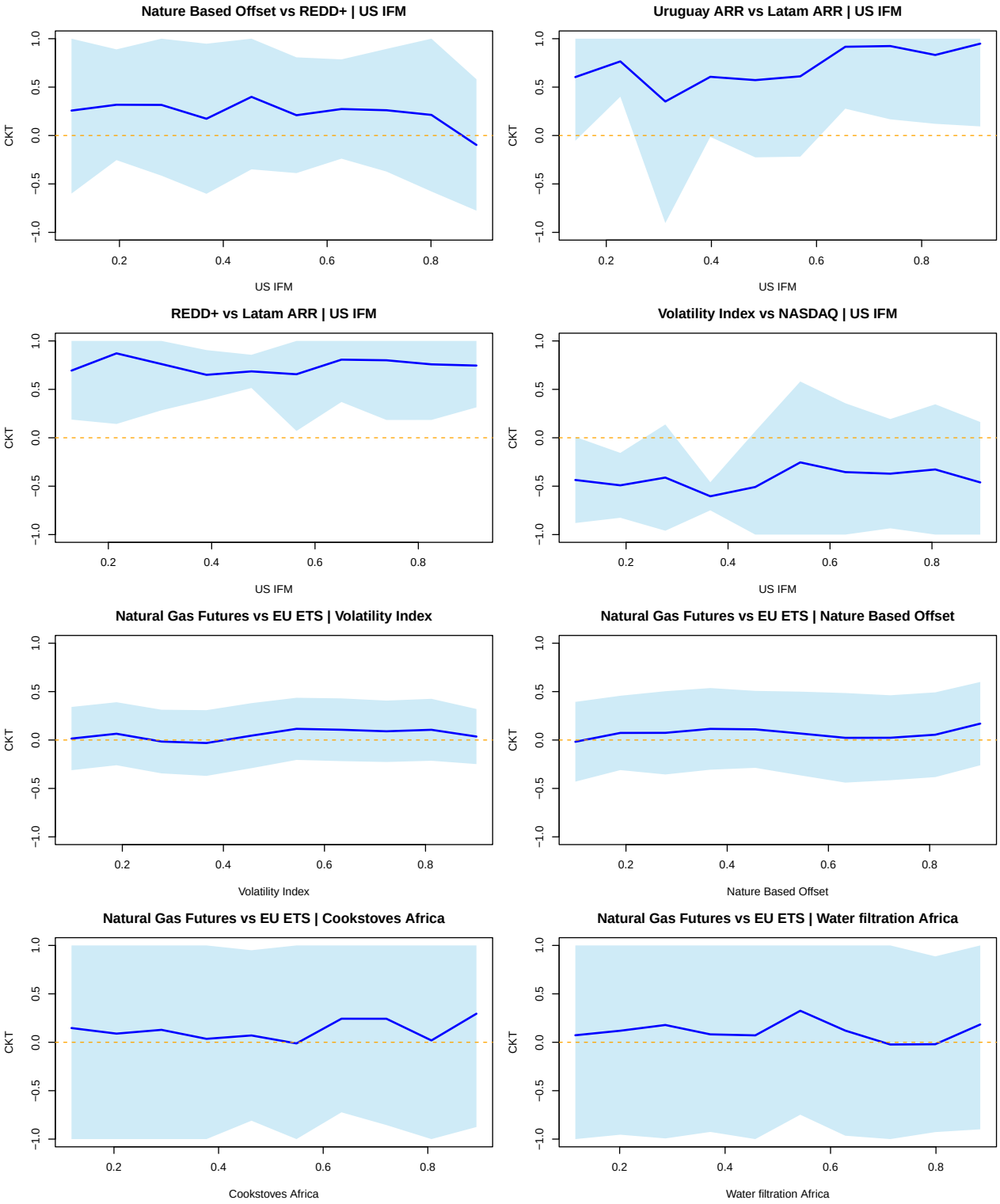


Figure 7.6: *Conditional Kendall's tau for asset pairs involving EU ETS and Natural Gas Futures, as well as pairs conditioned on US IFM credits, with 95% confidence intervals shown in blue*

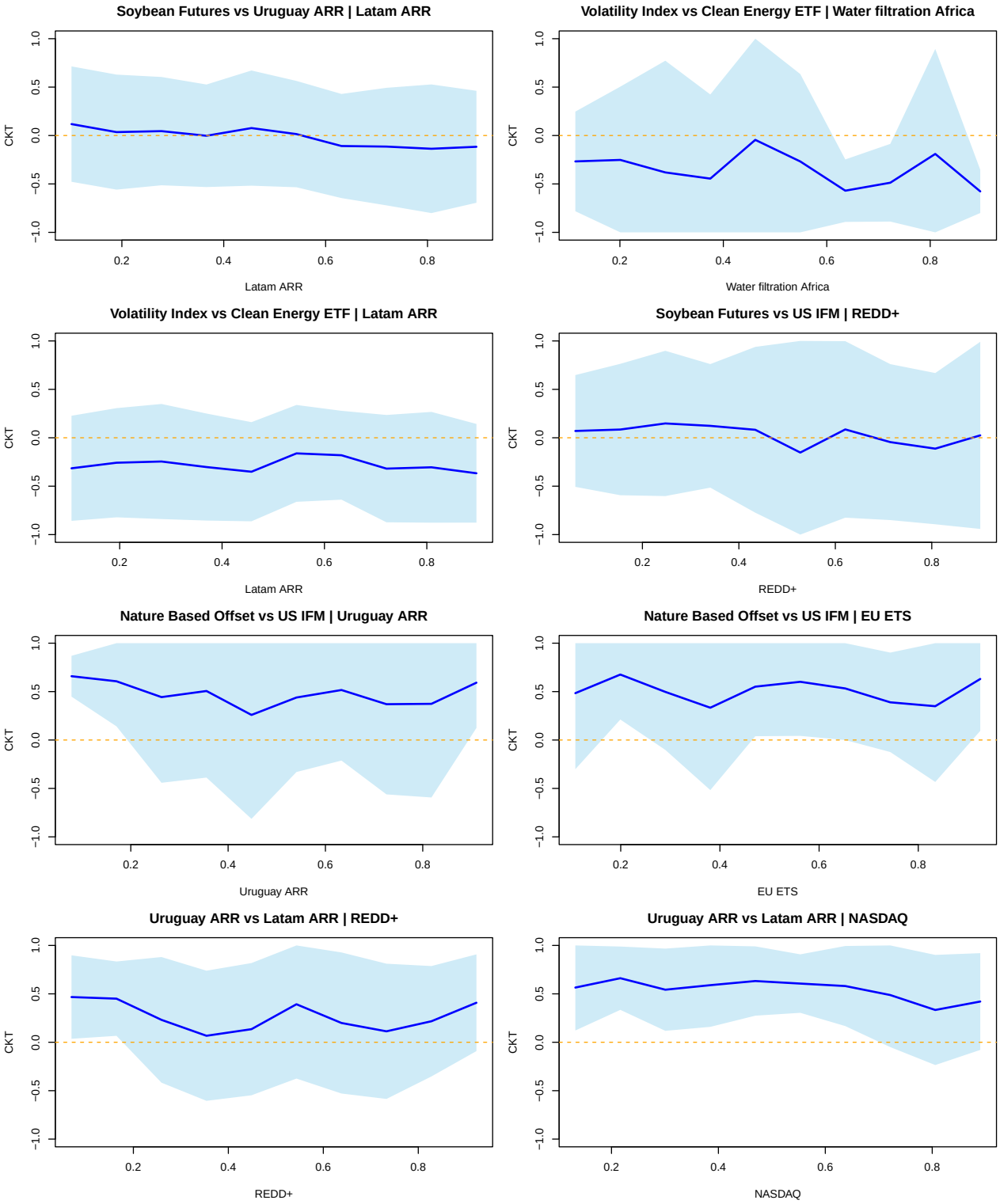


Figure 7.7: *Conditional Kendall's tau for asset pairs reflecting both logically possible correlations and those already present in the unconditional setting, with 95% confidence intervals shown in blue*

In Figure 7.6, when conditioning on US IFM, differences appear particularly in the tails. For instance, the relationship between Nature-Based Offset and REDD+ becomes weaker when US IFM performs well. In contrast, Uruguay ARR and Latam ARR show a strong positive association when US IFM performs well. This suggests that good US IFM outcomes coincide with tighter co-movement between these two series. Additionally, NASDAQ and the Volatility Index have a stronger negative association when US IFM performs poorly. In the case of Natural Gas and EU ETS, the overall association is close to zero; however, when conditioning on Cookstoves Africa or Water Filtration Africa, there is a slight upward spike in the association at mid-range values of the conditioning variable.

In Figure 7.7, we see that despite Soybean Futures, Uruguay ARR, and Latam ARR all belonging to the nature-based category, there is no meaningful correlation observed across the range of the conditioning variable. The relationship between the Volatility Index and Water Filtration Africa shows a tendency toward positive correlation when Water Filtration Africa performs either poorly or well, with little to no correlation in the mid-range. For Nature-Based Offset and US IFM credits conditioned on Uruguay ARR, as well as for Nature-Based Offset and US IFM credits conditioned on EU ETS, stronger associations appear in the distribution tails; however, in the EU ETS case, it remains questionable whether this effect is genuinely driven by EU ETS, especially given the absence of correlation between EU ETS and other selected assets in prior analyses. Lastly, Uruguay ARR and Latam ARR credits become less positively correlated when NASDAQ performs well.

In Figures 7.5–7.7, the estimated conditional Kendall's tau shows some variation across values of the conditioning variable. However, the confidence intervals are wide throughout, which means we cannot tell with high certainty whether the observed fluctuations reflect actual fluctuations of the true conditional Kendall's. In fact, the true dependence could plausibly be constant over the conditioning range. As a result, we cannot draw firm conclusions about the precise effect of the conditioning variable. See for more information on the simplifying assumption Appendix C.

7.3.2 Conditioning on two events

Now we turn into the conditioning on two events, where we have even more possible combinations. Indeed now we have With $n = 13$ assets, there are $\binom{13}{2} \times \binom{11}{2} = 4290$ combinations we can consider. We will show two examples for a nature-based pair of assets and a more tech-based pair of assets to show how conditioning on two events looks like. These will be presented in Figure 7.8.

As in the case of conditioning on one event, the conditional Kendall's tau estimates were obtained using the `CKT.kernel` function from the `CondCopulas` package in `R`. The estimation relies on a non-parametric kernel smoothing approach, where we used the Gaussian kernel and a fixed bandwidth of $h = 0.3$ based on visual inspection.

In Figure 7.8, the upper plot shows the conditional Kendall's tau between Latam ARR and Uruguay ARR, which are known to exhibit positive unconditional dependence (see Figure 7.1). The conditional correlation is particularly strong when both Soybean Futures and US IFM credits perform well. When these conditioning variables are in the mid-range, the correlation remains positive but is weaker compared to the high-performance scenarios. There is also a slight increase in correlation when both US IFM and Soybean Futures perform poorly, indicating that dependence tends to be more pronounced in the tails of the conditioning distribution.

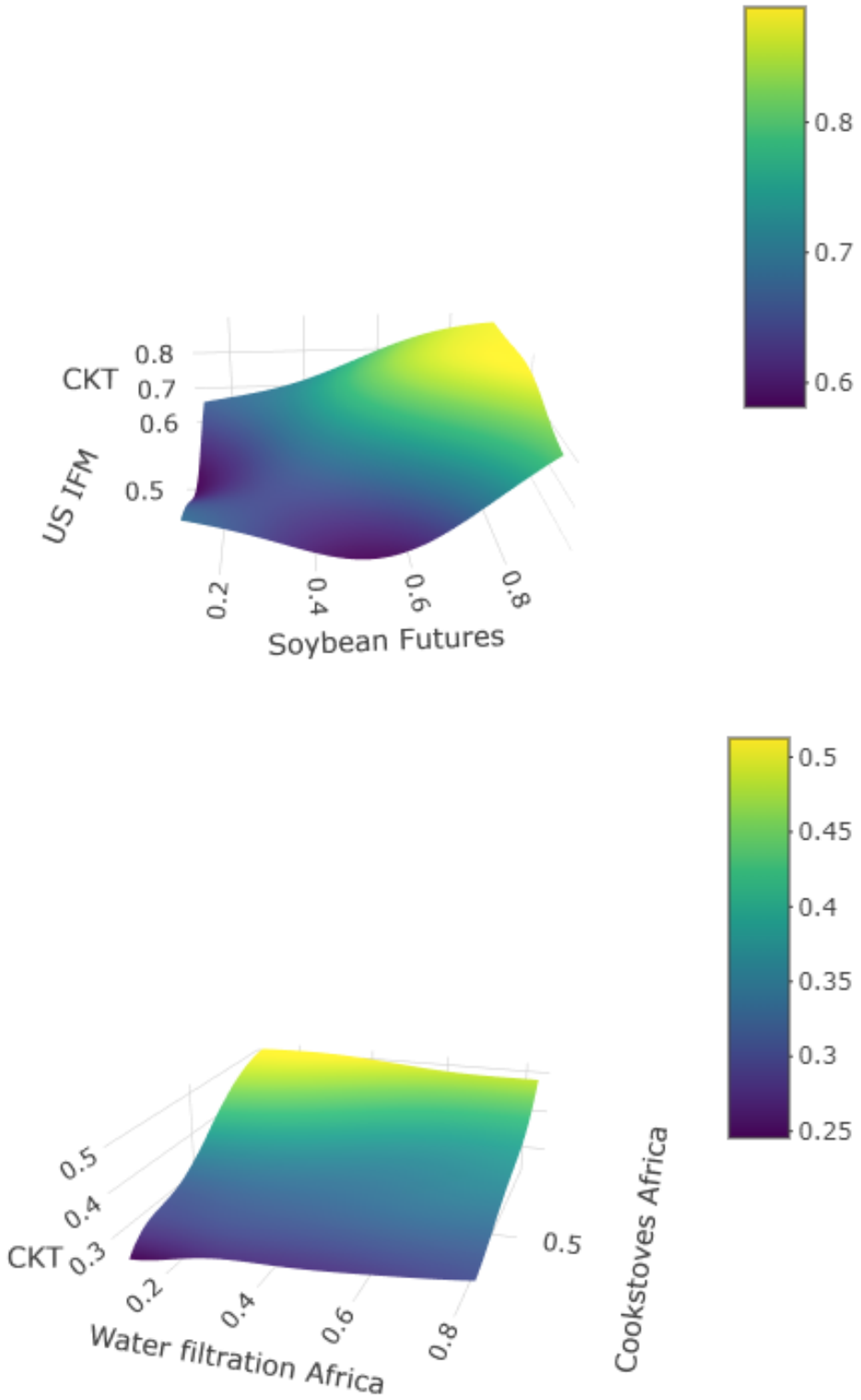


Figure 7.8: *Dependence plots conditioned on two events. The top plot shows the CKT between Latam ARR and Uruguay ARR, conditioned on Soybean Futures and US IFM. The bottom plot takes a more technology-oriented selection, showing the CKT between the Clean Energy ETF and NASDAQ, conditioned on given values for Cookstoves and Water Filtration Africa.*

The plot at the bottom presents the conditional Kendall's tau between the Clean Energy ETF and NASDAQ given Water Filtration Africa and Cookstoves Africa. In the unconditional setting, the Clean Energy ETF and NASDAQ are already positively correlated. The conditional analysis reveals that this positive correlation reaches its highest values when both conditioning assets perform well. Conversely, the correlation is lowest when Cookstoves Africa performs poorly, particularly when both Water Filtration Africa and Cookstoves Africa are at their lowest performance levels. For good-performing Water Filtration Africa combined with poor or medium-performing Cookstoves Africa, the correlation remains relatively low.

In the case of conditioning on two variables, confidence intervals are not provided by the function `CKT.kernel` from the `CondCopulas` package, but they are expected to be wide as well due to the increased dimensionality and reduced actual sample size. As a result, it remains difficult to determine whether the observed variation in conditional dependence reflects a true effect or whether the underlying structure is effectively constant.

8 Conclusion

The goal of this thesis was to explore dependence modeling in illiquid financial markets, with a particular focus on voluntary carbon credit markets. To this end, we introduced a novel zero-inflated GARCH-X model extending the recent work of [25] on zero-inflated GARCH processes. Our main contribution lies in adapting this framework to a multivariate setting, tailoring the zero-inflated GARCH to better capture the characteristics of illiquid financial assets such as carbon credits, and in establishing the consistency and asymptotic normality of its estimators.

In this thesis, we considered seven voluntary carbon credits, covering nature-based and technology-based projects, as well as avoidance and removal types. To investigate potential linkages with more liquid markets, we selected six additional assets, including commodities, an ETF, and major equity indices such as the NASDAQ. Together with Rabobank, we formulated hypotheses on why voluntary carbon credits might be correlated with the selected liquid assets. These hypotheses are discussed in greater detail in Chapter 4. An example of such a hypothesis is the relationship between nature-based credits and agricultural commodities, tech-based credits and clean energy ETFs, and carbon credits and the EU ETS, reflecting potential links between the voluntary and compliance markets. Another example is the relationship between natural gas and the EU ETS, which has been examined in previous studies.

Illiquid financial assets are characterized by a high frequency of zero returns, often referred to in the literature as zero-inflated data. Since voluntary carbon credits exhibit this property, they can be regarded as illiquid assets. To explicitly capture this feature, we introduced binary trading indicators $I_{j,t}$ to model the asset returns. To further model the characteristics of illiquid markets, we incorporated the lagged trading indicator $I_{j,t-1}$ as an exogenous covariate to account for the impact of trading activity on volatility. In addition, we used a time-step specification in which the time index is advanced only when trades occur, thereby focusing on moments of relevant information. These components together define the zero-inflated GARCH-X model presented in Chapter 4. Model parameters were estimated using quasi-maximum likelihood, a standard approach for GARCH-X processes. We proved that the quasi-maximum likelihood estimator is both consistent and asymptotically normal. To our knowledge, these theoretical properties, established in a multivariate zero-inflated setting such as ours, are not present in the existing literature.

The multivariate specification incorporates two different types of dependence; the multivariate extension combines trading indicators to capture cross-dependence in trading activity with copula-

based structures that model the joint dependence among asset returns. Importantly, given the limited sample size, our analysis focused on bivariate copulas rather than higher-dimensional dependence structures. Larger-scale dependence modeling could be pursued when more transaction-level carbon credit data becomes available. First, cross-dependence in trading activity is modeled through binary trading indicators using Markov networks. Second, cross-dependence across assets is modeled using a copula-GARCH approach applied to the residuals. Since zero-inflated returns do not produce residuals $\eta_{j,t}$, these cases were treated as undefined values. To apply copula techniques in the presence of zero-inflated data, we did two things. First, for each asset pair, we removed all observations where either residual was undefined. Second, we introduced a joint probability integral transform approach. In this construction, the univariate marginals are defined conditional on the simultaneous trading activity of each asset, rather than conditioning only on each asset's own trading activity. In Chapter 7, we proved that this joint conditioning PIT approach yields a consistent estimator for the copula of the asset pair under study. Furthermore, we estimated Kendall's tau both unconditionally and conditionally using kernel-based methods.

The empirical findings highlight several important results. Most notably, we found no significant dependence between voluntary carbon credits and the selected liquid financial assets. Although these assets were selected based on plausible economic hypotheses, the data did not support them. These results suggest that illiquid and mature markets may operate largely independently, although identifying the exact cause is challenging. The absence of correlation between global natural gas prices and the EU ETS may be due to our use of global rather than European-specific prices, given that the EU ETS operates within the European market. In contrast, we observed positive correlations within the carbon credit market itself, particularly among nature-based credits. This clustering likely reflects the absence of standardized pricing mechanisms in voluntary markets, where project developers, brokers, and other participants tend to reference each other's prices. Such peer-driven pricing behavior naturally gives rise to endogenous correlation structures. Importantly, the absence of dependence with the selected conventional financial assets studied in this thesis does not imply that no conventional asset is linked to carbon credits. Our framework can incorporate any chosen asset, allowing future research, for example by Rabobank, to investigate potential dependence between other conventional financial assets and carbon credits.

The limitations of our study are primarily data-driven. The reliance on bivariate copulas was necessitated by the relatively short time series and the low frequency of trades. More granular transaction-level data would enable higher-dimensional dependence modeling and potentially more robust estimation of time-varying structures. Although we have used the best available data to our knowledge, improvements in data quality and access to longer historical series are essential for drawing stronger conclusions.

A further important observation concerns the event-driven nature of illiquid markets. As illustrated by the recent Guardian article [34] that significantly affected the reputation of carbon credits, critical media coverage can sharply influence both demand and pricing. Modeling temporal shifts in dependence offers a promising direction for quantifying the impact of specific events. This can be done, for example, by comparing correlation structures before and after major events. This is particularly relevant in volatile, thinly traded markets where exogenous shocks may drive endogenous pricing dynamics.

So, this thesis extends the current literature on modeling illiquid financial assets by proposing a novel multivariate zero-inflated GARCH-X framework with proven asymptotic properties, specifically designed for markets with sparse trading activity. Methodologically, the framework integrates Markov network modeling of trading activity with copula-GARCH modeling of returns, and adapts copula techniques to zero-inflated data through a consistent joint PIT approach. Empirically, it provides new evidence on the independence of voluntary carbon markets from conventional financial assets and the internal correlation structure of these credits. Future research could extend this work by incorporating non-stationary covariates, modeling higher-dimensional dependence, and quantifying the effect of discrete events on market correlations. These extensions would further improve our understanding of how illiquid markets form prices and respond to information shocks.

References

[1] ABN AMRO. *Carbon Market Strategist: Prices Drop Following Unprecedented Tariffs*. Accessed April 24, 2025. ABN AMRO Research. 2024. URL: https://www.abnamro.com/research/en/our-research/carbon-market-strategist-prices-drop-following-unprecedented-tariffs.

[2] Riadh Aloui, Mohamed Safouane Ben Aïssa, and Duc Khuong Nguyen. "Conditional dependence structure between oil prices and exchange rates: A copula-GARCH approach". In: *Journal of International Money and Finance* 32 (2013), pp. 719–738. ISSN: 0261-5606. DOI: https://doi.org/10.1016/j.jimonfin.2012.06.006. URL: https://www.sciencedirect.com/science/article/pii/S0261560612001441.

[3] Autoriteit Financiële Markten (AFM). *Voluntary Carbon Markets: A Supervisory View*. Tech. rep. Occasional Paper on the functioning, risks, and supervisory considerations of voluntary carbon markets. Dutch Authority for the Financial Markets, 2023. URL: https://www.afm.nl/en.

[4] Alberto Ballesteros-Rodríguez, Juan De-Lucio, and Miguel-Ángel Sicilia. "Tokenized Carbon Credits in Voluntary Carbon Markets: The Case of KlimaDAO". In: *Frontiers in Blockchain* 7 (2024), p. 1474540. DOI: 10.3389/fbloc.2024.1474540. URL: https://www.frontiersin.org/articles/10.3389/fbloc.2024.1474540/full.

[5] Andrey Berdichevskiy et al. "Unlocking the Full Potential in Carbon Markets: Pathways to Growth and Sustainability in Asia". In: *Boston Consulting Group* (Feb. 2025). URL: https://www.bcg.com/publications/2025/southeast-asia-unlocking-the-full-potential-in-carbon-markets.

[6] I. Berkes, L. Horváth, and P. Kokoszka. "GARCH processes: Structure and estimation". In: *Bernoulli* 9.2 (2003), pp. 201–227.

[7] Fetsje Bijma, Marianne Jonker, and Aad van der Vaart. *An Introduction to Mathematical Statistics*. Amsterdam: Amsterdam University Press, 2017. ISBN: 9789048536115. DOI: doi:10.1515/9789048536115.

[8] Patrick Billingsley. *Probability and Measure*. 3rd. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons, 1995. ISBN: 978-0-471-00710-4.

[9] Christopher Blaufelder et al. *A Blueprint for Scaling Voluntary Carbon Markets to Meet the Climate Challenge*. Report. McKinsey & Company, Jan. 2021. URL: https://www.mckinsey.com/business-functions/sustainability/our-insights/a-blueprint-for-scaling-voluntary-carbon-markets-to-meet-the-climate-challenge.

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

[11] B. M. Brown. "Martingale central limit theorems". In: *Annals of Mathematical Statistics* 42 (1971), pp. 59–66.

[12] Yen-Chi Chen. *STAT 535: Statistical Machine Learning, Lecture 11: Graphs and Networks*. https://faculty.washington.edu/yenchic/19A_stat535/Lec11_UG.pdf. Lecture notes, University of Washington, Autumn 2019. 2019.

[13] Hung Chu. "Modelling the dependence structure of financial time series: considering copula-MGARCH models". Student ID S2348594; supervisors Prof. dr. E.C. Wit and Dr. M.A. Grzegorczyk. MSc thesis. Groningen, Netherlands: University of Groningen, Faculty of Science and Engineering, June 2017.

[14] Marc N. Conte and Matthew J. Kotchen. "Explaining the Price of Voluntary Carbon Offsets". In: *Climate Change Economics* 1.2 (2010). Accessed April 24, 2025, pp. 93–111. URL: https://resources.environment.yale.edu/kotchen/pubs/explain.pdf.

[15] Jake R Conway, Alexander Lex, and Nils Gehlenborg. "UpSetR: An R Package for the Visualization of Intersecting Sets and their Properties". In: *Bioinformatics* 33.18 (2017), pp. 2938–2940. DOI: 10.1093/bioinformatics/btx364.

[16] Claudia Czado. *Analyzing Dependent Data with Vine Copulas: A Practical Guide With R*. Vol. 222. Lecture Notes in Statistics. Garching, Germany: Springer Nature Switzerland AG, 2019. ISBN: 978-3-030-13784-7.

[17] Alexis Derumigny and Jean-David Fermanian. "About tests of the "simplifying" assumption for conditional copulas". In: *Dependence Modeling* 5 (2017), pp. 154–197.

[18] Alexis Derumigny and Jean-David Fermanian. *On kernel-based estimation of conditional Kendall's tau: finite-distance bounds and asymptotic behavior*. 2019. arXiv: 1810.06234 [math.ST]. URL: https://arxiv.org/abs/1810.06234.

[19] Alexis Derumigny and Johannes Schmidt-Hieber. "On Lower Bounds for the Bias-Variance Trade-off". In: *The Annals of Statistics* 51.4 (2023), pp. 1510–1533. DOI: 10.1214/23-AOS2282.

[20] Alexis Derumigny et al. *CondCopulas: Estimation and Inference for Conditional Copula Models*. R package version 0.1.4.1. 2024. DOI: 10.32614/CRAN.package.CondCopulas. URL: https://CRAN.R-project.org/package=CondCopulas.

[21] Juan M. Escanciano. "Quasi-maximum likelihood estimation of semi-strong GARCH models". In: *Econometric Theory* 25.3 (2009), pp. 561–570.

[22] R. A. Fisher. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd, 1934.

[23] Ildikó Flesch and Peter J. Lucas. "Markov Equivalence in Bayesian Networks". In: *Advances in Probabilistic Graphical Models*. Vol. 213. Studies in Fuzziness and Soft Computing. Springer, 2007, pp. 3–38. DOI: 10.1007/978-3-540-68996-6_1.

[24] C. Francq and J.M. Zakoian. *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, 2019. ISBN: 9781119313489.

[25] Christian Francq and Genaro Sucarrat. "Volatility Estimation When the Zero-Process is Nonstationary". In: *Journal of Business & Economic Statistics* 41.1 (2023), pp. 53–66. DOI: 10.1080/07350015.2021.1999821. URL: https://doi.org/10.1080/07350015.2021.1999821.

[26] Christian Francq and Le Quyen Thieu. *QML inference for volatility models with covariates*. MPRA Paper No. 63198. Mar. 2015. URL: https://mpra.ub.uni-muenchen.de/63198/.

[27] Christian Francq and jean-michel Zakoian. "Estimating GARCH Models by Quasi-Maximum Likelihood". In: (July 2010), pp. 141–184. DOI: 10.1002/9780470670057.ch7.

[28] Christian Franq and Jean-Michel Zakoian. "Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes". In: *Bernoulli* 10.4 (2004), pp. 605–637.

[29] Stephen Friedberg. *Linear Algebra*. Pearson Education, 2013. ISBN: 9781292026503.

[30] Julio Friedmann and Matthew D. Potts. *Carbon removal, reduction, and avoidance credits explained*. Accessed April 24, 2025. Carbon Direct. 2023. URL: https://www.carbondirect.com/insights/carbon-removal-reduction-and-avoidance-credits-explained.

[31] Jean-François Le Gall. *Brownian Motion, Martingales, and Stochastic Calculus*. Vol. 274. Graduate Texts in Mathematics. Springer International Publishing, 2016. ISBN: 978-3-319-31088-6. DOI: 10.1007/978-3-319-31089-3.

[32] E. García-Portugués. *Notes for Predictive Modeling*. https://bookdown.org/egarpor/PM-UC3M/. Version 5.9.8. 2022.

[33] Pierre Giot. "Relationships Between Implied Volatility Indexes and Stock Index Returns". In: *Journal of Portfolio Management - J PORTFOLIO MANAGE* 31 (Mar. 2005), pp. 92–100. DOI: 10.3905/jpm.2005.500363.

[34] Patrick Greenfield. "Revealed: More than 90% of rainforest carbon offsets by biggest certifier are worthless, analysis shows". In: *The Guardian* (2023). Accessed May 7, 2025. URL: https://www.theguardian.com/environment/2023/jan/18/revealed-forest-carbon-offsets-biggest-provider-worthless-verra-aoe.

[35] Keita Hamamoto. *Copula-Based Density Estimation Models for Multivariate Zero-Inflated Continuous Data*. Submitted to IEEE; Central Research Laboratory, Hitachi Ltd., Kokubunji, Japan. 2023. arXiv: 2304.00537 [stat.ME]. URL: https://arxiv.org/abs/2304.00537.

[36] Heejoon Han and Dennis Kristensen. "Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates". In: *Journal of business & economic statistics* 32.3 (2014), pp. 416–429.

[37] Heejoon Han and Dennis Kristensen. "Supplemental Material to "Asymptotic Theory for the QMLE in GARCH-X Models with Stationary and Non-Stationary Covariates"". In: *Journal of Business & Economic Statistics* (2014). Supplemental material containing proofs of lemmas, pp. 416–429. DOI: 10.1080/07350015.2014.897954. URL: https://doi.org/10.1080/07350015.2014.897954.

[38] M. Hofert et al. *Elements of Copula Modeling with R*. Use R! Springer, 2018. ISBN: 9783319896366.

[39] Sárka Hudecová and Michal Pešta. "Copula hurdle GARCH models for multivariate non-negative time series". In: *Statistical Papers* 66.4 (June 2025), pp. 1–19. DOI: 10.1007/s00362-025-01713-x. URL: https://ideas.repec.org/a/spr/stpapr/v66y2025i4d10.1007_s00362-025-01713-x.html.

[40] Søren Tolver Jensen and Anders Rahbek. "Asymptotic Inference for Nonstationary GARCH". In: *Econometric Theory* 20.6 (Dec. 2004). 24 pages, pp. 1203–1226. URL: https://www.jstor.org/stable/3533451.

[41] Pim Keer. "Hypothesis Testing in Contingency Tables". Mentors: H.P. Lopuhaä and Øyvind Bakke (NTNU). Master's Thesis. Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2023.

[42] Dennis Kristensen and Anders Rahbek. "Asymptotics of the QMLE for a Class of ARCH(q) Models". In: *Econometric Theory* 21.5 (2005), pp. 946–961. URL: http://www.jstor.org/stable/3533519.

[43] Dennis Kristensen and Yongcheol Shin. "Estimation of dynamic models with nonparametric simulated maximum likelihood". In: *Journal of Econometrics* 167 (2012), pp. 76–94.

[44] David A. Lesmond, Joseph P. Ogden, and Charles A. Trzcinka. "A New Estimate of Transaction Costs". In: *The Review of Financial Studies* 12.5 (1999), pp. 1113–1141. ISSN: 08939454, 14657368. URL: http://www.jstor.org/stable/2645977 (visited on 08/04/2025).

[45] Alexander Lex et al. “UpSet: Visualization of Intersecting Sets”. In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 1983–1992. DOI: 10.1109/TVCG.2014.2346248.

[46] Roger B. Nelsen. *An Introduction to Copulas*. 2nd ed. Springer Series in Statistics. New York, NY: Springer, 2006. ISBN: 978-0-387-28659-4. DOI: 10.1007/0-387-28678-0.

[47] Octopus Investments. *Net Zero: The Second Stage - The need for quality in the carbon credit market*. Octopus Investments, Feb. 2025. URL: https://octopusgroup.com/newsroom/latest-news/almost-60-of-uk-senior-decisionmakers-dont-know-what-carbon-credits-are/.

[48] Andrew Patton. “Modelling time-varying exchange rate dependence using the conditional copula”. In: *Econometrics eJournal* (2001). Available at SSRN: https://ssrn.com/abstract=264991.

[49] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1988. ISBN: 1558604790.

[50] Elisa Perrone, Edwin R. van den Heuvel, and Zhuozhao Zhan. “Kendall’s Tau Estimator for Bivariate Zero-Inflated Count Data”. In: *Statistics and Probability Letters* 199 (2023), p. 109858. DOI: 10.1016/j.spl.2023.109858. URL: https://doi.org/10.1016/j.spl.2023.109858.

[51] L.S.C. Piancastelli and R.B. Silva. “Multivariate Zero-Inflated INGARCH Models: Bayesian Inference and Composite Likelihood Approach”. In: *Statistics and Computing* 35.19 (2025). DOI: 10.1007/s11222-024-10549-8. URL: https://doi.org/10.1007/s11222-024-10549-8.

[52] R.S. Pimentel, M. Niewiadomska-Bugaj, and J.C. Wang. “Association of Zero-Inflated Continuous Variables”. In: *Statistics & Probability Letters* 96 (2015), pp. 61–67. DOI: 10.1016/j.spl.2014.09.002. URL: https://doi.org/10.1016/j.spl.2014.09.002.

[53] Nasim Pour, Sebastien Cross, and Joel Gould. *The EU is returning to international carbon credits. Here’s what it means*. Accessed: 2025-08-03. World Economic Forum. July 2025. URL: https://www.weforum.org/stories/2025/07/eu-return-to-international-carbon-credits/.

[54] Quantum Commodity Intelligence. *Carbon Market Data and Assessments*. https://www.qcintel.com/carbon/. Accessed: 18 June 2025. 2025.

[55] Rabobank. *Internal discussions on carbon credit pricing and market linkages*. Personal communication. Unpublished internal discussions with Rabobank representatives. 2025.

[56] Hamdi Raïssi. *On the Correlation Analysis of Illiquid Stocks*. Instituto de Estadística, PUCV, Valparaíso, Chile. ANID funding Fondecyt 1201898 is acknowledged. 2020. arXiv: 2008.06168 [q-fin.ST]. URL: https://arxiv.org/pdf/2008.06168.

[57] Isuru Panduka Ratnayake and V. A. Samaranayake. “An integer GARCH model for a Poisson process with time-varying zero-inflation”. In: *PLOS ONE* 18.5 (May 2023), pp. 1–32. DOI: 10.1371/journal.pone.0285769. URL: https://doi.org/10.1371/journal.pone.0285769.

[58] R. Van der Spek. “Fast Estimation of Kendall’s Tau and Conditional Kendall’s Tau Matrices under Structural Assumptions”. In: *TU Delft* (2022).

[59] Genaro Sucarrat. “garchx: Flexible and Robust GARCH-X Modeling”. In: *The R Journal* 13 (1 2021). https://rjournal.github.io/, pp. 335–350. ISSN: 2073-4859.

[60] Genaro Sucarrat and Álvaro Escribano. “Estimation of log-GARCH Models in the Presence of Zero Returns”. In: *European Journal of Finance* 23.9 (2017). Published online: 15 June 2017, pp. 809–827. DOI: 10.1080/1351847X.2017.1336452. URL: https://doi.org/10.1080/1351847X.2017.1336452.

[61] Genaro Sucarrat and Steffen Grønneberg. “Risk Estimation with a Time-Varying Probability of Zero Returns*”. In: *Journal of Financial Econometrics* 20.2 (June 2020), pp. 278–309. ISSN: 1479-8409. DOI: 10.

1093/jjfinec/nbaa014. eprint: https://academic.oup.com/jfec/article-pdf/20/2/278/43109421/nbaa014.pdf. URL: https://doi.org/10.1093/jjfinec/nbaa014.

[62] Magdalena Szumilas. "Explaining Odds Ratios". In: *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 19.3 (2010), pp. 227–229. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC2938757/pdf/ccap19_3p227.pdf.

[63] Gregory Trencher et al. "Demand for Low-Quality Offsets by Major Companies Undermines Climate Integrity of the Voluntary Carbon Market". In: *Nature Communications* 15 (2024). Published online: 10 August 2024, Accessed April 24, 2025. DOI: 10.1038/s41467-024-51151-w. URL: https://www.nature.com/articles/s41467-024-51151-w.

[64] Howard G. Tucker. "A Generalization of the Glivenko-Cantelli Theorem". In: *Annals of Mathematical Statistics* 30.3 (Sept. 1959), pp. 828–830. DOI: 10.1214/aoms/1177706212. URL: https://doi.org/10.1214/aoms/1177706212.

[65] United Nations Framework Convention on Climate Change. *Kyoto Protocol - Mechanisms under the Protocol*. Accessed April 24, 2025. UNFCCC. 2023. URL: https://unfccc.int/process/the-kyoto-protocol/mechanisms.

[66] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. First paperback edition 2000. Cambridge: Cambridge University Press, 1998. ISBN: 9780521496032.

[67] Hitendra Varsani and Rohit Gupta. *Carbon Markets: An Emerging Asset Class?* Accessed April 24, 2025. MSCI Research. 2023. URL: https://www.msci.com/www/blog-posts/carbon-markets-an-emerging/03256542753.

[68] Hitendra D. Varsani and Rohit Gupta. *Introducing the Carbon Market Age*. Accessed April 24, 2025. MSCI Research. June 2022. URL: https://www.msci.com/www/blog-posts/introducing-the-carbon-market/03227158119.

[69] Yongdeng Xu. *Extended multivariate EGARCH model: A model for zero-return and negative spillovers*. eng. Cardiff Economics Working Papers E2024/24. Cardiff, 2024. URL: https://hdl.handle.net/10419/309974.

[70] F. Yates. "Tests of significance for 2 × 2 contingency tables". In: *Journal of the Royal Statistical Society: Series A (General)* 147.3 (1984). [Online], pp. 426–449. DOI: 10.2307/2981577. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.2307/2981577.

Appendices

A List of assets

This chapter describes the different assets, both carbon credit and other commodities, in more detail. The description of the carbon credits comes from Quantum Commodity Intelligence.

A.1 Voluntary carbon credits

1. **VCS ARR Uruguay Vintage 2021**
   This voluntary carbon credit corresponds to afforestation, reforestation, and restoration (ARR) activities conducted in Uruguay under the Verified Carbon Standard (VCS). The credit is attributed to the vintage year 2021 and typically originates from eucalyptus monoculture plantations on degraded lands, as exemplified by the Guanaré project (VCS959). While this type of ARR project is used as a benchmark, similar projects across Latin America may be considered with appropriate normalization. The valuation reflects a standard deal size of 20,000 tonnes, with adjustments made for smaller or larger volumes.

2. **ACR US IFM**
   This credit represents a 2021 vintage voluntary carbon credit generated by an Improved Forest Management (IFM) project in the United States and registered under the American Carbon Registry (ACR). These projects aim to enhance carbon sequestration through more sustainable forest management practices. The valuation is based on a standard transaction size of 20,000 tonnes, with normalization applied for different volumes. An indicative example of such a project is ACR698.

3. **VCS REDD+**
   This credit pertains to avoided deforestation projects under the REDD+ framework, carrying additional Climate, Community, and Biodiversity (CCB) Gold certification. Registered under the Verra emissions registry, this assessment aggregates data from ten benchmark projects—including Envira Amazonia, Katingan, Kariba, and Southern Cardamom—spanning Latin America, Africa, and Asia. Credits are of vintage 2021 or earlier, with recent inclusion of additional projects such as Tambopata and Keo Seima from July 2023. This diversified composition ensures a representative valuation of high-quality REDD+ credits.

4. **Clean Cookstoves Africa Vintage 2021**
   This voluntary carbon credit is generated from clean cookstove projects in Africa, accredited under the Gold Standard and reflecting a vintage year of 2021. Projects in Kenya, Uganda, Somalia, Nigeria, and Ghana are typically assessed at par, with normalization for other countries. All stove types are considered—ranging from charcoal and biomass to LPG and water purifiers—though the lowest-cost technology generally sets the market value. These projects commonly contribute to at least four United Nations Sustainable Development Goals (SDGs). An indicative project is GS5642.

5. **Nature-Based Offset Exchange Index**
   This credit represents a broader market index of voluntary carbon credits from the Agriculture, Forestry and Other Land Use (AFOLU) sector, registered with Verra and including CCB accreditation. Eligible projects span a range of methodologies including REDD, Improved Forest Management, Wetland Restoration, and Avoided Grassland Conversion. The lowest-cost qualifying project sets the daily reference price, ensuring a market-reflective valuation. An indicative project included in this index is VCS934.

6. **GS Water Filtration Africa Vintage 2021**
   This carbon credit reflects emissions reductions from water filtration projects across select African countries, accredited under the Gold Standard and of vintage 2021. Countries such as Kenya, Malawi, Togo, and Burkina Faso form the core reference group, with normalization applied as necessary for other regions. The assessed projects typically address at least four SDGs and contribute to both climate mitigation and public health. An example of such a project is GS11207.

7. **VCS CCB / GS ARR Latin America 2021**
   This assessment covers ARR projects located in Central and South America that are registered either under the Verified Carbon Standard (with CCB certification) or the Gold Standard. The credits are of vintage 2021 and reflect activities aimed at ecosystem restoration and afforestation. A standard transaction size of 20,000 tonnes is assumed, with adjustments made for scale. An indicative project ID associated with this category is VCS2512.

A.2 List of commodity assets

Below, one can find a more detailed description of the selected conventional assets that we have used in this thesis:

1. **EU Emissions Trading System (EU ETS)**
   The EU ETS is the world's most established cap-and-trade carbon market. It regulates greenhouse gas emissions from power generators, industrial facilities, and aviation within the European Union. Each EU Allowance permits the emission of one metric ton of CO_2 equivalent. The benchmark pricing instrument is the EUA December futures contract, traded on platforms such as ICE Endex and EEX. This contract is widely regarded as the standard reference for

compliance carbon pricing in Europe and is sensitive to policy developments, energy prices, and industrial activity levels.

2. **iShares Global Clean Energy ETF (ICLN)**
   ICLN is an exchange-traded fund that tracks the performance of the S&P Global Clean Energy Index. It includes approximately 100 companies involved in renewable energy generation and related technologies such as solar, wind, hydroelectric, and fuel cells. With holdings spanning North America, Europe, and Asia, the ETF offers diversified exposure to the global energy transition. Key constituents include companies like Enphase Energy, Vestas Wind Systems, and Ørsted. Traded in USD on the NASDAQ, ICLN is a widely used benchmark for clean energy equity investment.

3. **NASDAQ Composite Index (IXIC)**
   The NASDAQ Composite is a broad-based equity index comprising over 3,000 common stocks listed on the NASDAQ Stock Exchange. It is heavily weighted toward the technology, biotech, and consumer growth sectors. Key constituents include Apple, Microsoft, Amazon, and Nvidia. The index is market-cap weighted and is considered the standard measure of performance for U.S. growth and technology equities. It is frequently used as a benchmark for mutual funds, ETFs, and derivatives linked to U.S. tech exposure.

4. **Natural Gas Futures, Jun-2025 (NG=F)**
   This contract represents the NYMEX Henry Hub Natural Gas Futures for June 2025, the most liquid and standardized pricing instrument for U.S. natural gas. Each contract corresponds to 10,000 million British thermal units (mmBtu), with delivery at the Henry Hub in Louisiana. The June contract is typically used as a reference for early summer demand, especially for cooling-related power generation. It is the primary benchmark for pricing physical and financial natural gas products in North America.

5. **Soybean Futures, Jul-2025 (ZS=F)**
   This is the benchmark CBOT futures contract for U.S. soybeans, representing a standardized trade of 5,000 bushels of No. 2 Yellow Soybeans, with July 2025 as the delivery month. The July contract is classified as a "new crop" month and is the primary pricing reference for expected harvest conditions in the U.S. Midwest. It is the standard price series used globally for soybean risk management and speculative positioning, with prices quoted in USD.

6. **CBOE Volatility Index (VIX)**
   The CBOE Volatility Index is a real-time index that measures market expectations for volatility over the next 30 calendar days, derived from S&P 500 Index option prices. Often called the "fear gauge," it increases during periods of market stress and declines during stable or bullish conditions. Though not directly tradable, the VIX serves as the benchmark measure of U.S. equity market volatility and underpins a variety of derivatives including futures and ETFs. It is closely followed by investors as an indicator of market sentiment.

B
Proofs

The following proofs are based on the supplemental material of Han and Kristensen's paper on asymptotic theory for the QMLE in GARCH-X Models with stationary and non-stationary covariates. See for more information [37].

B.1 Proof of Lemma 5.1

Lemma B.1.0. *Under Assumption 6.1, there exists a stationary and ergodic solution to Equation 6.1 at* $\boldsymbol{\theta}_j^\star$ *satisfying* $\mathbb{E}[\sigma_{j,t}^{2s}] < +\infty$ *and* $\mathbb{E}[r_{j,t}^{2s}] < +\infty$ *for some* $0 < s < 1$.

Proof. Let us define $a_{j,t} := \alpha_{j,1}^\star \eta_{j,t-1}^2 + \alpha_{j,2}^\star \geq 0$, and $b_{j,t} := \alpha_{j,0}^\star + \alpha_{j,3}^\star I_{j,t}^2 \geq 0$. We can rewrite the true data generating process as,

$$\sigma_{j,t}^2 = a_{j,t}\sigma_{j,t-1}^2 + b_{j,t}.$$

This is a stochastic recursion where $(a_{j,t}, b_{j,t})_{t\geq 1}$ is a stationary and ergodic sequence. The first part of the result now follows from [10]. This is because Assumption 6.1(ii) implies that the Lyapunov coefficient, i.e. $\mathbb{E}[\log(a_{j,t})]$, associated with the above stochastic recursion is negative, and that $\mathbb{E}\left[\log^+(b_{j,t})\right] < \infty$. Next, the stationary solution can be written as $\sigma_t^2 = b_{j,t} + \sum_{i=0}^\infty a_{j,t}\cdots a_{j,t-i}b_{j,t-i-1}$. Following [6], the negative Lyapunov coefficient implies that $\mathbb{E}[(a_0\cdots a_m)^{2s}] < 1$ for some $s > 0$ and $m \geq 1$. Hence, $\mathbb{E}[(a_{j,t}\cdots a_{j,t-i})^{2s}] \leq c\varrho^i$ for some $c < 1$ and $\varrho < 1$. Clearly, it holds that $\mathbb{E}\left[b_{j,t}^{2s}\right] < \infty$. Then,

$$\mathbb{E}\left[\sigma_{j,t}^{2s}\right] \leq \mathbb{E}[b_{j,t}^s] + \sum_{i=0}^\infty \mathbb{E}\left[(a_{j,t}\cdots a_{j,t-i})^s b_{j,t}^s\right] \tag{B.1}$$

$$\leq \mathbb{E}[b_{j,t}^s] + \sqrt{\mathbb{E}\left[b_{j,t}^{2s}\right]}\sum_{i=0}^\infty \sqrt{\mathbb{E}\left[(a_{j,t}\cdots a_{j,t-i})^{2s}\right]} \tag{B.2}$$

$$= \mathbb{E}[b_{j,t}^s] + c\sqrt{\mathbb{E}\left[b_{j,t}^{2s}\right]}(1-\varrho)^{-1} \tag{B.3}$$

$$< \infty. \tag{B.4}$$

Note that $\mathbb{E}[r_{j,t}^{2s}] < \infty$ follows from $r_{j,t} = I_{j,t}\sigma_{j,t}(\boldsymbol{\theta}_j)\eta_{j,t}$ together with Assumption 6.1(ii).

□

B.2 Proof of Lemma 5.2

Lemma B.2.0. *Under Assumption 6.1, with $s > 0$ given in Lemma 6.1, there exists some $K_s < +\infty$ such that*

$$\mathbb{E}\left[sup_{\boldsymbol{\theta}_j\in\boldsymbol{\Theta}_j}|s_{j,t}(\boldsymbol{\theta}_j) - s^*_{j,t}(\boldsymbol{\theta}_j)|^s\right] \leq K_s\alpha^{st}_{j,2}, \tag{B.5}$$

where

$$s^*_{j,t}(\boldsymbol{\theta}_j) = \frac{\tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j)}{\tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}^*_j)}, \quad \tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j) := \sum_{i=1}^{+\infty}\alpha^{i-1}_{j,2}(\alpha_{j,0} + \alpha_{j,1}r^2_{j,t} + \alpha_{j,3}I^2_{j,t-1}) \tag{B.6}$$

The process $\tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j)$ is stationary and ergodic with $\mathbb{E}[sup_{\boldsymbol{\theta}_j\in\boldsymbol{\Theta}_j}\tilde{\sigma}^{2s}_{j,t}(\boldsymbol{\theta}_j)] < +\infty$.

Proof. We can rewrite $\sigma^2_{j,t} = \alpha_{j,0} + \alpha_{j,1}\,\sigma^2_{j,t^*} + \alpha_{j,2}\,r^2_{j,t^*} + \alpha_{j,3}\,I_{j,t-1}$ as

$$\sigma^2_{j,t}(\boldsymbol{\theta}_j) := \alpha_{j,2}\sigma^2_{j,t-1}(\boldsymbol{\theta}_j) + w_{j,t}(\boldsymbol{\theta}_j), \tag{B.7}$$

which is an AR(1) model with stationary errors

$$w_t(\boldsymbol{\theta}_j) = \alpha_{j,0} + \alpha_{j,1}r^2_{j,t-1} + \alpha_{j,3}I^2_{j,t-1}. \tag{B.8}$$

The first part of the result now follows by [6]. From Lemma 6.1 together with Assumption 6.1(ii), we have

$$\mathbb{E}\Big[\sup_{\boldsymbol{\theta}_j\in\boldsymbol{\Theta}_j} w^s_{j,t}(\boldsymbol{\theta}_j)\Big] \leq \overline{\alpha}^s_{j,0} + \overline{\alpha}^s_{j,1}\mathbb{E}[r^{2s}_{j,t-1}] + \overline{\alpha}^s_{j,3}\mathbb{E}[I^{2s}_{j,t-1}] < \infty. \tag{B.9}$$

Thus,

$$\mathbb{E}\Big[\sup_{\boldsymbol{\theta}_j\in\boldsymbol{\Theta}_j} \sigma^{2s}_{0,t}(\boldsymbol{\theta}_j)\Big] \leq \sum_{i=0}^{\infty}(\alpha_{j,2})^{is}\mathbb{E}\Big[\sup_{\boldsymbol{\theta}_j\in\boldsymbol{\Theta}_j} w^s_{j,t}(\boldsymbol{\theta}_j)\Big] < \infty. \tag{B.10}$$

Therefore, we conclude that there exists a stationary and ergodic solution given by $\tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j)$, where $\tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j) := \sum_{i=0}^{+\infty}\alpha^i_{j,2}w_{t-i}(\boldsymbol{\theta}_j)$. Next, observe that $\sigma^2_t(\boldsymbol{\theta}_j) - \tilde{\sigma}^2_{j,t}(\boldsymbol{\theta}_j) = (\alpha_{j,2})\{\sigma^2_{j,t-1} - \tilde{\sigma}^2_{j,t-1}(\boldsymbol{\theta}_j)\} = (\alpha_{j,2})^t\{\overline{\sigma}^2_{j,0} - \tilde{\sigma}^2_{j,0}(\boldsymbol{\theta}_j)\}$, where $\bar{\sigma}^2_{j,0} > 0$ is the fixed, initial value, used to compute the likelihood function. The result now follows with

$$K_s = \mathbb{E}\left[\sup_{\boldsymbol{\theta}_j\in\boldsymbol{\Theta}_j}|\bar{\sigma}^2_0 - \tilde{\sigma}^2_{j,0}(\boldsymbol{\theta}_j)|^s\right], \tag{B.11}$$

which is finite, according to Lemma 6.1, since

$$\mathbb{E}\Big[\sup_{\boldsymbol{\theta}_j\in\boldsymbol{\Theta}} \tilde{\sigma}^{2s}_{j,0}(\boldsymbol{\theta}_j)\Big] < \infty. \tag{B.12}$$

□

B.3 Proof of Lemma 5.3

Lemma B.3.0. *Under Assumptions 6.1 - 6.2, the first and second derivatives of the volatility-ratio process $s_t(\boldsymbol{\theta})$, given by $\frac{\partial s_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and $\frac{\partial^2 s_t^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$, are stationary and ergodic for all $\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j$. Moreover, there exist stationary and ergodic sequences $B_{k,t} \in \mathcal{F}_{t-1}$, for $k = 0, 1, 2$, which are independent of $\boldsymbol{\theta}_j$, such that the following bounds hold for all $\boldsymbol{\theta}_j$ in a neighborhood of $\boldsymbol{\theta}_j^\star$:*

$$\frac{1}{s_t^*(\boldsymbol{\theta}_j)} \leq B_{0,t}, \quad \left\| \frac{\partial s_t^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j} \right\| \leq B_{1,t} \cdot s_t^*(\boldsymbol{\theta}_j), \quad \left\| \frac{\partial^2 s_t^*(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_j'} \right\| \leq B_{2,t} \cdot s_t^*(\boldsymbol{\theta}_j). \tag{B.13}$$

where it holds that:

$$\mathbb{E}\left[B_{1,t} + B_{2,t}^2\right] < \infty, \quad \textit{and} \quad \mathbb{E}\left[B_{0,t} \cdot \left(B_{1,t} + B_{2,t}^2\right)\right] < \infty. \tag{B.14}$$

Proof. Observe that

$$\frac{\partial \sigma_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,0}} = 1 + \alpha_{j,2} \frac{\partial \sigma_{j,t-1}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,0}} = \ldots = \sum_{i=0}^{t} \alpha_{j,2}^i, \tag{B.15}$$

$$\frac{\partial \sigma_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,1}} = r_{t-1}^2 + \alpha_{j,2} \frac{\partial \sigma_{t-1}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,1}} = \ldots = \sum_{i=0}^{t} \alpha_{j,2}^i r_{t-1-i}^2, \tag{B.16}$$

$$\frac{\partial \sigma_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,3}} = I_{j,t-1}^2 + \alpha_{j,2} \frac{\partial \sigma_{j,t-1}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,3}} = \ldots = \sum_{i=0}^{t} \alpha_{j,2}^i I_{j,t-1-i}^2, \tag{B.17}$$

$$\frac{\partial \sigma_t^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,2}} = \sigma_{j,t-1}^2(\boldsymbol{\theta}_j) + \alpha_{j,2} \frac{\partial \sigma_{j,t-1}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,2}} = \ldots = \sum_{i=0}^{t} \alpha_{j,2}^i \sigma_{j,t-1-i}^2(\boldsymbol{\theta}_j). \tag{B.18}$$

By the same arguments as in the proof of Lemma 6.2, these equations have stationary solutions. The proof for the second-order partial derivatives with respect to $\alpha_{j,0}$, $\alpha_{j,1}$, and $\alpha_{j,2}$ proceeds along the lines of [28, p. 619], since these do not involve $I_{j,t}$. Regarding the second-order derivatives involving $\alpha_{j,3}^\star$, using the above expressions of the first-order derivatives:

$$\frac{\partial^2 \sigma_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,0} \partial \alpha_{j,3}} = \alpha_{j,2} \frac{\partial^2 \sigma_{j,t-1}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,0} \partial \alpha_{j,3}} = \ldots = 0, \tag{B.19}$$

$$\frac{\partial^2 \sigma_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,1} \partial \alpha_{j,3}} = \alpha_{j,2} \frac{\partial^2 \sigma_{j,t-1}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,1} \partial \alpha_{j,3}} = \ldots = 0, \tag{B.20}$$

$$\frac{\partial^2 \sigma_{j,t}^2(\boldsymbol{\theta}_j)}{\partial (\alpha_{j,3})^2} = \alpha_{j,2} \frac{\partial^2 \sigma_{j,t-1}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,2} \partial \alpha_{j,3}} = \ldots = 0, \tag{B.21}$$

$$\frac{\partial^2 \sigma_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,2} \partial \alpha_{j,3}} = \frac{\partial \sigma_{j,t-1}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,3}} + \alpha_{j,2} \frac{\partial^2 \sigma_{j,t-1}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,2} \partial \alpha_{j,3}} = \ldots = \sum_{i=0}^{t} \alpha_{j,2}^i \frac{\partial \tilde{\sigma}_{j,t-1-i}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,3}}. \tag{B.22}$$

Again, there clearly exist stationary solutions to these equations. Moreover, by the same arguments as in [28, p. 622], there exist constants $c < \infty$ and $0 < \varrho < 1$ such that for all $\boldsymbol{\theta}_j$ in a neighborhood of

$\boldsymbol{\theta}_j^\star$ and all $0 < r \leq s$,

$$\frac{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j^\star)}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)} \leq c \sum_{i=0}^{\infty} \varrho^{ri} \overline{w}_{j,t}^r, \tag{B.23}$$

where $\overline{w}_{j,t} := \overline{\alpha}_{j,0} + \overline{\alpha}_{j,1} r_{t-1}^2 + \overline{\alpha}_{j,3} I_{j,t}^2$ is stationary and ergodic with $\mathbb{E}[\bar{w}_{j,t}^r] < \infty$. This in turn implies that $\sum_{i=0}^{\infty} \varrho^{is} \overline{w}_{j,t}^r$ is stationary and ergodic with first moment. Given the representations of the stationary solutions $\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j^\star)$ and $\partial(\sigma_{j,t})^2(\boldsymbol{\theta}_j^\star)/(\partial \boldsymbol{\theta}_j)$, it is easily shown that for some constant $c < \infty$ the following inequalities hold for all $\boldsymbol{\theta}_j$ in a neighborhood of $\boldsymbol{\theta}_j^\star$ [28, p. 619].

$$\frac{1}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j^\star)} \frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j^\star)}{\partial \alpha_{j,0}} \leq \frac{1}{\alpha_{j,0}^\star}, \quad \frac{1}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)} \frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,1}} \leq \frac{1}{\alpha_{j,1}^\star}, \quad \frac{1}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)} \frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,3}} \leq \frac{1}{\alpha_{j,3}^\star}, \tag{B.24}$$

$$\frac{1}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)} \frac{\partial \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,2}} \leq c \sum_{i=0}^{\infty} (\alpha_{j,2})^{ri} \overline{w}_{j,t}^r, \tag{B.25}$$

Finally, by the same arguments as in [28, p. 620], it also holds that

$$\frac{1}{\tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)} \frac{\partial^2 \tilde{\sigma}_{j,t}^2(\boldsymbol{\theta}_j)}{\partial \alpha_{j,2} \partial \alpha_{j,3}} \leq c \sum_{i=0}^{\infty} (\alpha_{j,2})^{ri} \overline{w}_{j,t}^r. \tag{B.26}$$

By inspection of the definitions of $B_{0,t}$, $B_{1,t}$, and $B_{2,t}$, one finds that the stated moment exists by choosing $r > 0$ sufficiently small.

□

C

Simplifying assumption

We investigate the simplifying assumption in our setting, using seven voluntary carbon credits and six conventional assets as described in Chapter 4. The simplifying assumption states that a conditional copula,

$$C_{12|Z}(u_1, u_2 \mid Z = z),$$

does not depend on the value of the conditioning variable z, i.e. the dependence between (X_1, X_2) given Z is constant across all realizations of Z. To formally test this property, we follow the nonparametric framework of Derumigny and Fermanian [17]. Specifically, we apply `CondCopulas::simpA.NP` with the Kolmogorov–Smirnov–type statistic `T1_KS_Cs3` (Equation (4) in [17]), where the conditional copula is estimated nonparametrically by kernel smoothing. In our implementation we specify a Gaussian kernel with bandwidth $h = 0.5$, and obtain p-values via a nonparametric bootstrap with $B = 100$ replications (`typeBoot = "boot.NP"`). For each asset triple $(X_1, X_2 \mid Z)$, the function outputs the test statistic and its bootstrap p-value.

The null hypothesis is the simplifying assumption, i.e. the conditional copula does not depend on the conditioning variable. Here, p-values smaller than 0.05 imply rejection, meaning that the dependence between X_1 and X_2 changes with Z. Larger p-values imply no evidence against simplification.

Table C.1 reports the outcomes of the simplifying assumption test, ordered by ascending p-values. We focus on the lowest p-values, as these may provide the strongest evidence against the simplifying assumption. Beyond the entries shown, all remaining asset triples yield p-values above 0.51, suggesting no indication of conditional variation in their copulas. Overall, the results imply that for the vast majority of asset combinations, we do not reject the simplifying assumption, meaning that the cross-asset dependence appears stable with respect to the conditioning variable. Only a small number of cases show weaker support, but even these remain far from conventional significance thresholds.

Table C.1: *Nonparametric test of the simplifying assumption for selected triples $(X_1, X_2 \mid Z)$. Reported are the KS test statistic and bootstrap p-value.*

X_1	X_2	Z	Stat.	p-value
VIX	Renewable Energy Contract	NASDAQ	0.0225	0.00
NASDAQ	EU ETS	Renewable Energy Contract	0.0165	0.08
NASDAQ	VCS Uruguay ARR	Renewable Energy Contract	0.0147	0.08
Nature Based Offset	VCS REDD+	ACR US IFM	0.0509	0.10
NASDAQ	Nature Based Offset	VIX	0.0233	0.11
Nature Based Offset	VCS REDD+	Renewable Energy Contract	0.0252	0.12
VIX	Natural Gas	EU ETS	0.0125	0.13
VIX	Nature Based Offset	Natural Gas	0.0236	0.13
Renewable Energy Contract	VCS Latam ARR	Natural Gas	0.0147	0.18
EU ETS	VCS REDD+	Renewable Energy Contract	0.0201	0.18
Renewable Energy Contract	NASDAQ	EU ETS	0.0149	0.20
Renewable Energy Contract	Nature Based Offset	Renewable Energy Contract	0.0240	0.22
NASDAQ	Renewable Energy Contract	VCS Uruguay ARR	0.0469	0.27
Natural Gas	ACR US IFM	VCS REDD+	0.0254	0.29
VIX	Natural Gas	Renewable Energy Contract	0.0122	0.29
VCS REDD+	Natural Gas	Renewable Energy Contract	0.0243	0.30
ACR US IFM	VCS Latam ARR	VCS REDD+	0.0596	0.31
Renewable Energy Contract	Nature Based Offset	VIX	0.0328	0.32
NASDAQ	ACR US IFM	Renewable Energy Contract	0.0103	0.34
ACR US IFM	Renewable Energy Contract	VCS Uruguay ARR	0.0104	0.34
Renewable Energy Contract	Natural Gas	VIX	0.0137	0.36
NASDAQ	VIX	VIX	0.0155	0.36
VIX	Renewable Energy Contract	Renewable Energy Contract	0.0146	0.36
Natural Gas	Renewable Energy Contract	Renewable Energy Contract	0.0221	0.36
ACR US IFM	VCS Latam ARR	Nature Based Offset	0.0590	0.36
Renewable Energy Contract	VCS REDD+	VCS REDD+	0.0121	0.37
NASDAQ	Renewable Energy Contract	Natural Gas	0.0102	0.45
NASDAQ	EU ETS	Natural Gas	0.0101	0.45
VIX	NASDAQ	VCS REDD+	0.0159	0.45
Renewable Energy Contract	NASDAQ	VCS Uruguay ARR	0.0133	0.46
Renewable Energy Contract	EU ETS	Natural Gas	0.0102	0.46
EU ETS	VCS Uruguay ARR	VCS Latam ARR	0.0246	0.47
Renewable Energy Contract	Natural Gas	NASDAQ	0.0131	0.48
EU ETS	VCS Uruguay ARR	Renewable Energy Contract	0.0144	0.48
Renewable Energy Contract	VCS Uruguay ARR	Natural Gas	0.0162	0.49
Natural Gas	VCS REDD+	EU ETS	0.0116	0.50
Renewable Energy Contract	VCS REDD+	Natural Gas	0.0125	0.51