
Electronic Thesis and Dissertation Repository

8-30-2022 10:00 AM

Copulas, maximal dependence, and anomaly detection in bi-variate time series

Ning Sun, *The University of Western Ontario*

Supervisor: Dr. Ricardas Zitikis, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Statistics and Actuarial Sciences

© Ning Sun 2022

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Statistical Methodology Commons](#)

Recommended Citation

Sun, Ning, "Copulas, maximal dependence, and anomaly detection in bi-variate time series" (2022).
Electronic Thesis and Dissertation Repository. 8845.
<https://ir.lib.uwo.ca/etd/8845>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

This thesis focuses on discussing non-parametric estimators and their asymptotic behaviors for indices developed to characterize bi-variate time series. There are typically two types of indices depending on whether the distributional information is involved. For the indices containing the distributional information of the bivariate stationary time series, we particularly focus on the index called the *tail order of maximal dependence* (TOMD), which is an improvement of the tail order. For the indices without distributional information of the bivariate time series, we focus on an anomaly detection index for univariate input-output systems.

This thesis integrates three articles. The first article (Chapter 2) proposes the *average block-minima estimator* for the TOMD and discusses theoretical aspects of this estimator under the independently identically distribution (i.i.d.) assumption, including asymptotic behavior and bias reduction. The performance of this estimator is justified by simulation studies using Marshall-Olkin copula and generalized Clayton copula, respectively. The second article (Chapter 3) examines the performance of the *average block-minima estimator* on stationary bivariate time series using simulation studies. Applications of the estimator on three groups of financial assets are employed to illustrate how the estimation method could be used in practice. The third article (Chapter 4) generalizes an existing anomaly detection index for input-output systems with i.i.d. inputs to those with stationary inputs. Theoretical evidence and illustrative examples are provided to validate the performance of the existing index for systems with stationary inputs.

Keywords: bivariate time series, copula, tail order of maximal dependence, non-parametric estimation, anomaly detection.

Summary for Lay Audience

Bi-variate time series is ubiquitous in real life such as financial risk management and reliability assessment of input-output systems. Depending on the issue of interests, various indices are proposed for bi-variate time series to serve the purpose of evaluation or comparison. Among all these indices, we are particularly interested in the following two: 1) *the tail order of maximal dependence* (TOMD); 2) an index of anomaly detection. The TOMD is an improvement of the existing tail order which is used to measure the extreme co-movements of random variables such as investment returns of assets. The index of anomaly detection is used to examine if the inputs and outputs of systems with certain original order contain persistent anomalies that may not change the regime such as network intrusion. In this research, we develop non-parametric estimation method for the TOMD and index of anomaly detection when the underlying data-generating process is bi-variate time series.

This research consists of three articles. The first article proposes the *average block-minima estimator* of the TOMD for independently identically distributed random pairs. The performance of the estimator is justified by both theoretical analysis and simulation studies. The second article examines the performance of the *average block-minima estimator* of the TOMD for stationary bi-variate time series using simulation studies. The third article extends the existing results for the anomaly detection index to the systems with bi-variate stationary inputs with theoretical validation.

Co-Authorship Statement

The materials presented in this thesis are based on three joint-authored research articles, the first which is published on ASTIN Bulletin: The Journal of the IAA. The second article is submitted to Mathematical Methods of Statistics and is now under review. The third article is published on Applied Stochastic Models for Business and Industry. I am the lead author for all of these articles. I gratefully thank my supervisor Dr. Zitikis and the other team member Dr. Yang for their important contributions to all of the aforementioned works, including suggesting research topics and solution approaches, crucial questions, extensive discussions, insightful comments and careful proofreading.

To my lord Jesus Christ.

Acknowledgements

First and foremost, I would like to thank the almighty God for the opportunity to work on my PhD in Western University.

I would like to express my gratitude to my supervisor Dr. Ričardas Zitikis for his long-standing guidance and support during my doctoral program. His efficient working style and strictness towards academic work benefit me a lot. Under his guide, I have learned many aspects in my research work. I appreciate the opportunity to work with him very much.

I would like to thank my thesis committee members, Dr. Marcos Escobar-Anel, Dr. Jiandong Ren, Dr. Xingfu Zou and Dr. Mohamedou Ould Haye for their contributions to this work. It is my greatest pleasure to work with such a group of leading researchers.

I would like to thank my beloved son Mike. I am so proud to be your mom. Also, I would like to thank my family for their encouragement through my hard times. Without their supports, I cannot finish my work so smoothly.

Finally, I would also like to thank all my colleagues and friends: Jing Bai, Lingzhi Chen, Junhe Chen, Jingyu Cui, Chenyu Fu, Yijuan Ge, Tian Gan, Yiming Huang, Wenjun Jiang, Yuanhao Lai, Ang Li, Yifan Li, Yuying Li, Dan Liu, Kexin Luo, Cong Nie, Fengqin Qin, Rongwei Rong, Jiang Wu, Junquan Xiao, Miao Yang, Qin Ye, Li Yi, Gecheng Yuan, Guandong David Zhang, Ruixi Zhang, Yichen Zhu.

Contents

Abstract	i
Summary for Lay Audience	ii
Co-Authorship Statement	iii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Figures	ix
List of Tables	xi
1 Introduction	1
1.1 Indices based on the copula	2
1.2 Distribution free indices	3
1.3 Research objectives	4
Bibliography	4
2 A statistical methodology for assessing the maximal strength of tail dependence	6
2.1 Introduction	6
2.2 Estimating the tail order of maximal dependence	8
2.2.1 Preliminaries: the TOMD	8
2.2.2 Estimating the TOMD: an idea	9
2.2.3 In pursuit of efficiency: an improved estimator	11
2.3 Estimator’s statistical properties and performance	12
2.3.1 Theoretical results	12
2.3.2 An illustrative simulation study	13
2.4 Estimating the TOMD when ℓ is present	15
2.4.1 Estimating the TOMD	15
2.4.2 An illustrative simulation study	16
2.5 Conclusion	18
Appendices	18

2.A	Auxiliary results	19
2.B	Proofs of Theorems 2.3.1–2.3.5	22
2.C	Histograms and normal fits	27
Bibliography		30
3	Tail maximal dependence in bivariate models: estimation and applications	32
3.1	Introduction	32
3.2	Tail-order estimators	33
3.3	Justification of the estimators	36
3.3.1	Estimating the TOMD	36
3.3.2	Estimating the TODD	38
3.3.3	Estimator’s performance justification	38
3.3.4	Illustrative insights into the cdf F_q^*	44
3.4	An illustrative simulation study	46
3.5	Extreme co-movements of financial instruments	48
3.5.1	Foreign currency exchange rates	49
3.5.2	Stock market indices	50
3.5.3	Diverse financial instruments	51
3.6	Conclusion	53
Appendices		53
3.A	Thresholds and pseudo observations	53
3.B	Testing the validity of bound (3.3.14)	55
3.C	Testing the boundary case of bound (3.3.14)	57
Bibliography		59
4	Detecting systematic anomalies affecting systems when inputs are stationary time series	63
4.1	Introduction	63
4.2	Setting the stage: basic notation	65
4.3	Two actual illustrations	66
4.3.1	Dow Jones and Australian All Ordinaries Indices	66
4.3.2	Sales with a leading indicator	68
4.4	Introducing a controlled experiment	69
4.5	Anomaly detection: a foundation	71
4.6	The experiment: parameter choices and results	74
4.6.1	Precise service range	76
4.6.2	Strict service range	76
4.6.3	Satisfactory service range	76
4.7	Anomaly-free systems	77
4.7.1	Asymptotics of the Λ_n numerator	78
4.7.2	Asymptotics of the Λ_n denominator	80
4.7.3	Back to the index I_n^0	81
4.8	Anomaly-affected orderly systems	82

4.9 A summary and potential extensions	84
Appendices	85
4.A Graphical illustrations	85
4.B Technical details	86
Bibliography	100
5 Future work	105
5.1 About the TOMD κ^*	105
5.2 About the anomaly detection index I	106
Bibliography	106
Curriculum Vitae	107

List of Figures

2.1	The M-O copula (left) and its MTD path (right) on a 10,000 point scatterplot for the specified M-O parameters (a, b) with $\kappa^* = 1.80, 1.52, \text{ and } 1.20$ (top to bottom).	7
2.2	Fits of simulated $\hat{\kappa}_{m_q}^*(m, \theta_0, q)$ values to the normal distribution when $m = 5$ and $q = 0.05$	17
2.3	Fits of simulated $\hat{\kappa}_{m_q}^*(m, \theta_0, q)$ values to the normal distribution when $m = 5$ and $q = 0.1$	18
2.4	The area B_h with its associated curves and points.	25
2.5	The rectangle E_h with its associated curves and points.	26
2.6	Fits of simulated $\hat{\kappa}_n^*(m, \theta_0)$ to the normal distribution for iid data.	28
2.7	Fits of simulated $\hat{\kappa}_n^*(m, \theta_0)$ to the normal distribution for iid data.	29
3.1	The area $B_{q,h}$ and associated curves.	43
3.2	The rectangle $E_{q,h} \subset B_{q,h}$ and associated curves.	44
3.3	Fits of simulated $\widehat{\kappa}_{m_q}^*(5, \theta_0, 0.05)$ to the normal distribution when $q = 0.05$	48
3.4	Fits of simulated $\widehat{\kappa}_{m_q}^*(5, \theta_0, 0.1)$ to the normal distribution when $q = 0.1$	48
3.5	Scatterplots of pseudo observations of foreign currency exchange rates.	50
3.6	Scatterplots of pseudo observations of stock market indices.	51
3.7	The pseudo observations under consideration of diverse financial instruments.	52
3.8	Original x_t 's (left-hand panels) and the pairs of extreme pseudo-observations (right-hand panels) for foreign currency exchange from January 4, 1971, to October 25, 2019.	54
3.9	Original x_t 's (left-hand panels) and the pairs of extreme pseudo-observations (right-hand panels) for stock market indices from January 4, 1971, to February 28, 2020.	55
3.10	Original x_t 's (left-hand panels) and the pairs of extreme pseudo-observations (right-hand panels) for diverse financial instruments from February 5, 1971, to March 3, 2020.	56
4.1	The percentage returns of DJ and AO.	67
4.2	The indices I_n and $B_{n,2}$ corresponding to DJ and AO with respect to the sample sizes $n = 20, \dots, 149$	68
4.3	The original 150-day data of the leading indicator and the sales.	69
4.4	The 1-lag differences of the leading indicator and the sales.	69
4.5	The indices I_n and $B_{n,2}$ of the 150-day sales data with respect to $n = 20, \dots, 149$	70
4.6	The transfer function h_c with $a = 114$ and $b = 126$ corresponding to the automatic voltage regulator with the transfer window 120 ± 6 volts (i.e., $\pm 5\%$).	70

4.7	ARMA(1, 1) inputs $(X_t)_{t \in \mathbb{Z}}$ as specified by model (4.6.1).	74
4.8	The anomaly-affected indices I_n and $B_{n,2}$ for the precise service range with respect to the sample sizes $2 \leq n \leq 300$ for ARMA(1, 1) inputs.	77
4.9	The anomaly-free indices I_n^0 and $B_{n,2}^0$ for the strict service range with respect to the sample sizes $2 \leq n \leq 300$ for ARMA(1, 1) inputs.	78
4.10	The anomaly-free indices I_n^0 and $B_{n,2}^0$ for the satisfactory service range with respect to the sample sizes $2 \leq n \leq 300$ for ARMA(1, 1) inputs.	78
4.11	The anomaly-affected indices I_n and $B_{n,2}$ for the strict service range with respect to $2 \leq n \leq 300$ for ARMA(1, 1) inputs and iid Lomax(1.2, 1) anomalies.	86
4.12	The anomaly-affected indices I_n and $B_{n,2}$ for the strict service range with respect to $2 \leq n \leq 300$ for ARMA(1, 1) inputs and iid Lomax(11, 1) anomalies.	87
4.13	The anomaly-affected indices I_n and $B_{n,2}$ for the satisfactory service range with respect to $2 \leq n \leq 300$ for ARMA(1, 1) inputs and iid Lomax(1.2, 1) anomalies.	88
4.14	The anomaly-affected indices I_n and $B_{n,2}$ for the satisfactory service range with respect to $2 \leq n \leq 300$ for ARMA(1, 1) inputs and iid Lomax(11, 1) anomalies.	89

List of Tables

2.1	The parameter (a, b) choices with κ^* values and MTD strength.	13
2.2	Summary statistics for $\hat{\kappa}_n^*(m, \theta_0)$ based on iid data of size $n = 10,000$ with the p -values of the A-D and C-vM tests for normality shaded when they fall below 0.05.	14
2.3	Summary statistics for $\hat{\kappa}_n^*(m, \theta_0)$ based on iid data of size $n = 20,000$ with the p -values of the A-D and C-vM tests for normality shaded when they fall below 0.05.	15
2.4	Summary of simulation results when $m = 5$ and $q = 0.05$	17
2.5	Summary of simulation results when $m = 5$ and $q = 0.1$	17
3.4.1	Summary of simulation results when $q = 0.05$	47
3.4.2	Summary of simulation results when $q = 0.1$	48
3.5.1	The upper-triangle entries of each panel report estimated tail orders, and the lower-triangle entries report the corresponding sample sizes.	49
3.5.2	The upper-triangle entries of each panel report estimated tail orders, and the lower-triangle entries report the corresponding sample sizes.	50
3.5.3	The upper-triangle entries of each panel report estimated tail orders, and the lower-triangle entries report the corresponding sample sizes.	52
3.A.1	The percentages of p -values retaining the null of white noise at the significance level $\alpha = 0.05$ alongside the sample sizes m_q (in parentheses) for appropriate choices of q	53
3.B.1	Testing H_0 vs H_1 of pseudo observations of foreign currency exchange rates. . .	57
3.B.2	Testing H_0 vs H_1 of pseudo observations of stock market indices.	58
3.B.3	Testing H_0 vs H_1 of pseudo observations of diverse financial instruments. . . .	58
3.C.1	Testing H_0^* vs H_1^* of pseudo observations of foreign currency exchange rates. . .	59
3.C.2	Testing H_0^* vs H_1^* of pseudo observations of stock market indices.	59
3.C.3	Testing H_0^* vs H_1^* of pseudo observations of diverse financial instruments. . . .	59

Chapter 1

Introduction

This study focuses on the bi-variate series $(X_t, Y_t)_{t \in \mathbb{Z}}$ with $(X_t, Y_t) \stackrel{d}{=} (X, Y)$ for every $t \in \mathbb{Z}$. By “focusing” we mean that the particular case that X_t is independent of Y_s for every pair $(s, t) \in \mathbb{Z}^2$ is excluded automatically. Hence there remains two basic situations of our interest: the first situation is that (X_t, Y_t) has a common dependence structure for every $t \in \mathbb{Z}$; the other situation is that (X_t, Y_t) does not only have a common dependence structure but also actually satisfies a common function relationship.

For the first situation, the common dependence structure as well as its measurement are usually of the interest. By assuming that (X, Y) is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with continuous marginal cumulative distribution functions (CDF)

$$F(x) := \mathbb{P}(X \leq x) \quad \text{and} \quad G(y) := \mathbb{P}(Y \leq y)$$

for $x, y \in \mathbb{R}$, there exists a unique bi-variate copula $C : [0, 1]^2 \rightarrow [0, 1]$ such that

$$\mathbb{P}(X \leq x, Y \leq y) = C(F(x), G(y))$$

[1]. This famous result leads to a great amount of interest particularly in the copula C .

In the recent decades, there is a continuously growing literature related to copulas due to the fast development of risk management studies. Statistical and actuarial researchers are particularly interested in copulas because they usually have to deal with several random variables simultaneously but have very limited information about the relation among those random variables beforehand. As the copulas determine the dependence among random variables, it is not surprising that risk management researches focus on copulas in order to obtain seemingly independent categories of assets for risk diversification.

For the second situation, suppose the relation between two random variables are very well-known, for example, the two random variables are known to be comonotone. In this case no one would bother with the copula (because in this case we have $C(u, v) = \min(u, v)$ for $(u, v) \in [0, 1]^2$), instead, people may be more interested in monitoring the stationarity and reliability of this relationship. In this case, there is no longer risk diversification and hence the

copula plays no role. As a result, researchers in the fields such as system reliability may attempt to develop analysis tools without copulas.

Note that whether or not to include the copula C , we are actually playing with functions. If certain dimension reduction techniques may translate these functions into scalar indices without loss of too much information, it would be more useful to risk management in practice not only because scalars are comparable while functions are not, but also because estimators as well as their inference results may be available given the observed data.

1.1 Indices based on the copula

Indices constructed based on functions are invented based on functionals. If the dependence between X and Y is the research object, the indices usually defined solely by the copula C . In the studies of risk management, there are typically two types of such indices. The first type is defined via integrals on $[0, 1]^2$, including the rank-based coefficients of dependence such as the Spearman's ρ :

$$\rho := \rho(C) = 12 \int_0^1 \int_0^1 [C(u, v) - uv] du dv,$$

Gini's coefficient:

$$\gamma := \gamma(C) = 2 \int_0^1 \int_0^1 (|u + v - 1| - |u - v|) dC(u, v),$$

and the Kendall's τ :

$$\tau := \tau(C) = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1.$$

The rank-based coefficients of dependence are considered as remedies for the linear correlation coefficient [2]. The other type is defined via taking limits, including all *tail dependence coefficients* (TDC). The most widely used and well-studied measures of extreme co-movements are the tail dependence [3]:

$$\lambda := \lambda(C) = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u},$$

the weak tail dependence [4]:

$$\chi := \chi(C) = \lim_{u \rightarrow 0^+} \frac{2 \log u}{\log C(u, u)} - 1,$$

and the tail order $\kappa := \kappa(C)$ [5, 6] defined via

$$C(u, u) = \ell(u)u^\kappa$$

when $u \downarrow 0$, where ℓ is a slowly varying at 0 function [e.g., 7].

Both types of indices are available for risk management arrangements such as risk diversification. The integral type of indices might be considered when expected portfolio returns

are taken into account, for example, the portfolio optimization methods using rank correlation [8, pp.158-171]. If only the extreme co-movements of asset returns are of interest, then the limiting type of indices will be considered, for instance, the studies of financial contagion [9] at the market level and portfolio diversification [10] at the asset level.

1.2 Distribution free indices

When the distribution of (X, Y) is not of interest, in particular, if the relation between X and Y is systematically determined as

$$Y = h(X)$$

where h is a known function, then researches focus on the reliability of h , i.e. to check if $Y = h(X)$ really holds given the observed data. A typical example of this kind of problem is the reliability of an input-output system. For such kind of systems, Gribkova and Zitikis [11] propose an index that indicates the existence of anomalies if the index converge to 0.5 or converge to some value other than 0.5 otherwise. The idea is roughly as follows: suppose h has certain smoothness, for example, h is absolutely continuous with respect to the Lebesgue measure on some interval $[a_X, b_X]$ with its Radon-Nikodym derivative h^* satisfying

$$h(b) - h(a) = \int_a^b h^*(x) dx$$

for all $a_X \leq a \leq b \leq b_X$. Then the following fact:

$$\begin{cases} h_+^*(x) - h_-^*(x) = h^*(x) \\ h_+^*(x) + h_-^*(x) = |h^*(x)| \end{cases},$$

implies

$$h_+^*(x) = \frac{h^*(x) + |h^*(x)|}{2}$$

for every x on $[a_X, b_X]$. Hence we have

$$\frac{\int_{a_X}^{b_X} h_+^*(x) dx}{\int_{a_X}^{b_X} |h^*(x)| dx} = \frac{1}{2} \left(1 + \frac{\int_{a_X}^{b_X} h^*(x) dx}{\int_{a_X}^{b_X} |h^*(x)| dx} \right) = \frac{1}{2} \left(1 + \frac{h(b_X) - h(a_X)}{\int_{a_X}^{b_X} |h^*(x)| dx} \right).$$

If

$$\int_{a_X}^{b_X} |h^*(x)| dx = \infty$$

while

$$-\infty < h(b_X) - h(a_X) < \infty,$$

then we obtain

$$\frac{\int_{a_X}^{b_X} h_+^*(x) dx}{\int_{a_X}^{b_X} |h^*(x)| dx} = \frac{1}{2}.$$

Hence an index for anomaly detection in an input-output system is conceptually given by:

$$I := I(h) = \frac{\int_{a_X}^{b_X} h_+(x) dx}{\int_{a_X}^{b_X} |h^*(x)| dx}.$$

Note that in practice there is no need to validate the exact relationship h between the input X and the output Y when anomalies are the concern. The anomalies could be detected by examining whether some of the properties of h are violated. For example, if h is of bounded variation on $[a_X, b_X]$ while anomalies behave like Brownian motions, then these anomalies could be detected if the estimation of I is closed to 0.5.

1.3 Research objectives

This research focus on the non-parametric estimators for an improved variant of the tail order κ as well as the distribution free index I . In particular, we are interested in the asymptotic behaviors of our target estimators when our observed data are stationary bivariate time series. The reason for studying bivariate time series is because the observed data are usually financial time series when studying the dependence between risks while for input-output systems the observed data may also show certain temporal dependent patterns such as seasonality.

The recently developed variant of the tail order is the *tail order of maximal dependence* (TOMD) denoted as κ^* , which fixes the issue that the tail order may sometimes underestimate the tail dependence between two random variables. However, κ^* is not widely acknowledged in practice due to lack of nonparametric estimation methods, while other TDCs do have such kind of methods [e.g., 12–14, and references therein]. This research proposes such a non-parametric estimator for the TOMD and examines how this estimator works particularly for the bivariate time series data.

For distribution free index I , the non-parametric version is given by Gribkova and Zitikis [11], which also analyzes its asymptotic behaviors when the input-output pairs are identically independently distributed (i.i.d.). In this research, we focus on generalized their results to bivariate time series data.

The articles integrated include the following: first, we develop a non-parametric estimation method for κ^* and test its performance in Chapter 2; next, we shall test the performance of our proposed method using bivariate time series data in Chapter 3. In Chapter 4, we include our extension to the results of Gribkova and Zitikis [11]. Chapter 5 lists potential future work based on our findings.

Bibliography

- [1] M. Sklar, Distribution functions in dimensions and their margins, Publications de l'Institut Statistique de l'Université de Paris 8 (1959) 229–231. 1

- [2] A. J. McNeil, R. Frey, P. Embrechts, *Quantitative Risk Management*, Princeton University Press, 2015. 2
- [3] H. Joe, Parametric Families of Multivariate Distributions with Given Margins, *Journal of Multivariate Analysis* 46 (2) (1993) 262–282. 2
- [4] S. Coles, J. Heffernan, J. Tawn, Dependence Measures for Extreme Value Analyses, *Extremes* 2 (4) (1999) 339–365. 2
- [5] A. Ledford, Statistics for near independence in multivariate extreme values, *Biometrika* 83 (1) (1996) 169–187. 2
- [6] L. Hua, H. Joe, Tail order and intermediate tail dependence of multivariate copulas, *Journal of Multivariate Analysis* 102 (10) (2011) 1454–1471. 2
- [7] N. H. Bingham, C. M. Goldie, J. L. Teugels, *Regular Variation*, Cambridge University Press, 1987. 2
- [8] J. Wang (Ed.), *Encyclopedia of Business Analytics and Optimization*, IGI Global, 2014. 3
- [9] M. Pericoli, M. Sbracia, A Primer on Financial Contagion, *Journal of Economic Surveys* 17 (4) (2003) 571–608. 3
- [10] F. Durante, R. Pappadà, N. Torelli, Clustering of financial time series in risky scenarios, *Advances in Data Analysis and Classification* 8 (4) (2013) 359–376. 3
- [11] N. Gribkova, R. Zitikis, A user-friendly algorithm for detecting the influence of background risks on a model, *Risks* 6 (3) (2018) 100. 3, 4
- [12] G. Draisma, H. Drees, A. Ferreira, L. D. Haan, Bivariate tail estimation: dependence in asymptotic independence, *Bernoulli* 10 (2). 4
- [13] G. Frahm, M. Junker, R. Schmidt, Estimating the tail-dependence coefficient: Properties and pitfalls, *Insurance: Mathematics and Economics* 37 (1) (2005) 80–100.
- [14] P. Krupskii, H. Joe, Nonparametric estimation of multivariate tail probabilities and tail dependence coefficients, *Journal of Multivariate Analysis* 172 (2019) 147–161. 4

Chapter 2

A statistical methodology for assessing the maximal strength of tail dependence

2.1 Introduction

Extreme co-movements manifest in a variety of problems associated with risk management and are typically measured using tail dependence indices [e.g., 1–4, and references therein]. Definitions of these indices rely on the identity

$$C(u, u) = \ell(u)u^\kappa \quad \text{when } u \downarrow 0, \quad (2.1.1)$$

where $C : [0, 1]^2 \rightarrow [0, 1]$ is the bivariate copula of underlying dependent risks, ℓ is a slowly varying at 0 function [e.g., 5], and $\kappa \in [1, \infty)$ is the lower tail order, which we pedantically call the lower tail order of *diagonal* dependence. The corresponding upper tail order is obtained by replacing C in definition (2.1.1) by the corresponding survival copula, and for this reason we do not discuss this analogous “upper” case. Hence, throughout the rest of this chapter, we drop the qualifier “lower” and simply call κ the tail order of diagonal dependence (TODD).

Equation (2.1.1) describes asymptotic behavior of the copula C when its arguments shrink to the origin $(0, 0)$ along the *diagonal* path $(u, u)_{0 \leq u \leq 1}$, which may or may not reflect the maximal strength of tail dependence carried by the copula, as illustrated by Furman et al. [6]. Hence, it becomes natural to seek for a path (e.g., Figure 2.1), called path of maximal tail dependence (MTD), that gives rise to the tail order of *maximal* dependence, henceforth shorthanded as TOMD and denoted by κ^* . When the diagonal path is not an MTD path, as is the case for many copulas [6] including the Marshall-Olkin (M-O) copula, the classical measures of tail dependence based on equation (2.1.1) inevitably underestimate the actual (i.e., maximal) tail dependence. It therefore becomes natural and important (e.g., when assessing the strength of co-movements of financial instruments), to aim at assessing the strength of tail dependence based on an MTD path, which may or may not be unique, may or may not be diagonal.

Having said this, we ought to make it clear that the classical *diagonal*-based indices of tail dependence, such as κ noted above, are enormously useful by conveying a wealth of information, as elucidated in many scholarly writings, such as the monographs by Nelsen [7], Joe [4], Durante and Sempi [8], and Mcneil et al. [9]; see also the references therein. We also note in this regard the recently introduced notion of p -concentration, which is a diagonal-based

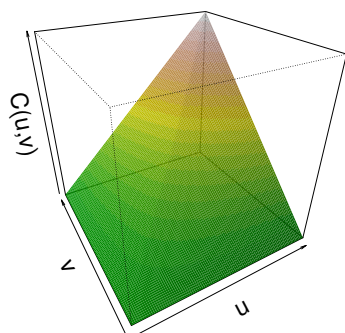
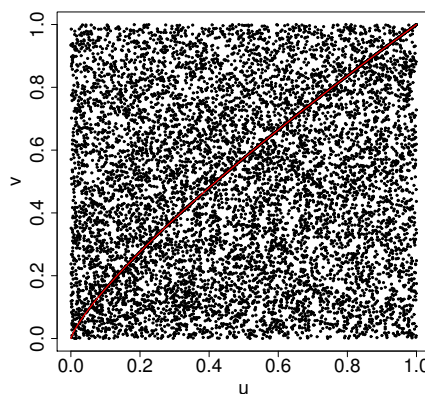
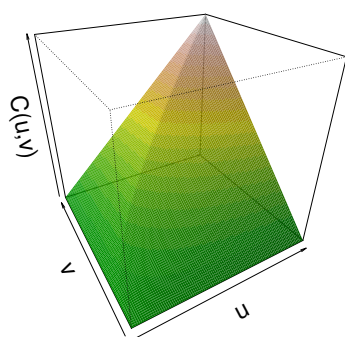
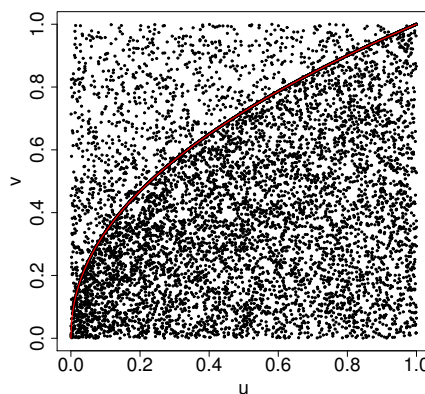
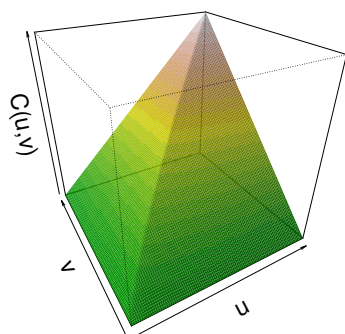
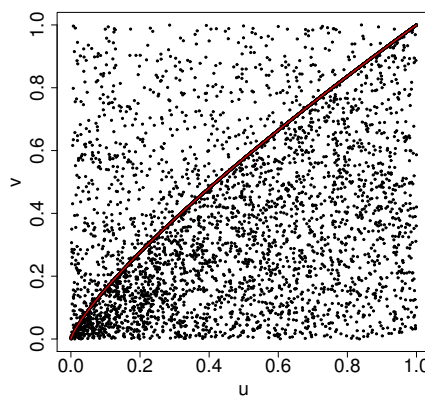
(a) M-O for $(a, b) = (0.18, 0.225)$.(b) MTD for $(a, b) = (0.18, 0.225)$.(c) M-O for $(a, b) = (0.3529, 0.75)$.(d) MTD for $(a, b) = (0.3529, 0.75)$.(e) M-O for $(a, b) = (0.72, 0.9)$.(f) MTD for $(a, b) = (0.72, 0.9)$.

Figure 2.1: The M-O copula (left) and its MTD path (right) on a 10,000 point scatterplot for the specified M-O parameters (a, b) with $\kappa^* = 1.80, 1.52, \text{ and } 1.20$ (top to bottom).

measure of dependence that plays a pivotal role in characterizing the risk measure Expected Shortfall [10].

The aforementioned κ as well as other tail dependence indices associated with equation (2.1.1) have empirical estimators and well-established statistical inference theories [e.g., 11–13, and references therein]. The tail order κ^* , as far as we are aware of, does not yet have an empirical estimator and thus statistical inference theory. This hinders practical use of κ^* when assessing the strength of tail dependence. The purpose of this chapter is to construct an empirical estimator of the tail order κ^* and discuss its practical implementation.

The main idea behind the construction hinges on the realization that it is possible to circumvent the need for having closed-form expressions for MTD paths. Explaining how this can be done makes up the contents of Section 2.2, where we gradually introduce and discuss the estimator. In Section 2.3 we explore statistical properties of the estimator, including its asymptotic properties, bias and its reduction. In the same section, we also test the performance of the estimator on simulated independent and identically distributed (iid) pairs that arise from the Marshall-Olkin copula. In Section 2.4 we explain how to mitigate the influence of slowly varying functions, such as ℓ in equation (2.1.1), on the TOMD estimator, where we also illustrate the technique in a simulation study based on the generalized Clayton copula. Section 2.5 concludes this chapter with a summary of main contributions and suggestions for future research. Proofs and other technicalities with illustrative graphs are in Appendix 2.5.

2.2 Estimating the tail order of maximal dependence

This section is subdivided into three parts. In Section 2.2.1 we recall the definition of the tail order κ^* of *maximal* dependence, which we call TOMD for short. It is a counterpart of the earlier discussed tail order κ of *diagonal* dependence, TODD for short. In Section 2.2.2 we introduce the main idea of estimating κ^* without relying on explicit expressions of the underlying paths of maximal tail dependence. In Section 2.2.3 we propose an empirical estimator of κ^* .

2.2.1 Preliminaries: the TOMD

Let $C : [0, 1]^2 \rightarrow [0, 1]$ be any copula. By definition [6], a path of tail dependence is $(\varphi(u), u^2/\varphi(u))_{0 \leq u \leq 1}$, where $\varphi : [0, 1] \rightarrow [0, 1]$ is a function that satisfies the following admissibility conditions:

1. $\varphi(u) \in [u^2, 1]$ for every $u \in [0, 1]$;
2. $\varphi(u) \rightarrow 0$ and $u^2/\varphi(u) \rightarrow 0$ when $u \downarrow 0$.

For example, $\varphi(u) = u$ is such a function, and it leads to the diagonal path $(u, u)_{0 \leq u \leq 1}$ that we mentioned earlier in relation to the classical tail order κ .

Let \mathcal{A} denote the set of all admissible functions. An MTD path, which is not necessarily unique, is $(\varphi^*(u), u^2/\varphi^*(u))_{0 \leq u \leq 1}$ that arises from $\varphi^* \in \mathcal{A}$ given by

$$\varphi^*(u) = \arg \max_{\varphi \in \mathcal{A}} C(\varphi(u), u^2/\varphi(u)).$$

Denote

$$\Pi^*(u) = C(\varphi^*(u), u^2/\varphi^*(u)).$$

If there exists κ^* such that

$$\Pi^*(u) = \ell^*(u)u^{\kappa^*} \quad \text{when } u \downarrow 0 \quad (2.2.1)$$

for a slowly varying at 0 function ℓ^* , then κ^* is called [6] the tail order of *maximal* dependence (TOMD). As illustrated in the aforementioned paper, there can be several MTD paths but there is only one TOMD. We also know from the paper that the TOMD may or may not coincide with the TODD, and the Gaussian copula is one of those rare cases when the two tail orders are the same [14].

2.2.2 Estimating the TOMD: an idea

To introduce an idea for estimating the TOMD, we focus on the case

$$\Pi^*(u) = u^{\kappa^*}, \quad u \in [0, 1], \quad (2.2.2)$$

where $\kappa^* \geq 1$ is a parameter. That is, for the sake of simplicity, but only until Section 2.4, we drop the slowly varying function ℓ^* from our main considerations, although we shall touch upon a practical aspect of this simplification sooner, at the end of the current section.

Property (2.2.2) is of course satisfied by the Marshall-Olkin (M-O) copula

$$C_{a,b}(u, v) = \min(u^{1-a}v, uv^{1-b}), \quad u, v \in [0, 1], \quad (2.2.3)$$

where $a, b \in [0, 1]$ are parameters, and we shall therefore conveniently use it in a simulation study, whose results will be reported in Section 2.3.2. Recall that the TOMD of this copula is [6]

$$\kappa^* = 2 - \frac{2ab}{a+b}. \quad (2.2.4)$$

For theory and applications of the M-O copula with further references, we refer to Cherubini et al. [15, 16], and Mcneil et al. [9]. For recent actuarial applications of the copula with generalizations and related references, we refer to Su and Furman [17].

Next, under assumption (2.2.2), we have

$$\kappa^* = \frac{\log \Pi^*(u)}{\log u} \leq \frac{\log C(\varphi(u), u^2/\varphi(u))}{\log u} \quad (2.2.5)$$

for all $u \in [0, 1]$ and every $\varphi \in \mathcal{A}$. With the notation $x = \varphi(u)$ and $y = u^2/\varphi(u)$, we therefore have

$$\kappa^* \leq \frac{\log C(x, y)}{\log \sqrt{xy}} = \frac{2 \log C(x, y)}{\log x + \log y}$$

for all $x, y \in [0, 1]$ such that $xy = u^2$. The latter bound suggests that given random pairs $(U_1, V_1), \dots, (U_n, V_n) \sim C$, which we assume to be independent throughout this chapter, the TOMD κ^* under model (2.2.2) can be estimated by

$$\hat{\kappa}_n^* := \min_{i=1, \dots, n} \frac{2 \log C_n(U_i, V_i)}{\log U_i + \log V_i}, \quad (2.2.6)$$

where

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{1} \{U_i \leq u, V_i \leq v\}$$

with $\mathbb{1}$ denoting the indicator function.

We shall find it more convenient, however, to work with a slight modification of the estimator $\hat{\kappa}_n^*$. Namely, let

$$\hat{\kappa}_n^*(\theta) := \min_{i=1, \dots, n} \frac{2T_\theta \circ C_n(U_i, V_i)}{\log U_i + \log V_i}, \quad (2.2.7)$$

where

$$T_\theta(t) = \begin{cases} (t^\theta - 1)/\theta & \text{when } \theta > 0, \\ \log(t) & \text{when } \theta = 0, \end{cases}$$

for all $t \in (0, 1]$. In the following paragraph we shall explain the technical aspect that has lead us to the introduction of the function T_θ into considerations.

The estimator $\hat{\kappa}_n^*(\theta)$ is the ratio of a quantity based on T_θ for some small $\theta > 0$ in the numerator, and a quantity based on $T_0 = \log$ in the denominator. Having this combination of T_θ 's does not affect the numerical performance of the estimator but considerably simplifies the proofs. To illustrate the technical challenges that we face, we can just look at the uniform empirical distribution function $C_n(u, 1)$, $0 < u < 1$. To assess how close $T_\theta \circ C_n(u, 1)$ and $T_\theta \circ C(u, 1)$ are, we employ the mean value theorem and look at the smallness of

$$\frac{|C_n(u, 1) - C(u, 1)|}{C(u, 1)^{1-\theta}(1 - C(u, 1))^{1-\theta}},$$

uniformly over $0 < u < 1$. When $n \rightarrow \infty$, we have uniform convergence to 0 as long as $\theta > 0$, but there cannot be such convergence when $\theta = 0$. These facts follow immediately from classical results of Lai [18] on the weighted Glivenko-Cantelli Theorem. For a multivariate version of this theorem, we refer to Mason [19]. For further references and advances on the topic, we refer to Berghaus et al. [20].

We are now ready to discuss practical implications of working with model (2.2.2) instead of (2.2.1). As we have seen, the simplification has helped us to elucidate the main idea of the estimator, but from the practical point of view, one would of course like to work with a great variety of copulas, and not just with the M-O copula, or its symmetrized version [e.g., 6]. To see how to overcome this obstacle, we go back to general model (2.2.1) for which statement (2.2.5) turns into

$$\begin{aligned} \kappa^* &= \frac{\log \Pi^*(u) - \log \ell(u)}{\log u} \\ &\leq \frac{\log C(\varphi(u), u^2/\varphi(u))}{\log u} - \frac{\log \ell(u)}{\log u} \approx \frac{\log C(\varphi(u), u^2/\varphi(u))}{\log u} \end{aligned} \quad (2.2.8)$$

with the approximation holding for sufficiently small $u > 0$, due to [e.g., 5, Proposition 1.3.6(i), p. 16]

$$\varrho(u) := \frac{\log \ell(u)}{\log u} \rightarrow 0 \quad \text{when } u \downarrow 0. \quad (2.2.9)$$

The right-hand side of (2.2.8) is the same as that of (2.2.5). Hence, the above introduced estimator of TOMD can work for general copulas as long as the ratio $\varrho(u)$ is small, for which $u > 0$ has to be small. This can be achieved by using only those random pairs $(U_1, V_1), \dots, (U_n, V_n)$ that are close to the origin $(0, 0)$. We shall therefore employ a conditioning argument later in this chapter.

We should point out right at the outset that the restriction of data to only those pairs that are close to the origin $(0, 0)$ is not an issue from the theoretical point of view but becomes a challenge from the practical point of view. Indeed, the closer we are to the origin, the fewer random pairs we have, and thus we possibly have a large statistical estimation error. If, however, we use all n observations, then the statistical error becomes minimal but the model bias due to possibly large ratio $\varrho(u)$ might be large, unless we of course deal with the M-O copula, for which $\varrho(u) = 0$ for all $u \in (0, 1)$. Hence, when dealing with real data, a very delicate balancing act needs to be accomplished, but the task is certainly doable as we shall see in Section 2.4.

2.2.3 In pursuit of efficiency: an improved estimator

Simulations have shown that the estimators $\hat{\kappa}_n^*$ and $\hat{\kappa}_n^*(\theta)$ do work, but they are rather inefficient. For this reason, we next introduce a modification of $\hat{\kappa}_n^*(\theta)$.

Let the pairs $(U_1, V_1), \dots, (U_n, V_n)$ be iid, and since we are working under model (2.2.2), we can utilize all the pairs. (Under model (2.2.1), we would need to restrict ourselves to only those pairs that are in a neighbourhood of $(0, 0)$; the smaller the neighbourhood, the smaller the bias.) Next we choose any fixed $m \in \{1, \dots, n\}$ and randomly assign the pairs $(U_1, V_1), \dots, (U_n, V_n)$ into $k = \lceil n/m \rceil$ groups, so that there can be at most one group (perhaps none) with less than m pairs, whereas all the other groups contain exactly m pairs. Denote these k groups by G_1, \dots, G_k . We define *the average block-minima estimator* of the TOMD κ^* by

$$\hat{\kappa}_n^*(m, \theta) = \frac{1}{k} \sum_{j=1}^k \min_{i \in G_j} \frac{2T_\theta \circ C_n(U_i, V_i)}{\log U_i + \log V_i}. \quad (2.2.10)$$

Note that the earlier introduced estimator $\hat{\kappa}_n^*(\theta)$ is $\hat{\kappa}_n^*(m, \theta)$ when $m = n$, and thus $k = 1$. In view of the aforementioned deficiencies of $\hat{\kappa}_n^*(\theta)$, however, in what follows (unless explicitly noted otherwise) we work with fixed m , which can in principle be allowed to depend on n , but this would require more complex arguments, which we wish to avoid at this stage of our research.

In a nutshell, the estimator $\hat{\kappa}_n^*(m, \theta)$ works because:

1. for every integer $m \geq 1$ and real $\theta > 0$, and when $n \rightarrow \infty$, the estimator $\hat{\kappa}_n^*(m, \theta)$ converges to the limit

$$\kappa^*(m, \theta) = \mathbb{E} \left[\min_{i=1, \dots, m} \frac{2T_\theta \circ C(U_i, V_i)}{\log U_i + \log V_i} \right]; \quad (2.2.11)$$

2. the TOMD κ^* can be written as

$$\kappa^* = \text{ess inf} \frac{2 \log C(U, V)}{\log U + \log V} \approx \text{ess inf} \frac{2T_\theta \circ C(U, V)}{\log U + \log V}, \quad (2.2.12)$$

with the approximation holding for small $\theta > 0$.

Thus, κ^* can be approximated by $\kappa^*(m, \theta)$ for sufficiently large $m \geq 1$ and small $\theta > 0$.

Rigorous formulations and assumptions needed for the validity of the above two facts make up the contents of two theorems in the next section. Namely, Theorem 2.3.1 deals with asymptotic results, whereas Theorem 2.3.5 deals with bias reduction. Their proofs, which are rather involved, are in Appendix 2.5, where we also placed graphs that visualize the asymptotic distribution (i.e., normality) of the estimator.

2.3 Estimator's statistical properties and performance

We split this section into two parts: Section 2.3.1 deals with theoretical aspects of the estimator $\hat{\kappa}_n^*(m, \theta)$, and Section 2.3.2 illustrates the estimator's performance on simulated data.

2.3.1 Theoretical results

Throughout, we use the function

$$g(u, v) = \frac{\mathbb{1}\{0 < C(u, v) < 1\}}{C(u, v)(1 - C(u, v))},$$

which is defined for all $u, v \in [0, 1]$.

Theorem 2.3.1 *Let C be positively quadrant dependent (PQD), that is, $C(u, v) \geq uv$ for every $u, v \in [0, 1]$, and let $\theta \in (0, 1)$ be any constant. If $\mathbb{E}[g(U, V)^{1-\theta}] < \infty$, then for every integer $m \geq 1$, we have*

$$\hat{\kappa}_n^*(m, \theta) \xrightarrow{p} \kappa^*(m, \theta) \quad (2.3.1)$$

when $n \rightarrow \infty$, where \xrightarrow{p} denotes convergence in probability.

We next discuss how to verify the finiteness of $\mathbb{E}[g(U_1, V_1)^{1-\theta}]$, for which we employ Kendall's cumulative distribution function (cdf) K_C [21]. It is the cdf of the random variable $C(U, V)$ when the pair (U, V) follows the joint cdf C . For example, when C is the M-O copula, the Kendall's cdf is (details in Appendix 2.A)

$$K_{a,b}(t) = t - \left(2 - \frac{2}{\kappa^*}\right)t \log(t) \quad (2.3.2)$$

for all $t \in (0, 1]$, where κ^* is the TOMD given equation (2.2.4).

Note 2.3.2 *The appearance of the TOMD κ^* and not of the classical TODD $\kappa = \min\{a, b\}$ in formula (2.3.2) for $K_{a,b}$ lends additional support for the naturalness of the TOMD as an index of tail dependence.*

Lemma 2.3.3 *Let K_C denote the Kendall's cdf, and let $\tau \in (0, 1)$ be any constant. If the integral $\int_{(0,1]} t^{-\tau} dK_C(t)$ is finite, then the moment $\mathbb{E}[g(U, V)^\tau]$ is finite, where the random pair (U, V) follows the joint cdf C .*

To illustrate, when C is the comonotonic copula, we have $\Pi^*(u) = u$ and $K_C(t) = t$, and when C is the independence copula, we have $\Pi^*(u) = u^2$ and $K_C(t) = t - t \log t$ [22]. In both cases, Lemma 2.3.3 and thus Theorem 2.3.1 apply. The following corollary to Lemma 2.3.3 is convenient.

Corollary 2.3.4 *If the derivative K'_C of the Kendall's cdf exists in a neighbourhood of 0 and is regularly varying of order $\alpha \in (-1, 0]$, that is, $K'_C(t) = t^\alpha \ell_C(t)$, where ℓ_C is a slowly varying at 0 function, then the moment $\mathbb{E}[g(U, V)^\tau]$ is finite for all $\tau \in (0, 1 + \alpha)$.*

Theorem 2.3.1 does not imply that $\hat{\kappa}_n^*(m, \theta)$ is a consistent estimator of κ^* because $\kappa^*(m, \theta)$ is not generally equal to κ^* . Nevertheless, the following theorem shows that $\kappa^*(m, \theta)$ approaches κ^* when m gets sufficiently large and $\theta > 0$ sufficiently small. This may suggest to use the maximally large m , that is, $m = n$, but this would lead to $k = \lceil n/m \rceil = 1$ and thus turn the estimator $\hat{\kappa}_n^*(m, \theta)$ into just one summand, thus effectively eliminating the important averaging effect that reduces the estimator's statistical error.

Theorem 2.3.5 *Let C be PQD, and let the functions $\varphi^*(u)$ and $\psi^*(u) := u^2/\varphi^*(u)$ be strictly increasing. Then $\kappa^*(m, \theta)$ can be made as close to κ^* as desired by taking sufficiently large $m \geq 1$ and sufficiently small $\theta > 0$.*

We next illustrate these theoretical results on simulated data.

2.3.2 An illustrative simulation study

To illustrate the performance of $\hat{\kappa}_n^*(m, \theta)$, we simulate data from the M-O copula (equation (2.2.3)) whose TOMD is given by equation (2.2.4). Hence, $\Pi_{a,b}^*(u) = u^{\kappa^*}$. Since

$$K'_{a,b}(t) = \left(\frac{2}{\kappa^*} - 1\right) - \left(2 - \frac{2}{\kappa^*}\right) \log(t) \quad (2.3.3)$$

is slowly varying at 0, Theorem 2.3.1 applies. Furthermore, since the M-O copula is PQD, Theorem 2.3.5 also applies. Based on these results, and always setting the parameter θ to

$$\theta_0 := 10^{-6},$$

we expect the simulated values of $\hat{\kappa}_n^*(m, \theta_0)$ to be close to the true theoretical ones. We verify this conjecture using a simulated experiment.

We set the M-O parameter (a, b) to the values specified in Table 2.1, with the corresponding

(a, b)	(0.18, 0.225)	(0.3529, 0.75)	(0.72, 0.9)
κ^*	1.80	1.52	1.20
MTD strength	weak	medium	strong

Table 2.1: The parameter (a, b) choices with κ^* values and MTD strength.

values of κ^* calculated according to equation (2.2.4). Our adopted classification of the MTD strength according to the values of $\kappa^* \in [1, 2]$ is of course subjective: we say that MTD is

weak when $\kappa^* \in [5/3, 2]$, medium when $\kappa^* \in [4/3, 5/3]$, and strong when $\kappa^* \in [1, 4/3]$. To study the effect of m on the estimator $\hat{\kappa}_n^*(m, \theta_0)$, we set $m = 3, 5, 7, 10$, and 20 . We measure the estimator's performance using the mean (Mean) and the standard deviation (StDev) based on 1,000 simulated values of $\hat{\kappa}_n^*(m, \theta_0)$. We also calculate the p -values based on the Anderson–Darling (A-D) and Cramér–von Mises (C-vM) goodness-of-fit tests when the simulated values of $\hat{\kappa}_n^*(m, \theta_0)$ are fitted to the normal distribution. The results for the sample sizes $n = 10,000$ and $n = 20,000$ are reported in Tables 2.2–2.3. The corresponding histograms and fitted

m	$k = \lceil n/m \rceil$	κ^*	Mean	StDev	A-D	C-vM
3	3334	1.80	1.8267	0.0186	0.7448	0.6912
		1.52	1.5466	0.0172	0.7114	0.5860
		1.20	1.1946	0.0151	0.6905	0.7290
5	2000	1.80	1.8073	0.0191	0.6053	0.5469
		1.52	1.5182	0.0181	0.3325	0.2770
		1.20	1.1819	0.0165	0.1222	0.1910
7	1429	1.80	1.7969	0.0195	0.4578	0.4273
		1.52	1.5065	0.0190	0.1679	0.1476
		1.20	1.1761	0.0179	0.0221	0.0469
10	1000	1.80	1.7872	0.0200	0.2899	0.2859
		1.52	1.4972	0.0203	0.0497	0.0502
		1.20	1.1700	0.0200	0.0026	0.0075
20	500	1.80	1.7700	0.0219	0.0790	0.0959
		1.52	1.4817	0.0244	0.0018	0.0040
		1.20	1.1563	0.0258	0.0000	0.0002

Table 2.2: Summary statistics for $\hat{\kappa}_n^*(m, \theta_0)$ based on iid data of size $n = 10,000$ with the p -values of the A-D and C-vM tests for normality shaded when they fall below 0.05.

normal densities are depicted in Figures 2.6–2.7. Based on these numerical results and graphs, we make the following observations.

First, the simulated means are quite close to the corresponding values of κ^* . Note also that the differences between them get larger when the values of κ^* get smaller, which may suggest that the stronger the tail dependence is, the more observations are needed to accurately estimate κ^* . Tables 2.2–2.3 suggest that $m = 3$ is not sufficiently large for the cases $\kappa^* = 1.80$ and $\kappa^* = 1.52$, because the differences between the simulated means and the theoretical values do not seem to diminish when the sample size increases. The choice $m = 5$ seems appropriate as it leads to fairly accurate estimates of κ^* in all the cases considered. Setting m larger than 5 does not seem to significantly improve the estimates. Furthermore, we see from Figures 2.6–2.7 that the simulated distributions of $\hat{\kappa}_n^*(m, \theta_0)$ exhibit uni-modality, even normality, which suggests that the simulated means get reasonably close to the theoretical values of κ^* .

m	$k = \lceil n/m \rceil$	κ^*	Mean	StDev	A-D	C-vM
3	6667	1.80	1.8296	0.0132	0.7623	0.6902
		1.52	1.5493	0.0124	0.9995	0.9984
		1.20	1.1983	0.0109	0.4018	0.3724
5	4000	1.80	1.8117	0.0136	0.8813	0.8560
		1.52	1.5226	0.0132	0.9594	0.9294
		1.20	1.1878	0.0120	0.0515	0.0608
7	2857	1.80	1.8025	0.0139	0.8974	0.8980
		1.52	1.5125	0.0138	0.5907	0.6090
		1.20	1.1836	0.0130	0.0077	0.0128
10	2000	1.80	1.7944	0.0144	0.9443	0.9683
		1.52	1.5052	0.0148	0.1778	0.2272
		1.20	1.1793	0.0145	0.0006	0.0017
20	1000	1.80	1.7813	0.0158	0.8076	0.8567
		1.52	1.4938	0.0178	0.0032	0.0088
		1.20	1.1697	0.0187	0.0000	0.0000

Table 2.3: Summary statistics for $\hat{\kappa}_n^*(m, \theta_0)$ based on iid data of size $n = 20,000$ with the p -values of the A-D and C-vM tests for normality shaded when they fall below 0.05.

2.4 Estimating the TOMD when ℓ is present

In this section, we adjust the methodology discussed in the previous sections in order to estimate TOMD when, unlike in the case of the M-O copula, a slowly varying function might be present. At the core of this adjustment is a conditioning argument. Namely, we restrict ourselves to only those observations that are near the origin $(0, 0)$, which effectively shifts our attention from the original copula to the one conditioned on observations below certain pre-specified thresholds. Since the conditional copula is the ratio of probabilities, this leads us – via equation (2.2.1) – to a ratio of the type $\ell(qu)/\ell(q)$, where $q > 0$ is a small parameter whose value could, for example, be a confidence level set by convention (e.g., 0.05), or by regulation [e.g., 23, 24]. The ratio $\ell^*(qu)/\ell^*(q)$ is close to 1 when $q > 0$ is small, due to the very definition of slowly varying at 0 function [e.g., 5, Definition, p. 6].

2.4.1 Estimating the TOMD

In view of the equation $\Pi^*(u) = \ell^*(u)u^{\kappa^*}$ when $u \downarrow 0$, the conditional maximal tail probability is

$$\frac{C(\varphi^*(qu), q^2 u^2 / \varphi^*(qu))}{C(\varphi^*(q), q^2 / \varphi^*(q))} = \frac{\ell^*(qu)(qu)^{\kappa^*}}{\ell^*(q)q^{\kappa^*}} \approx u^{\kappa^*}$$

when $0 < q \approx 0$. Hence, for any φ , we have the approximate bound

$$\frac{C(\varphi(qu), q^2 u^2 / \varphi(qu))}{C(\varphi(q), q^2 / \varphi(q))} \lesssim u^{\kappa^*}.$$

By letting $\tilde{u} = \varphi(qu)/\varphi^*(q)$ and $\tilde{v} = u^2\varphi^*(q)/\varphi(qu)$, where

$$\varphi^*(q) = \arg \max_x C(x, q^2/x),$$

the above approximate bound turns into

$$\frac{C(\varphi^*(q)\tilde{u}, q^2\tilde{v}/\varphi^*(q))}{C(\varphi^*(q), q^2/\varphi^*(q))} \lesssim u^{\kappa^*}.$$

Note the equation $u^{\kappa^*} = (\tilde{u}\tilde{v})^{\kappa^*/2}$. With the notation

$$F_q^*(\tilde{u}, \tilde{v}) := \frac{C(\varphi^*(q)\tilde{u}, q^2\tilde{v}/\varphi^*(q))}{C(\varphi^*(q), q^2/\varphi^*(q))},$$

we arrive at the bound

$$\kappa^* \lesssim \frac{2 \log F_q^*(\tilde{u}, \tilde{v})}{\log \tilde{u} + \log \tilde{v}}.$$

Hence, given random pairs $(U_1, V_1), \dots, (U_n, V_n)$ and a small risk level $q > 0$, the TOMD κ^* can be estimated via the following procedure:

1. using all the pairs available in the entire unit square $[0, 1]^2$, compute an estimate of $\varphi^*(q)$ by maximizing $C_n(x, q^2/x)$ with respect to $x \in [q^2, 1]$, where C_n is the empirical copula constructed from all the n pairs that are available in the entire unit square $[0, 1]^2$;
2. extract the set $\mathcal{M}_{q,n}$ of all those pairs (U_i, V_i) that are in the rectangle $[0, \varphi^*(q)] \times [0, q^2/\varphi^*(q)]$, and let $m_{q,n} := \#(\mathcal{M}_{q,n})$ denote the number of pairs in $\mathcal{M}_{q,n}$;
3. randomly assign the pairs that are in $\mathcal{M}_{q,n}$ into $\lceil m_{q,n}/m \rceil$ disjoint groups, whose index sets we denote by $G_1, \dots, G_{\lceil m_{q,n}/m \rceil} \subseteq \{1, 2, \dots, n\}$, and they are such that there is at most one group with less than m elements, with all the other groups having exactly m elements;
4. compute the average block-minima estimator of the TOMD κ^* by the formula

$$\hat{\kappa}_{m_{q,n}}^*(m, \theta, q) = \frac{1}{\lceil m_{q,n}/m \rceil} \sum_{j=1}^{\lceil m_{q,n}/m \rceil} \min_{i \in G_j} \frac{2T_\theta \circ F_{q, \mathcal{M}_{q,n}}^*(\tilde{U}_i, \tilde{V}_i)}{\log \tilde{U}_i + \log \tilde{V}_i},$$

where $\tilde{U}_i = U_i/\varphi^*(q)$, $\tilde{V}_i = V_i\varphi^*(q)/q^2$, and

$$F_{q, \mathcal{M}_{q,n}}^*(u, v) = \frac{1}{m_{q,n}} \sum_{(U_k, V_k) \in \mathcal{M}_{q,n}} \mathbb{1}\{\tilde{U}_k \leq u, \tilde{V}_k \leq v\}.$$

2.4.2 An illustrative simulation study

To demonstrate how the estimator $\hat{\kappa}_{m_q}^*(m, \theta, q)$ works in practice, we simulate $n = 500,000$ independent pairs $(U_i, V_i) \sim C_{\gamma_0, \gamma_1}$ from the generalized Clayton copula

$$C_{\gamma_0, \gamma_1}(u, v) = u^{\gamma_1/\gamma^*} \left(u^{-1/\gamma^*} + v^{-1/\gamma_0} - 1 \right)^{-\gamma_0}, \quad u, v \in [0, 1],$$

where $\gamma_0 > 0$ and $\gamma_1 \geq 0$ are the model parameters, with their values specified in Table 2.4, and $\gamma^* := \gamma_0 + \gamma_1$. As calculated by [6], the TOMD for this copula is

$$\kappa^* = 1 + \frac{\gamma_1}{\gamma_1 + 2\gamma_0}.$$

We repeat the procedure 1,000 times and thus obtain 1,000 values of $\hat{\kappa}_{m_q}^*(m, \theta_0, q)$ with the

(γ_0, γ_1)	κ^*	Mean	StDev	A-D	C-vM
(0.1, 0.8)	1.8	1.7914	0.0366	0.9565	0.9507
(0.4, 0.8)	1.5	1.5071	0.0219	0.9596	0.9183
(0.4, 0.2)	1.2	1.2086	0.0127	0.8601	0.8792

Table 2.4: Summary of simulation results when $m = 5$ and $q = 0.05$.

parameter values set to $m = 5$ and $q = 0.05$. Summary statistics are reported in Table 2.4, from which we see that the A-D and C-vM goodness-of-fit p -values retain the null hypothesis of normality. Fits of the estimator values to the normal distribution are depicted in Figure 2.2.

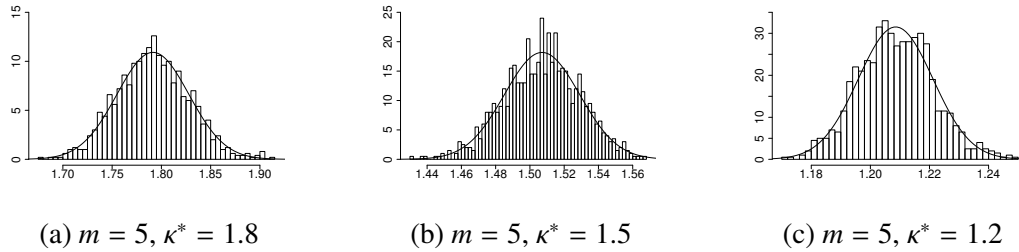


Figure 2.2: Fits of simulated $\hat{\kappa}_{m_q}^*(m, \theta_0, q)$ values to the normal distribution when $m = 5$ and $q = 0.05$.

For comparison, using the same parameter (γ_0, γ_1) values as before, and the same $m = 5$ but now with the increased threshold $q = 0.1$ and thus a larger m_q , we obtain 1,000 values of the TOMD estimator $\hat{\kappa}_{m_q}^*(m, \theta_0, q)$ whose summary statistics are reported in Table 2.5. The

(γ_0, γ_1)	κ^*	Mean	StDev	A-D	C-vM
(0.1, 0.8)	1.8	1.8072	0.0200	0.9436	0.8645
(0.4, 0.8)	1.5	1.5142	0.0136	0.7527	0.7505
(0.4, 0.2)	1.2	1.2111	0.0081	0.7281	0.7883

Table 2.5: Summary of simulation results when $m = 5$ and $q = 0.1$.

A-D and C-vM goodness-of-fit p -values retain the null hypothesis of normality. Fits of the estimator values to the normal distribution are depicted in Figure 2.3.

By increasing the risk level q from 0.05 to 0.1 we have increased the number m_q of pairs and thus potentially reduced the estimator's statistical error, but this could have increased the deterministic bias tackled in Theorem 2.3.5. As we have noted earlier, it is a delicate task to strike the right balance between the competing forces of having small $q > 0$ and large m_q .

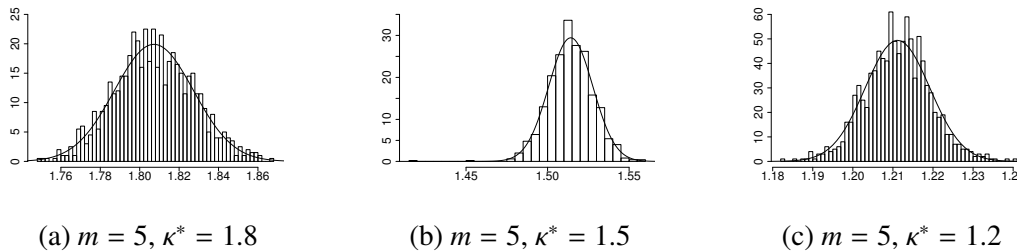


Figure 2.3: Fits of simulated $\hat{\kappa}_{m_q}^*(m, \theta_0, q)$ values to the normal distribution when $m = 5$ and $q = 0.1$.

2.5 Conclusion

We have proposed an empirical estimator of the tail order of maximal dependence. The estimator successfully avoids the complexity of deriving explicit formulas of the underlying paths of maximal tail dependence. To test the estimator's practical performance, we have conducted a simulation study and discussed our findings. There are still a number of interesting problems for future research, including these:

1. Explore estimator's statistical properties and practical performance under various dynamical DGP's (i.e., data generating processes) such as those arising from short- and long-range memory time series.
2. Relax the currently imposed fixedness of m by describing those sequences $m = m_n \in \{1, \dots, n\}$ which grow to infinity together with $n \rightarrow \infty$ and maintain asymptotic unbiasedness of the proposed estimator.
3. When searching for efficient and reliable estimators, explore different block designs and their combinations using, e.g., generalized means, as illustrated by Vovk and Wang [25] in a statistically different but philosophically closely related context.
4. Explore the roles of the functions T_θ and $T_0 = \log$ when defining estimators.
5. Derive asymptotic normality/distribution of the estimator, and explore ways for practical assessment of its standard error.
6. In addition to κ^* , [6] also discuss other indices related to maximal tail dependence, such as λ_L^* and χ_L^* . We have indications, but no proofs yet, that techniques of this chapter could be adapted to estimate those indices as well.

Appendix

In Appendix 2.A we present a number of technical results that will later be used for proving Theorems 2.3.1–2.3.5 in Appendix 2.B.

2.A Auxiliary results

We start with three calculus-type propositions, which are included here to make the proofs in subsequent appendices more self-contained and thus easier checked.

Proposition 2.A.1 *For all real $x_i, y_i \in \mathbb{R}$ and integer $m \geq 1$, we have*

$$\left| \min_{i=1, \dots, m} x_i - \min_{i=1, \dots, m} y_i \right| \leq \max_{i=1, \dots, m} |x_i - y_i|.$$

Proof Let j and k be those that satisfy the equations $\min_{i=1, \dots, m} x_i = x_j$ and $\min_{i=1, \dots, m} y_i = y_k$. If $x_j \geq y_k$, then $|x_j - y_k| = x_j - y_k \leq x_k - y_k$, but if $x_j < y_k$, then $|x_j - y_k| = y_k - x_j \leq y_j - x_j$. This finishes the proof.

Proposition 2.A.2 *Let $\theta \in (0, 1)$. The bound $x^{1-\theta}|y^\theta - x^\theta| \leq |y - x|$ holds for all $x, y \in (0, \infty)$.*

Proof Denote $\Delta := x^{1-\theta}|y^\theta - x^\theta|$. If $y > x$, then $\Delta = x^{1-\theta}y^\theta - x \leq y - x = |y - x|$, but if $x > y$, then $\Delta = x - x^{1-\theta}y^\theta \leq x - y = |y - x|$. This finishes the proof.

Proposition 2.A.3 *For all $x \leq 0$, we have $|e^x - 1 - x| \leq 0.5x^2$.*

Proof The functions $f(x) = e^x - 1 - x$ and $g(x) = 0.5x^2$ are non-negative, and satisfy $f(0) = 0 = g(0)$. Furthermore, their difference $g(x) - f(x)$ is non-increasing. Hence,

$$g(x) - |f(x)| = g(x) - f(x) \geq g(0) - f(0) = 0,$$

which completes the proof.

For any copula C , its Kendall's cdf is, by definition, the distribution of the random variable $C(U, V)$, where (U, V) follows the joint cdf C .

Proposition 2.A.4 *The Kendall's cdf $K_{a,b}$ of the M-O copula $C_{a,b}$ is given by formula (2.3.2).*

Proof Let (U, V) denote the random pair whose joint cdf is the M-O copula $C_{a,b}$. Then [e.g., 15, p. 128]

$$\mathbb{P}(U \leq u \mid V = v) = u^{1-a} \mathbb{1}\{u \in [v^{b/a}, 1]\} + (1-b)uv^{-b} \mathbb{1}\{u \in [0, v^{b/a}]\}$$

for all $u, v \in [0, 1]$. Using this equation, we next derive formula (2.3.2) for the Kendall's cdf $K_{a,b}$ by establishing the corresponding formula for the Kendall's survival function $\bar{K}_{a,b} = 1 - K_{a,b}$. Namely, for $t \in (0, 1)$, we have

$$\begin{aligned} \bar{K}_{a,b}(t) &= \mathbb{P}(C_{a,b}(U, V) > t) \\ &= \mathbb{P}(U^{1-a}V > t, UV^{1-b} > t) \\ &= \int_t^1 \mathbb{P}(U > \max((t/v)^{1/(1-a)}, tv^{b-1}) \mid V = v) dv \end{aligned}$$

$$\begin{aligned}
&= \int_t^1 \left(1 - \max(t/v, t^{1-a}v^{(b-1)(1-a)}) \mathbb{1}\{\max((t/v)^{1/(1-a)}, tv^{b-1}) \in [v^{b/a}, 1]\} \right. \\
&\quad \left. - (1-b) \max((t/v)^{1/(1-a)}, tv^{b-1})v^{-b} \mathbb{1}\{\max((t/v)^{1/(1-a)}, tv^{b-1}) \in [0, v^{b/a}]\} \right) dv.
\end{aligned}$$

Since $(t/v)^{1/(1-a)} \geq v^{b/a}$ is equivalent to the bound $t^{1/(b/a-b+1)} \geq v$, and $tv^{b-1} \geq v^{b/a}$ is also equivalent to the same bound $t^{1/(b/a-b+1)} \geq v$, we have the following equivalence relationships:

$$\max((t/v)^{1/(1-a)}, tv^{b-1}) \in [v^{b/a}, 1] \iff t^{1/(b/a-b+1)} \geq v \iff t/v \geq t^{1-a}v^{(b-1)(1-a)}.$$

Hence,

$$\begin{aligned}
\bar{K}_{a,b}(t) &= 1 - t - t \int_t^{t^{1/(b/a-b+1)}} v^{-1} dv - (1-b)t \int_{t^{1/(b/a-b+1)}}^1 v^{-1} dv \\
&= 1 - t - t \int_t^1 v^{-1} dv + bt \int_{t^{1/(b/a-b+1)}}^1 v^{-1} dv \\
&= 1 - t + t \log(t) - \frac{ab}{a+b-ab} t \log(t) \\
&= 1 - t + \left(2 - \frac{a+b}{a+b-ab} \right) t \log(t) \\
&= 1 - t + \left(2 - \frac{2}{\kappa^*} \right) t \log(t).
\end{aligned}$$

This completes the proof of formula (2.3.2) and establishes Proposition 2.A.4.

Proof of Lemma 2.3.3 Note that $K_C(t) \geq t$ for all $t \in (0, 1)$. Hence,

$$\begin{aligned}
\mathbb{E}[g(U, V)^\tau \mathbb{1}\{C(U, V) > 1/2\}] &= \int_{(1/2, 1]} \frac{1}{t^\tau (1-t)^\tau} dK_C(t) \\
&\leq 2^\tau \int_{(1/2, 1]} \frac{1}{(1-K_C(t))^\tau} dK_C(t) < \infty
\end{aligned}$$

for every $\tau \in (0, 1)$. The expectation $\mathbb{E}[g(U, V)^\tau \mathbb{1}\{C(U, V) < 1/2\}]$ is finite whenever $\int_{(0, 1/2]} t^{-\tau} dK_C(t) < \infty$, which establishes Lemma 2.3.3.

Proposition 2.A.5 *Let (U, V) follow the joint distribution C . If the moment $\mathbb{E}[g(U, V)^\tau]$ is finite for some $\tau \in (0, 1)$, then, when $n \rightarrow \infty$,*

$$\Delta_n(\tau) := \sup_{u, v \in [0, 1]} g(u, v)^\tau |C_n(u, v) - C(u, v)| \xrightarrow{p} 0.$$

Proof We begin by noting that for every sufficiently large constant c (e.g., $c > 4^\tau$), there exists $\gamma_c > 0$ such that

$$c = (1/2 + \gamma_c)^{-\tau} (1/2 - \gamma_c)^{-\tau},$$

and $\gamma_c \rightarrow 1/2$ when $c \rightarrow \infty$. Hence, $g(u, v)^\tau > c$ if and only if

$$C(u, v) \in (0, 1/2 - \gamma_c) \cup (1/2 + \gamma_c, 1).$$

Therefore,

$$\Delta_n(\tau) \leq cI_{1n}(c) + I_{2n}(c) + I_{3n}(c), \quad (2.A.1)$$

where

$$\begin{aligned} I_{1n}(c) &= \sup_{u,v \in [0,1]} |C_n(u, v) - C(u, v)|, \\ I_{2n}(c) &= \sup_{u,v \in [0,1]: C(u,v) < 1/2 - \gamma_c} g(u, v)^\tau |C_n(u, v) - C(u, v)|, \\ I_{3n}(c) &= \sup_{u,v \in [0,1]: C(u,v) > 1/2 + \gamma_c} g(u, v)^\tau |C_n(u, v) - C(u, v)|. \end{aligned}$$

For every fixed $c > 0$, the two-dimensional version of the Glivenko-Cantelli theorem implies

$$I_{1n}(c) \xrightarrow{\mathbb{P}} 0 \quad (2.A.2)$$

when $n \rightarrow \infty$. As to $I_{2n}(c)$, we estimate it from above by $J_n(c) + J(c)$, where

$$\begin{aligned} J_n(c) &= \sup_{u,v \in [0,1]: C(u,v) < 1/2 - \gamma_c} g(u, v)^\tau C_n(u, v), \\ J(c) &= \sup_{u,v \in [0,1]: C(u,v) < 1/2 - \gamma_c} g(u, v)^\tau C(u, v). \end{aligned}$$

To tackle $J_n(c)$, we first note that when $C(u, v) < 1/2 - \gamma_c$, the function $g(u, v)^\tau$ is non-increasing in each argument. Hence,

$$\begin{aligned} J_n(c) &\leq \frac{1}{n} \sum_{i=1}^n \sup_{u,v \in [0,1]: C(u,v) < 1/2 - \gamma_c} g(u, v)^\tau \mathbb{1}\{U_i \leq u, V_i \leq v\} \\ &\leq \frac{1}{n} \sum_{i=1}^n g(U_i, V_i)^\tau \sup_{u,v \in [0,1]: C(u,v) < 1/2 - \gamma_c} \mathbb{1}\{U_i \leq u, V_i \leq v\} \\ &\leq \frac{1}{n} \sum_{i=1}^n g(U_i, V_i)^\tau \sup_{u,v \in [0,1]: C(u,v) < 1/2 - \gamma_c} \mathbb{1}\{C(U_i, V_i) \leq C(u, v)\} \\ &\leq \frac{1}{n} \sum_{i=1}^n g(U_i, V_i)^\tau \mathbb{1}\{C(U_i, V_i) < 1/2 - \gamma_c\}. \end{aligned}$$

By the law of large numbers, for every $\epsilon > 0$ we can find a sufficiently large $c < \infty$ such that

$$\mathbb{P}(J_n(c) \geq \epsilon) \rightarrow 0$$

when $n \rightarrow \infty$, because $\gamma_c \rightarrow 1/2$ when $c \rightarrow \infty$ and thus $\mathbb{E}[g(U, V)^\tau \mathbb{1}\{C(U_1, V_1) < 1/2 - \gamma_c\}]$ can be made as small as desired due to $\mathbb{E}[g(U, V)^\tau] < \infty$. Furthermore,

$$J(c) \leq \sup_{u,v \in [0,1]: C(u,v) < 1/2 - \gamma_c} \frac{C(u, v)^{1-\tau}}{(1 - C(u, v))^\tau} \leq \frac{(1/2 - \gamma_c)^{1-\tau}}{(1/2 + \gamma_c)^\tau} \rightarrow 0$$

when $c \rightarrow \infty$. Hence, for every $\epsilon > 0$, we can find a sufficiently large $c < \infty$ such that, when $n \rightarrow \infty$,

$$\mathbb{P}(I_{2n}(c) \geq \epsilon) \rightarrow 0. \quad (2.A.3)$$

Finally, to tackle $I_{3n}(c)$, we start with the bound $I_{3n}(c) \leq K_n(c) + K(c)$, where

$$\begin{aligned} K_n(c) &= \sup_{u,v \in [0,1]: C(u,v) > 1/2 + \gamma_c} g(u, v)^\tau (1 - C_n(u, v)), \\ K(c) &= \sup_{u,v \in [0,1]: C(u,v) > 1/2 + \gamma_c} g(u, v)^\tau (1 - C(u, v)). \end{aligned}$$

We have

$$K_n(c) \leq \frac{1}{n} \sum_{i=1}^n \sup_{u,v \in [0,1]: C(u,v) > 1/2 + \gamma_c} g(u, v)^\tau (\mathbb{1}\{U_i > u\} + \mathbb{1}\{V_i > v\}).$$

When $C(u, v) > 1/2 + \gamma_c$, the function $g(u, v)^\tau$ is nondecreasing in each argument and thus

$$\begin{aligned} K_n(c) &\leq \frac{1}{n} \sum_{i=1}^n g(U_i, 1)^\tau \sup_{u \in [0,1]: u > 1/2 + \gamma_c} \mathbb{1}\{U_i > u\} + \frac{1}{n} \sum_{i=1}^n g(1, V_i)^\tau \sup_{v \in [0,1]: v > 1/2 + \gamma_c} \mathbb{1}\{V_i > v\} \\ &\leq \frac{1}{n} \sum_{i=1}^n g(U_i, 1)^\tau \mathbb{1}\{U_i > 1/2 + \gamma_c\} + \frac{1}{n} \sum_{i=1}^n g(1, V_i)^\tau \mathbb{1}\{V_i > 1/2 + \gamma_c\}. \end{aligned}$$

By the law of large numbers, for every $\epsilon > 0$ we can find a sufficiently large $c < \infty$ such that

$$\mathbb{P}(K_n(c) \geq \epsilon) \rightarrow 0$$

when $n \rightarrow \infty$, because $\gamma_c \rightarrow 1/2$ when $c \rightarrow \infty$ and thus the expectations $\mathbb{E}[g(U, 1)^\tau \mathbb{1}\{U > 1/2 + \gamma_c\}]$ and $\mathbb{E}[g(1, V)^\tau \mathbb{1}\{V > 1/2 + \gamma_c\}]$ can be made as small as desired, due to

$$\mathbb{E}[g(U, 1)^\tau] = \mathbb{E}[g(1, V)^\tau] = \text{Beta}(1 - \tau, 1 - \tau) < \infty.$$

Furthermore,

$$K(c) \leq \sup_{u,v \in [0,1]: C(u,v) > 1/2 + \gamma_c} \frac{(1 - C(u, v))^{1-\tau}}{C(u, v)^\tau} \leq \frac{(1/2 - \gamma_c)^{1-\tau}}{(1/2 + \gamma_c)^\tau} \rightarrow 0$$

when $c \rightarrow \infty$. Consequently, for every $\epsilon > 0$ we can find a sufficiently large $c < \infty$ such that

$$\mathbb{P}(I_{3n}(c) \geq \epsilon) \rightarrow 0 \tag{2.A.4}$$

when $n \rightarrow \infty$. Bound (2.A.1) and statements (2.A.2)–(2.A.4) establish Proposition 2.A.5.

2.B Proofs of Theorems 2.3.1–2.3.5

Proof of Theorem 2.3.1 Since there is at most one group G_j whose cardinality is less than m , without loss of generality we can, and thus do, assume $n = mk$. Denote

$$\hat{\xi}_{i,n,\theta} = \frac{2T_\theta \circ C_n(U_i, V_i)}{\log U_i + \log V_i} \quad \text{and} \quad \xi_{i,\theta} = \frac{2T_\theta \circ C(U_i, V_i)}{\log U_i + \log V_i}$$

for all $i = 1, \dots, n$. By Proposition 2.A.1, for every $j = 1, \dots, k$, we have

$$\left| \min_{i \in G_j} \hat{\xi}_{i,n,\theta} - \min_{i \in G_j} \xi_{i,\theta} \right| \leq \max_{i \in G_j} |\hat{\xi}_{i,n,\theta} - \xi_{i,\theta}|$$

$$\begin{aligned}
&= \frac{2}{\theta} \max_{i \in G_j} \left| \frac{C_n(U_i, V_i)^\theta - C(U_i, V_i)^\theta}{\log U_i + \log V_i} \right| \\
&\leq \frac{2}{\theta} \max_{i \in G_j} \frac{|C_n(U_i, V_i) - C(U_i, V_i)|}{C(U_i, V_i)^{1-\theta} |\log U_i + \log V_i|}, \tag{2.B.1}
\end{aligned}$$

where the last inequality is due to Proposition 2.A.2. Since C is PQD, we have $uv \leq C(u, v)$ and thus $|\log(uv)| \geq |\log C(u, v)|$ for all $u, v \in [0, 1]$. Bound (2.B.1) implies

$$\left| \min_{i \in G_j} \hat{\xi}_{i,n,\theta} - \min_{i \in G_j} \xi_{i,\theta} \right| \leq \frac{2}{\theta} \max_{i \in G_j} \frac{|C_n(U_i, V_i) - C(U_i, V_i)|}{C(U_i, V_i)^{1-\theta} |\log C(U_i, V_i)|}. \tag{2.B.2}$$

Furthermore, since $|\log C(u, v)| \geq 1 - C(u, v)$ for every $u, v \in [0, 1]$, bound (2.B.2) implies

$$\begin{aligned}
\left| \min_{i \in G_j} \hat{\xi}_{i,n,\theta} - \min_{i \in G_j} \xi_{i,\theta} \right| &\leq \frac{2}{\theta} \max_{i \in G_j} \frac{|C_n(U_i, V_i) - C(U_i, V_i)|}{C(U_i, V_i)^{1-\theta} [1 - C(U_i, V_i)]} \\
&\leq \frac{2}{\theta} \Delta_n (1 - \theta) \max_{i \in G_j} [1 - C(U_i, V_i)]^{-\theta}, \tag{2.B.3}
\end{aligned}$$

where $\Delta_n(\cdot)$ is defined in Proposition 2.A.5. Hence,

$$\begin{aligned}
\left| \frac{1}{k} \sum_{j=1}^k \min_{i \in G_j} \hat{\xi}_{i,n,\theta} - \frac{1}{k} \sum_{j=1}^k \min_{i \in G_j} \xi_{i,\theta} \right| &\leq \frac{1}{k} \sum_{j=1}^k \left| \min_{i \in G_j} \hat{\xi}_{i,n,\theta} - \min_{i \in G_j} \xi_{i,\theta} \right| \\
&\leq \Delta_n (1 - \theta) \frac{2}{k\theta} \sum_{j=1}^k \max_{i \in G_j} [1 - C(U_i, V_i)]^{-\theta} \\
&\leq \Delta_n (1 - \theta) \frac{2}{k\theta} \sum_{j=1}^k \sum_{i \in G_j} [1 - C(U_i, V_i)]^{-\theta} \\
&\leq \Delta_n (1 - \theta) \frac{2m}{n\theta} \sum_{i=1}^n [1 - C(U_i, V_i)]^{-\theta}. \tag{2.B.4}
\end{aligned}$$

Since the moment $\mathbb{E}[g(U_1, V_1)^{1-\theta}]$ is finite by assumption, the quantity $\Delta_n(1 - \theta)$ converges to 0 in probability. Hence, the entire right-hand side of bound (2.B.4) converges to 0 in probability if the average $n^{-1} \sum_{i=1}^n [1 - C(U_i, V_i)]^{-\theta}$ converges to a finite number. This follows from the classical law of large numbers because

$$\begin{aligned}
\mathbb{E}[(1 - C(U, V))^{-\theta}] &= \int_{(0,1)} \frac{1}{(1-t)^\theta} dK_C(t) \\
&\leq \int_{(0,1)} \frac{1}{(1-K_C(t))^\theta} dK_C(t) < \infty.
\end{aligned}$$

Therefore, the left-hand side of bound (2.B.4) converges to 0 in probability, which means

$$\hat{\kappa}_n^*(m, \theta) - \frac{1}{k} \sum_{j=1}^k \min_{i \in G_j} \xi_{i,\theta} \xrightarrow{p} 0$$

when $k \rightarrow \infty$ ($n \rightarrow \infty$), thus concluding the proof of Theorem 2.3.1.

Proof of Theorem 2.3.5 First we show that for every $m \geq 1$, we have

$$\kappa^*(m, \theta) \rightarrow \kappa^*(m, 0) \quad (2.B.5)$$

when $\theta \downarrow 0$. For this, we first employ Proposition 2.A.1 and have

$$|\kappa^*(m, \theta) - \kappa^*(m)| \leq \mathbb{E} \left(\max_{i=1, \dots, m} \left| \frac{2T_\theta \circ C(U_i, V_i)}{\log U_i + \log V_i} - \frac{2T_0 \circ C(U_i, V_i)}{\log U_i + \log V_i} \right| \right). \quad (2.B.6)$$

Next we apply Proposition 2.A.3 with $x = \theta \log(C(U_i, V_i)) \leq 0$ and have

$$\begin{aligned} |\kappa^*(m, \theta) - \kappa^*(m)| &\leq \theta \mathbb{E} \left(\max_{i=1, \dots, m} \left| \frac{(\log C(U_i, V_i))^2}{\log U_i + \log V_i} \right| \right) \\ &\leq \theta \mathbb{E} \left(\max_{i=1, \dots, m} |\log(U_i V_i)| \right), \end{aligned}$$

where the last inequality is due to C being PQD. Hence,

$$\begin{aligned} |\kappa^*(m, \theta) - \kappa^*(m, 0)| &\leq \theta \mathbb{E} \left(\max_{i=1, \dots, m} (-\log U_i) \right) + \theta \mathbb{E} \left(\max_{i=1, \dots, m} (-\log V_i) \right) \\ &= 2m\theta \int_0^\infty x(1 - e^{-x})^{m-1} e^{-x} dx. \end{aligned}$$

For every m , the right-hand side of the latter bound can be made as small as desired by choosing as sufficiently small $\theta > 0$. This concludes the proof of statement (2.B.5).

We next prove that $\kappa^*(m, 0)$ can be made as close to κ^* as desired by choosing a sufficiently large m . We start with the equation

$$\kappa^*(m, 0) = \int_0^2 \mathbb{P}(\xi > x)^m dx, \quad (2.B.7)$$

where

$$\xi := \frac{2 \log C(U, V)}{\log U + \log V} \in [0, 2] \quad (2.B.8)$$

with $(U, V) \sim C$. Since $\Pi^*(u) = u^{\kappa^*}$ for all $u \in [0, 1]$ by assumption, we have

$$\kappa^* \leq \frac{2 \log C(u, v)}{\log u + \log v}$$

for all $u, v \in [0, 1]$, and thus $\kappa^* \leq \text{ess inf } \xi$. Hence, equation (2.B.7) becomes

$$\kappa^*(m, 0) = \kappa^* + \int_{\kappa^*}^2 \mathbb{P}(\xi > x)^m dx, \quad (2.B.9)$$

from which we see that we only need to show that the integral on the right-hand side converges to 0 when $m \rightarrow \infty$. In doing so, we can assume without loss of generality that $\kappa^* < 2$, which effectively eliminates the independence copula, that is, from now on we can, and thus do, assume that C is not the independence copula.

By the Lebesgue dominated convergence theorem, the integral converges to 0 provided that $\mathbb{P}(\xi > x) \in (0, 1)$ for all $x \in (\kappa^*, 2)$. In other words, we need to show that

$$\mathbb{P}(\xi \leq \kappa^* + h) > 0 \quad \text{for every } h \in (0, 2 - \kappa^*). \quad (2.B.10)$$

We start with the bound

$$\begin{aligned} \mathbb{P}(\xi \leq \kappa^* + h) &= \mathbb{P}\left(\frac{2 \log C(U, V)}{\log U + \log V} \leq \frac{2 \log w_0^{\kappa^* + h}}{\log w_0^2}\right) \\ &\geq \mathbb{P}((U, V) \in B_h), \end{aligned} \quad (2.B.11)$$

where the area B_h (see Figure 2.4) is defined by the equation

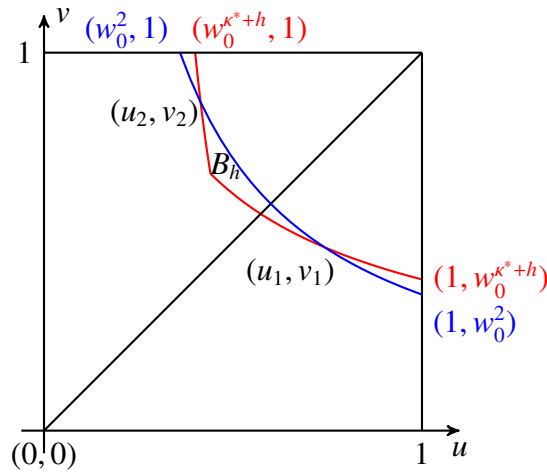


Figure 2.4: The area B_h with its associated curves and points.

$$B_h = \left\{ (u, v) \in [0, 1]^2 : uv \leq w_0^2, C(u, v) > w_0^{\kappa^* + h} \right\}.$$

Since C is PQD but not the independence copula, we can find $u_0, v_0 \in (0, 1)$ such that

$$u_0 v_0 < C(u_0, v_0) = (u_0 v_0)^{\kappa^*/2}.$$

With the notation $w_0 = \sqrt{u_0 v_0}$, we have $u_0 = \varphi^*(w_0)$, $v_0 = w_0^2 / \varphi^*(w_0)$, and

$$C\left(\varphi^*(w_0), w_0^2 / \varphi^*(w_0)\right) = w_0^{\kappa^*}.$$

Since $\kappa^* < 2$, $\varphi^*(w_0)$ is neither w_0^2 nor 1. Therefore, there is $\eta \in (0, 2)$ such that $\varphi^*(w_0) = w_0^\eta$ due to $w_0^2 < \varphi^*(w_0) < 1$. Define the function $\Phi_{w_0}^* : [0, 2] \rightarrow [0, 1]$ by the formula

$$\Phi_{w_0}^*(x) = C\left(w_0^x, w_0^{2-x}\right).$$

Note the values:

$$\Phi_{w_0}^*(0) = w_0^2, \quad \Phi_{w_0}^*(\eta) = w_0^{\kappa^*} \quad \text{and} \quad \Phi_{w_0}^*(2) = w_0^2.$$

Moreover, for $0 \leq x_1 \leq x_2 \leq 2$, we have

$$\begin{aligned} |\Phi_{w_0}^*(x_2) - \Phi_{w_0}^*(x_1)| &\leq |w^{x_2} - w^{x_1}| + |w^{2-x_2} - w^{2-x_1}| \\ &= (w_0^{x_1} + w_0^{2-x_2})(1 - w_0^{x_2-x_1}) \\ &\leq 2(1 - w_0^{x_2-x_1}), \end{aligned}$$

which shows that $\Phi_{w_0}^*$ is continuous and thus uniformly continuous on the compact interval $[0, 2]$. Therefore, we can find $x_1 \in (0, \eta)$ and $x_2 \in (\eta, 2)$ such that

$$C(u_1, v_1) = C(u_2, v_2) = w_0^{\kappa^*+h}$$

where $u_i = w_0^{x_i}$ and $v_i = w_0^{2-x_i}$ for $i = 1$ and 2 . This implies that the curve $uv = w_0^2$ and the contour $C(u, v) = w_0^{\kappa^*+h}$ intersect at the points (u_1, v_1) and (u_2, v_2) .

We next show that into the area B_h we can squeeze a rectangle E_h of positive Lebesgue area such that $\mathbb{P}((U, V) \in E_h) > 0$. In view of bound (2.B.11), this will imply statement (2.B.10) needed to complete the proof of Theorem 2.3.5. Denote

$$u_h = \varphi^*(w_0^{(\kappa^*+h)/\kappa^*}) \quad \text{and} \quad v_h = w_0^{2(\kappa^*+h)/\kappa^*} / \varphi^*(w_0^{(\kappa^*+h)/\kappa^*}). \quad (2.B.12)$$

We have $C(u_h, v_h) = \Pi^*(w_0^{(\kappa^*+h)/\kappa^*}) = w_0^{\kappa^*+h}$. Define the rectangle by the equation

$$E_h := (u_h, u_0] \times (v_h, v_0] \subset B_h$$

(see Figure 2.5). The functions $\varphi^*(u)$ and $\psi^*(u) := u^2/\varphi^*(u)$ are strictly increasing, and so

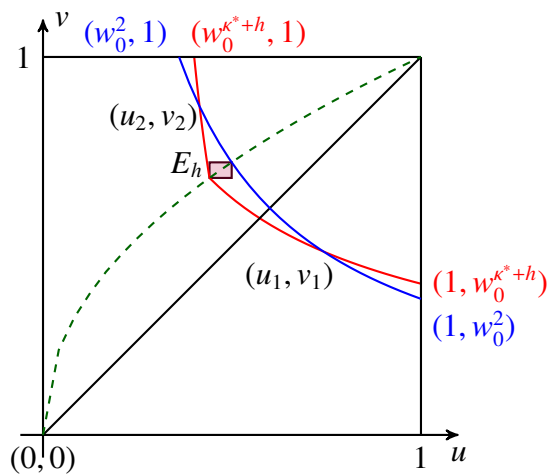


Figure 2.5: The rectangle E_h with its associated curves and points.

$u_h < u_0, v_h < v_0$. Moreover, since

$$w_0^{2(\kappa^*+h)/\kappa^*} \leq \varphi^*(w_0^{(\kappa^*+h)/\kappa^*}) \leq 1,$$

we have

$$(u_h, v_h) = (w_0^{\eta_h}, w_0^{2(\kappa^*+h)/\kappa^* - \eta_h})$$

for some $\eta_h \in [0, 2(\kappa^* + h)/\kappa^*]$. Since $u_h < u_0$ and $v_h < v_0$, we have the bounds $\eta_h > \eta$ and

$$\frac{2(\kappa^* + h)}{\kappa^*} - \eta_h > 2 - \eta,$$

which imply

$$\eta < \eta_h < \eta + \frac{2h}{\kappa^*}. \quad (2.B.13)$$

Hence,

$$\begin{aligned} \mathbb{P}((U, V) \in E_h) &= C(u_0, v_0) - C(u_0, v_h) - C(u_h, v_0) + C(u_h, v_h) \\ &= w_0^{\kappa^*} - C(u_0, v_h) - C(u_h, v_0) + w_0^{\kappa^* + h} \\ &\geq w_0^{\kappa^*} - (w_0^\eta w_0^{2(\kappa^* + h)/\kappa^* - \eta_h})^{\kappa^*/2} - (w_0^{\eta_h} w_0^{2 - \eta})^{\kappa^*/2} + w_0^{\kappa^* + h} \\ &= w_0^{\kappa^*} - w_0^{\kappa^* + h + \kappa^*(\eta - \eta_h)/2} - w_0^{\kappa^* - \kappa^*(\eta - \eta_h)/2} + w_0^{\kappa^* + h} \\ &= w_0^{\kappa^*} (1 - w_0^{-\kappa^*(\eta - \eta_h)/2}) + w_0^{\kappa^* + h} (1 - w_0^{\kappa^*(\eta - \eta_h)/2}) \\ &= (w_0^{\kappa^* + \kappa^*(\eta_h - \eta)/2} - w_0^{\kappa^* + h}) (w_0^{\kappa^*(\eta - \eta_h)/2} - 1), \end{aligned}$$

which is positive, because bounds (2.B.13) imply that the two factors on the right-hand side of the bound are strictly positive. This completes the proof of Theorem 2.3.5.

2.C Histograms and normal fits

In this appendix we depict histograms and fitted normal densities of the values of $\hat{\kappa}_n^*(m, \theta_0)$ (given by equation (2.2.10)) that arise from simulated iid data, as discussed in Section 2.3.

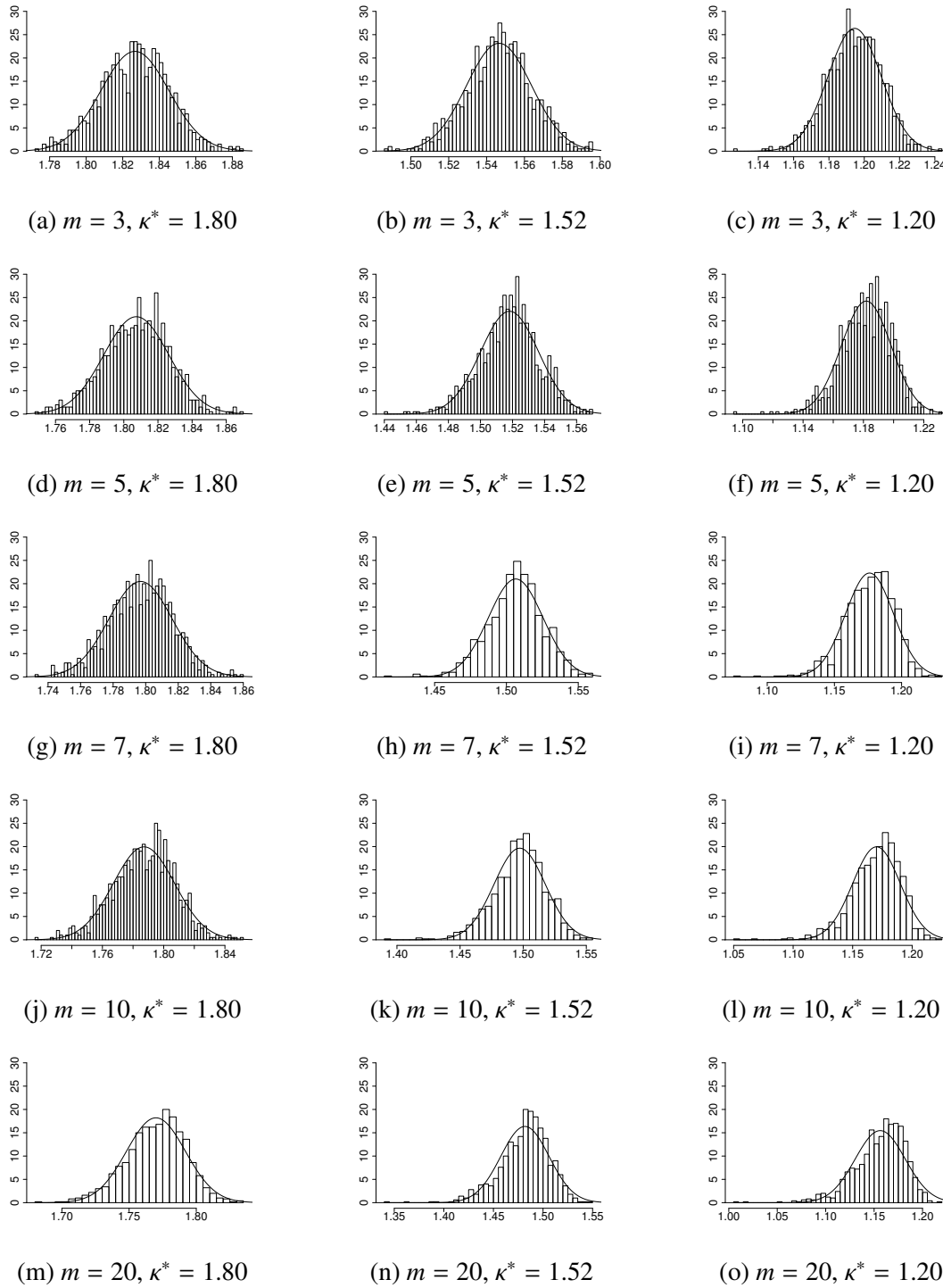


Figure 2.6: Fits of simulated $\hat{\kappa}_n^*(m, \theta_0)$ to the normal distribution for iid data.

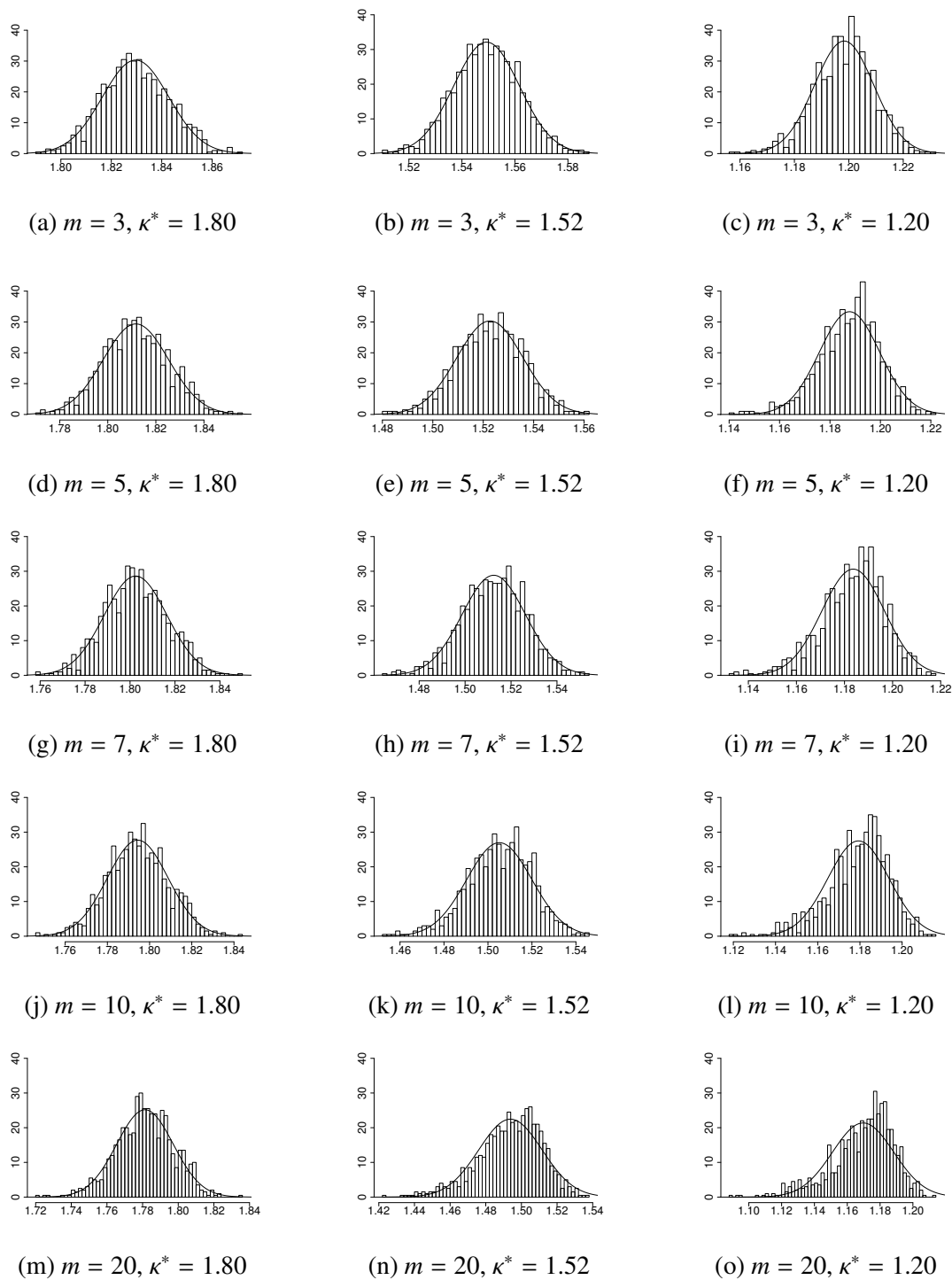


Figure 2.7: Fits of simulated $\hat{\kappa}_n^*(m, \theta_0)$ to the normal distribution for iid data.

Bibliography

- [1] H. Joe, Parametric families of multivariate distributions with given margins, *Journal of Multivariate Analysis* 46 (2) (1993) 262–282. 6
- [2] S. Coles, J. Heffernan, J. Tawn, Dependence measures for extreme value analyses, *Extremes* 2 (4) (1999) 339–365.
- [3] L. Hua, H. Joe, Tail order and intermediate tail dependence of multivariate copulas, *Journal of Multivariate Analysis* 102 (10) (2011) 1454–1471.
- [4] H. Joe, *Dependence Modeling with Copulas*, Chapman and Hall/CRC, 2014. 6
- [5] N. H. Bingham, C. M. Goldie, J. L. Teugels, *Regular Variation*, Cambridge University Press, 1987. 6, 10, 15
- [6] E. Furman, J. Su, R. Zitikis, Paths and indices of maximal tail dependence, *ASTIN Bulletin* 45 (3) (2015) 661–678. 6, 8, 9, 10, 17, 18
- [7] R. B. Nelsen, *An Introduction to Copulas*, Springer New York, 2006. 6
- [8] F. Durante, C. Sempi, *Principles of Copula Theory*, Chapman and Hall/CRC, 2015. 6
- [9] A. J. Mcneil, R. Frey, P. Embrechts, *Quantitative Risk Management*, Princeton University Press, ISBN 0691166277, 2015. 6, 9
- [10] R. Wang, R. Zitikis, An axiomatic foundation for the expected shortfall, *Management Science* 67 (3) (2021) 1413–1429. 7
- [11] G. Draisma, H. Drees, A. Ferreira, L. D. Haan, Bivariate tail estimation: dependence in asymptotic independence, *Bernoulli* 10 (2). 8
- [12] G. Frahm, M. Junker, R. Schmidt, Estimating the tail-dependence coefficient: Properties and pitfalls, *Insurance: Mathematics and Economics* 37 (1) (2005) 80–100.
- [13] P. Krupskii, H. Joe, Nonparametric estimation of multivariate tail probabilities and tail dependence coefficients, *Journal of Multivariate Analysis* 172 (2019) 147–161. 8
- [14] E. Furman, A. Kuznetsov, J. Su, R. Zitikis, Tail dependence of the Gaussian copula revisited, *Insurance: Mathematics and Economics* 69 (2016) 97–103. 9
- [15] U. Cherubini, E. Luciano, W. Vecchiato, *Copula Methods in Finance*, John Wiley & Sons Ltd, 2013. 9, 19
- [16] U. Cherubini, F. Durante, S. Mulinacci (Eds.), *Marshall Olkin Distributions - Advances in Theory and Applications*, Springer International Publishing, 2015. 9
- [17] J. Su, E. Furman, Multiple risk factor dependence structures: Copulas and related properties, *Insurance: Mathematics and Economics* 74 (2017) 109–121. 9

- [18] T. L. Lai, Convergence rates in the strong law of large numbers for random variables taking values in Banach spaces, *Bulletin of the Institute of Mathematics Academia Sinica* 2 (1) (1974) 67–85. 10
- [19] D. M. Mason, Some characterizations of almost sure bounds for weighted multidimensional empirical distributions and a Glivenko-Cantelli theorem for sample quantiles, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 59 (4) (1982) 505–513. 10
- [20] B. Berghaus, A. Bücher, S. Volgushev, Weak convergence of the empirical copula process with respect to weighted metrics, *Bernoulli* 23 (1). 10
- [21] R. B. Nelsen, J. J. Quesada-Molina, J. A. Rodríguez-Lallena, M. Úbeda Flores, Kendall distribution functions, *Statistics & Probability Letters* 65 (3) (2003) 263–268. 12
- [22] R. B. Nelsen, J. J. Quesada-Molina, J. A. Rodríguez-Lallena, M. Úbeda Flores, Distribution functions of copulas: a class of bivariate probability integral transforms, *Statistics & Probability Letters* 54 (3) (2001) 277–282. 13
- [23] Basel Committee on Banking Supervision, Minimum Capital Requirements for Market Risk (January, 2016), URL <https://www.bis.org/bcbs/publ/d352.htm>, 2016. 15
- [24] Basel Committee on Banking Supervision, Minimum Capital Requirements for Market Risk (February, 2019), URL <https://www.bis.org/bcbs/publ/d457.htm>, 2019. 15
- [25] V. Vovk, R. Wang, Combining p -values via averaging, *Biometrika* 107 (4) (2020) 791–808. 18

Chapter 3

Tail maximal dependence in bivariate models: estimation and applications

3.1 Introduction

The phenomenon of extreme co-movements manifests in a variety of problems associated with risk management. To illustrate, at the market level, this phenomenon arises when dealing with financial contagion [1]. At the asset level, studies of extreme co-movements facilitate risk mitigation arrangements, such as portfolio diversification [2]. These and many other applications have inspired prolific searches for, and studies of, methods for assessing dependence within extreme co-movements [e.g., 3, 4, and references therein]. For this purpose, researchers have often employed tail dependence indices [e.g., 5–9, and references therein]. Well-developed statistical inference results for the indices [e.g., 10–12, and references therein] have greatly facilitated their practical uses.

Definitions of the indices rely on the behaviour of the copula $C : [0, 1]^2 \rightarrow [0, 1]$, which arises from a pair of financial instruments, along the diagonal path $(u, u)_{0 \leq u \leq 1}$ near the point $\mathbf{0} := (0, 0)$. In particular, the parameter $\kappa \in [1, \infty)$ in the equation

$$C(u, u) = \ell(u)u^\kappa, \tag{3.1.1}$$

assuming that it holds with a slowly varying at 0 function ℓ [e.g., 13], is called the lower tail order, which we henceforth call the tail order of *diagonal* dependence (TODD). The modifier *diagonal* has been added to emphasize the equality of the two copula arguments on the left-hand side of equation (3.1.1).

Indices of tail dependence such as κ have played prominent roles in quantifying risks, as seen from the many scholarly writings on the topic, including the monographs by Nelsen [14], Joe [8], Durante and Sempi [15], and McNeil et al. [9]. For a recently introduced notion of p -concentration, which is a diagonal-based measure of dependence that has played a pivotal role in characterizing the regulatory risk measure Expected Shortfall, we refer to Wang and Zitikis [16] and references therein.

Interestingly, the asymptotic behaviour of $C(u, u)$ may or may not reflect the *maximal* strength of tail dependence, as has been pointed out and illustrated by Furman et al. [17, 18]. Hence, it becomes natural to seek a path, called a path of maximal tail dependence (MTD),

along which the copula C behaves like $\ell^*(u)u^{\kappa^*}$ for the smallest possible κ^* and some slowly varying at 0 function ℓ^* . We note that this MTD path may not be unique, and may not coincide with the diagonal path $(u, u)_{0 \leq u \leq 1}$. Henceforth, we call κ^* the tail order of *maximal* dependence (TOMD), whose rigorous description will be given in the next section.

For the TOMD κ^* , basic statistical inference results have been developed by Sun et al. [19] for independent and identically distributed (iid) paired data. In practice, however, co-movements often arise from time-indexed stochastic processes, such as time series [e.g., 4, and references therein], whose generated data are inherently dependent. The iid-based results of Sun et al. [19], therefore, need to be adjusted and even extended. Fortunately, the block-structure of the estimator of Sun et al. [19] facilitates the task, due to the fact that practically relevant time series give rise to nearly-independent data blocks when the gaps between the blocks are sufficiently wide. Nevertheless, serious work remains to be done, and this is our main goal in the current chapter.

The rest is organized as follows. In Section 3.2 we introduce TOMD and TODD estimators. In Section 3.3 we discuss the estimators and prepare them for the analysis of real time series data. In Section 3.4 we develop a dependent-data generating process and then illustrate and assess the estimators' performance. In Section 3.5 we explore the strength of extreme co-movements of a number of financial instruments, such as historical exchange rates of several currencies, stock market indices, and mixed financial instruments. Section 3.6 concludes this chapter with a summary. To facilitate readability of this chapter, a number of statistical tests, tables and graphs have been relegated to Appendix 3.6, which consists of Sections 3.A–3.C.

3.2 Tail-order estimators

For any bivariate copula $C : [0, 1]^2 \rightarrow [0, 1]$, a path of tail dependence [17] is defined as $(\varphi(u), u^2/\varphi(u))_{0 \leq u \leq 1}$ for any function $\varphi : [0, 1] \rightarrow [0, 1]$ that satisfies two admissibility conditions:

1. $\varphi(u) \in [u^2, 1]$ for every $u \in [0, 1]$,
2. $\varphi(u) \rightarrow 0$ and $u^2/\varphi(u) \rightarrow 0$ when $u \downarrow 0$.

Denote the set of all admissible functions by \mathcal{A} .

An admissible function $\varphi^* \in \mathcal{A}$ that maximizes the functional $\varphi \mapsto C(\varphi(u), u^2/\varphi(u))$ gives rise to a path $(\varphi^*(u), u^2/\varphi^*(u))_{0 \leq u \leq 1}$ that we call a path of maximal tail dependence (MTD). There can be several maximizing functions φ^* and thus several MTD paths. For any of them, define $\Pi^* : [0, 1] \rightarrow [0, 1]$ by

$$\Pi^*(u) = C(\varphi^*(u), u^2/\varphi^*(u)).$$

If there exists κ^* and a slowly varying at 0 function ℓ^* such that

$$\Pi^*(u) = \ell^*(u)u^{\kappa^*}, \tag{3.2.1}$$

then κ^* is called the lower tail order of *maximal* dependence (TOMD). This κ^* is unique for the copula C , irrespective of the fact that there can be several admissible functions φ^* leading to it.

As illustrated by Furman et al. [17], the TOMD κ^* may or may not coincide with the TODD κ , with the Gaussian copula providing one of those rare examples when $\kappa^* = \kappa$ [18].

Intuitively, the role of a maximal path, written generally as $(\varphi^*(u), \psi^*(u))_{0 \leq u \leq 1}$, is for each $u \in (0, 1]$ to define the rectangle

$$\mathcal{R}_u(\mathbf{0}) := [0, \varphi^*(u)] \times [0, \psi^*(u)]$$

which

1. is a subset of the unit square $[0, 1]^2$;
2. has the same area as $[0, u]^2$, and thus necessarily implies $\psi^*(u) = u^2/\varphi^*(u)$;
3. when a large number of pairs are simulated from the copula C , the rectangle $\mathcal{R}_u(\mathbf{0})$ contains at least as many simulated pairs as any other rectangle in the first quadrant of the plane with one of its vertices being $\mathbf{0}$ and the area equal to u^2 .

We now introduce an empirical estimator of the TOMD κ^* . Suppose that the underlying model is a bi-variate stationary time series (X_i, Y_i) , $i \in \mathbb{Z}$, and let the pairs $(X_1, Y_1), \dots, (X_n, Y_n)$ be observable. Hence, our data are $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^2$. Next we separate the marginal distributions from the dependence structure, about which we learn from the bivariate pseudo-observations $(u_1, v_1), \dots, (u_n, v_n) \in [0, 1]^2$, which have been generated from the copula C . These pseudo-observations give rise to a scatterplot in the unit square $[0, 1]^2$, as well as to the empirical copula

$$C_n(u, v) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{u_i \leq u, v_i \leq v\}, \quad u, v \in [0, 1].$$

In order to mitigate the potential influence of ℓ^* (equation (3.2.1)) on the TOMD estimation, we focus on those observed pairs that fall into the rectangle

$$\mathcal{R}_{q,n}(\mathbf{0}) := [0, \varphi_n^*(q)] \times [0, q^2/\varphi_n^*(q)],$$

where

$$\varphi_n^*(q) = \arg \max_{x \in [q^2, 1]} C_n(x, q^2/x).$$

Various choices of $q \in (0, 1]$ will be discussed later in this chapter, for simulated and real data. We now only note that when $q = 1$, the rectangle is the unit square $[0, 1]^2$.

Hence, from all the pseudo observations $(u_1, v_1), \dots, (u_n, v_n) \in [0, 1]^2$, we single out those that are in $\mathcal{R}_{q,n}(\mathbf{0})$. We identify them by their indices, which we collect into the set

$$\mathcal{M}_{q,n} := \{i : (u_i, v_i) \in \mathcal{R}_{q,n}(\mathbf{0})\} \subseteq \{1, \dots, n\}.$$

Let $m_{q,n} := \#\mathcal{M}_{q,n}$ denote the cardinality of $\mathcal{M}_{q,n}$, that is, the number of the pairs (u_i, v_i) residing in $\mathcal{R}_{q,n}(\mathbf{0})$. Obviously, $m_{q,n} \leq n$ and so when we make the assumption that $m_{q,n}$ is sufficiently large, we also implicitly say that the underlying sample size n is large. The quantity

$$\Pi_n^*(q) := \frac{m_{q,n}}{n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{(u_i, v_i) \in \mathcal{R}_{q,n}(\mathbf{0})\}$$

is an empirical proxy for $\Pi^*(q) = \mathbb{P}((U, V) \in \mathcal{R}_q(\mathbf{0}))$.

Note 3.2.1 *This chapter is long, and we have tried – whenever reasonable – to avoid pedantic details of the proof. For example, results like $\Pi_n^*(q)$ being a proxy for $\Pi^*(q)$ would normally go without a proof, but to illustrate this instance, suppose that we wish to check the result. Since*

$$\Pi_n^*(q) = \max_{x \in [q^2, 1]} C_n(x, q^2/x), \quad \Pi^*(q) = \max_{x \in [q^2, 1]} C(x, q^2/x)$$

for $q \in (0, 1]$, we have

$$\begin{aligned} |\Pi_n^*(q) - \Pi^*(q)| &\leq \max_{x \in [q^2, 1]} |C_n(x, q^2/x) - C(x, q^2/x)| \\ &\leq \sup_{u, v \in [0, 1]} |C_n(u, v) - C(u, v)| \xrightarrow{P} 0 \end{aligned}$$

when $n \rightarrow \infty$, where \xrightarrow{P} denotes convergence in probability. As we see from the classical results of [20], and Kiefer et al. [21], the convergence holds even almost surely. Much work has been done since these papers in order to relax the iid assumption, and we refer to, e.g., Dehling et al. [22], Davydov and Zitikis [23], and Kontorovich and Weiss [24] for results and references on the topic.

Next, we fix any $m \geq 1$ such that $m \leq m_{q,n}$, which holds (Note 3.2.1) with as large a probability as desired, assuming that n is sufficiently large. Then we assign the pairs (u_i, v_i) , $i \in \mathcal{M}_{q,n}$, into $\lceil m_{q,n}/m \rceil$ disjoint groups so that there can be at most one group with less than m pairs and all the other groups containing exactly m pairs, where $\lceil \cdot \rceil$ is the classical ceiling function. We collect the indices of the grouped pairs into the (disjoint) sets $G_{j,q,n}$, $1 \leq j \leq \lceil m_{q,n}/m \rceil$, thus producing a partition of $\mathcal{M}_{q,n}$. We note that choosing an appropriate value of m is a delicate problem in practice, and we shall discuss it in great detail when working with simulated and real data later in this chapter.

The average block-minima estimator of the TOMD κ^* is [cf. 19]

$$\widehat{\kappa}_{m_{q,n}}^*(m, \theta, q) = \frac{1}{\lceil m_{q,n}/m \rceil} \sum_{j=1}^{\lceil m_{q,n}/m \rceil} \min_{i \in G_{j,q,n}} \frac{2T_\theta \circ F_{q, \mathcal{M}_{q,n}}(u_i, v_i)}{\log u_i + \log v_i - 2 \log q}, \quad (3.2.2)$$

where, for all $u, v \in [0, 1]$,

$$F_{q, \mathcal{M}_{q,n}}(u, v) = \frac{1}{m_{q,n}} \sum_{i \in \mathcal{M}_{q,n}} \mathbb{1}\{u_i \leq u, v_i \leq v\}, \quad (3.2.3)$$

and, for all $t \in (0, 1]$,

$$T_\theta(t) = \begin{cases} (t^\theta - 1)/\theta & \text{when } \theta > 0, \\ \log(t) & \text{when } \theta = 0. \end{cases}$$

Note 3.2.2 *We may think of various other ways of combining the minima (or some other functionals replacing them) that are on the right-hand side of definition (3.2.2). For an illuminating discussion of a philosophically-related problem, we refer to Vovk and Wang [25].*

For comparison, we also estimate the TODD κ . Its estimator is defined as follows [cf. 26]. Denote

$$\mathcal{N}_{q,n} = \{i : (u_i, v_i) \in \mathcal{S}_q(\mathbf{0})\} \subseteq \{1, \dots, n\},$$

where $\mathcal{S}_q(\mathbf{0}) = [0, q]^2$. Let $n_{q,n} := \#(\mathcal{N}_{q,n})$, the cardinality of $\mathcal{N}_{q,n}$, that is, $n_{q,n}$ is the number of pairs (u_i, v_i) in the square $\mathcal{S}_q(\mathbf{0})$. For each $i \in \mathcal{N}_{q,n}$, denote $w_i := q \min\{u_i^{-1}, v_i^{-1}\}$. Before we introduce an estimator of TODD κ , we simplify the notation by re-enumerating w_i , $i \in \mathcal{N}_{q,n}$ into $w_1, \dots, w_{n_{q,n}}$ whose order statistics we denote by $w_{1:n_{q,n}} \geq \dots \geq w_{n_{q,n}:n_{q,n}}$. (Technically, we would need to change the notation of the re-enumerated w 's but it worth leaving them as they are for the sake of notational simplicity.) Note that $n_{q,n} = \max\{i : w_{i:n_{q,n}} \geq 1\}$. We define the estimator of the TODD κ by

$$\widehat{\kappa}_{n_{q,n}}^{\text{OLS}} = \frac{\sum_{i=1}^{n_{q,n}} (\log w_{i:n_{q,n}} - \overline{\log w}) \log(i - 0.5)}{\sum_{i=1}^{n_{q,n}} (\log w_{i:n_{q,n}} - \overline{\log w})^2}, \quad (3.2.4)$$

where $\overline{\log w}$ denotes the average of all $\log w_1, \dots, \log w_{n_{q,n}}$. We refer to Gabaix and Ibragimov [26] for an illuminating discussion of why 0.5 needs to be subtracted from i in $\log(i - 0.5)$ when defining the estimator.

3.3 Justification of the estimators

In practice we cannot know whether slowly varying functions are present or not in equations (3.1.1) and (3.2.1). This poses a challenge. In this section, therefore, we adjust the methodology of Sun et al. [19] so that we could tackle TOMD and TODD estimation under the uncertainty with respect to slowly varying functions. An important feature that will permeate our following considerations is a conditioning argument, whose main idea is based on the fact that extreme financial losses are those that are below a certain (small) threshold $q \in (0, 1)$, which could have been set by convention or regulation [e.g., 27, 28].

Hence, we are dealing with conditional copulas below q , which are ratios of probabilities. Via equation (3.2.1), we reduce these ratios to quantities like $\ell^*(qu)/\ell^*(q)$, which are close to 1 when $q > 0$ is small, due to the very definition of slowly varying at 0 functions [e.g., 13, Definition, p. 6]:

$$\frac{\ell^*(qu)}{\ell^*(q)} \rightarrow 1 \quad \text{when } q \downarrow 0, \quad (3.3.1)$$

and this limit holds irrespective of $u > 0$.

When choosing a small $q > 0$ in practice, we should be mindful of the fact that the smaller the threshold q , the smaller the number $m_{q,n}$ of pairs falling into the rectangle $\mathcal{R}_{q,n}(\mathbf{0})$. This obviously impedes statistical inference, and so we need to strike a balance between the values of q and $m_{q,n}$. We shall give a considerable thought to this issue when dealing with real data later in this chapter.

3.3.1 Estimating the TOMD

We first recall a procedure for estimating the TOMD κ^* developed by Sun et al. [19]. By doing so, we also introduce the necessary notation for our following considerations. We start with

equation (3.2.1) and express the conditional maximal tail probability as

$$\frac{\Pi^*(qu)}{\Pi^*(q)} = \frac{C(\varphi^*(qu), q^2 u^2 / \varphi^*(qu))}{C(\varphi^*(q), q^2 / \varphi^*(q))} = \frac{\ell^*(qu)(qu)^{\kappa^*}}{\ell^*(q)q^{\kappa^*}} \approx u^{\kappa^*} \quad (3.3.2)$$

when $0 < q \approx 0$, due to statement (3.3.1). Hence, we have the approximate bound

$$\frac{C(\varphi(qu), q^2 u^2 / \varphi(qu))}{C(\varphi^*(q), q^2 / \varphi^*(q))} \lesssim u^{\kappa^*} \quad (3.3.3)$$

that holds for every $\varphi \in \mathcal{A}$. With the notations $\tilde{u} = \varphi(qu)/\varphi^*(q)$ and $\tilde{v} = u^2 \varphi^*(q)/\varphi(qu)$, we turn approximate bound (3.3.3) into

$$\frac{C(\varphi^*(q)\tilde{u}, q^2 \tilde{v} / \varphi^*(q))}{C(\varphi^*(q), q^2 / \varphi^*(q))} \lesssim u^{\kappa^*}.$$

Since $u^{\kappa^*} = (\tilde{u}\tilde{v})^{\kappa^*/2}$, this leads to

$$\kappa^* \lesssim \frac{2 \log F_q^*(\tilde{u}, \tilde{v})}{\log \tilde{u} + \log \tilde{v}}, \quad (3.3.4)$$

where $F_q^* : [0, 1]^2 \rightarrow [0, 1]$ is defined by

$$F_q^*(u, v) = \frac{C(u\varphi^*(q), vq^2 / \varphi^*(q))}{C(\varphi^*(q), q^2 / \varphi^*(q))}. \quad (3.3.5)$$

This gives a theoretical basis for building an empirical estimator of the TOMD κ^* .

Namely, given pseudo observations $(u_1, v_1), \dots, (u_n, v_n) \in [0, 1]^2$, and also a small and fixed $q > 0$, the TOMD κ^* is estimated using the following four-step procedure [cf. 19]:

1. using all the pairs available in the unit square $[0, 1]^2$, compute an empirical estimate of $\varphi^*(q)$ by maximizing $C_n(x, q^2/x)$ with respect to $x \in [q^2, 1]$;
2. extract the set $\mathcal{P}_{q,n}$ of those pairs (u_k, v_k) that are in the rectangle $\mathcal{R}_{q,n}(\mathbf{0})$, then collect the indices of the pairs into the set $\mathcal{M}_{q,n}$, and denote the cardinality of $\mathcal{M}_{q,n}$ by $m_{q,n}$;
3. randomly assign the pairs of $\mathcal{P}_{q,n}$ into $\lceil m_{q,n}/m \rceil$ disjoint groups of pairs, whose indices partition $\mathcal{M}_{q,n}$ into groups $G_{1,q,n}, \dots, G_{\lceil m_{q,n}/m \rceil, q,n}$ with at most one of them having fewer than m elements, while all the other groups having exactly m elements;
4. compute the average block-minima estimator of the TOMD κ^* using formula (3.2.2).

To appreciate formula (3.2.2) in view of the above procedure, note that when $i \in \mathcal{M}_{q,n}$ we have

$$F_{q, \mathcal{M}_{q,n}}(u_i, v_i) = \frac{1}{m_{q,n}} \#\{k \in \mathcal{M}_{q,n} : u_k \leq u_i, v_k \leq v_i\} = F_{q, \mathcal{M}_{q,n}}^*(\tilde{u}_i, \tilde{v}_i),$$

where $\tilde{u}_i = u_i / \varphi_n^*(q)$, $\tilde{v}_i = v_i \varphi_n^*(q) / q^2$, and

$$F_{q, \mathcal{M}_{q,n}}^*(u, v) := \frac{1}{m_{q,n}} \sum_{k \in \mathcal{M}_{q,n}} \mathbb{1}\{\tilde{u}_k \leq u, \tilde{v}_k \leq v\}. \quad (3.3.6)$$

Hence, estimator (3.2.2) can be rewritten as

$$\widehat{\kappa}_{m_{q,n}}^*(m, \theta, q) = \frac{1}{\lceil m_{q,n}/m \rceil} \sum_{j=1}^{\lceil m_{q,n}/m \rceil} \min_{i \in G_{j,q,n}} \frac{2T_\theta \circ F_{q, \mathcal{M}_{q,n}}^*(\tilde{u}_i, \tilde{v}_i)}{\log \tilde{u}_i + \log \tilde{v}_i}. \quad (3.3.7)$$

3.3.2 Estimating the TODD

For comparison, we estimate the TODD κ . With some modifications, we follow the spirit of Section 3.3.1. Namely, in view of equation (3.1.1), we have

$$\frac{C(qu, qu)}{C(q, q)} = \frac{\ell(qu)(qu)^\kappa}{\ell(q)q^\kappa} \approx u^\kappa \quad (3.3.8)$$

when $0 < q \approx 0$, due to statement (3.3.1). To see an analogy between equations (3.3.8) and (3.3.2), set $\varphi^*(u) = u$ in statement (3.3.2). Next, we rewrite approximate equation (3.3.8) as

$$\mathbb{P}(\max\{U, V\} \leq qu \mid \max\{U, V\} \leq q) \approx u^\kappa. \quad (3.3.9)$$

With the notation $z := u^{-1}$, equation (3.3.9) turns into

$$\mathbb{P}(W \geq z \mid W \geq 1) \approx z^{-\kappa}, \quad (3.3.10)$$

where $W = q \min\{U^{-1}, V^{-1}\}$.

To implement approximate equation (3.3.10) on data, we need its empirical version. For this, modulus some adjustments, we follow the ideas of Gabaix and Ibragimov [26]. In terms of the order statistics $w_{1:n_{q,n}} \geq \dots \geq w_{n_{q,n}:n_{q,n}}$, the empirical version of statement (3.3.10) becomes

$$i/n_{q,n} \approx w_{i:n_{q,n}}^{-\kappa}, \quad i = 1, \dots, n_{q,n}. \quad (3.3.11)$$

Taking logarithms of both sides of approximate equation (3.3.11), we obtain

$$\log i \approx \log n_{q,n} - \kappa \log w_{i:n_{q,n}}.$$

We now arrive at an estimator of κ as the slope of the least squares regression line fitted to the scatterplot of the pairs $(\log(i - 0.5), \log w_{i:n_{q,n}})$, $i = 1, \dots, n_{q,n}$. This gives estimator (3.2.4).

3.3.3 Estimator's performance justification

We use all n available pseudo observations (u_i, v_i) to get estimates of C , $\varphi^*(q)$, $\mathcal{R}_q(\mathbf{0})$, and $\Pi^*(q)$. Since in our applications the sample size n is large, it is reasonable to assume that the estimates are close – or as close as we can possibly get them – to their population versions, and we thus adopt the same notations for them, that is, we skip the subscript n .

The number m_q of the observed pairs in the estimated rectangle is relatively small, and thus their variability matters. We thus face the question of whether or not m_q is sufficiently large in order to make reliable statistical decisions. To answer this question, we theoretically explore the convergence of the simplified estimator

$$\widehat{\kappa}_{m_q}^*(m, 0, q) = \frac{1}{\lceil m_q/m \rceil} \sum_{j=1}^{\lceil m_q/m \rceil} \min_{i \in G_{j,q}} \frac{2T_0(C(U_{i,q}, V_{i,q})/\Pi^*(q))}{\log U_{i,q} + \log V_{i,q} - 2 \log q} \quad (3.3.12)$$

when $m_q \rightarrow \infty$, where $(U_{i,q}, V_{i,q})$, $i \in \mathcal{M}_q$, are random iid pairs, each following the conditional cdf $F_q : \mathcal{R}_q(\mathbf{0}) \rightarrow [0, 1]$ defined by

$$F_q(u, v) := \mathbb{P}(U \leq u, V \leq v \mid (U, V) \in \mathcal{R}_q(\mathbf{0}))$$

$$= \frac{1}{\Pi^*(q)} C\left(\min\{u, \varphi^*(q)\}, \min\{v, q^2/\varphi^*(q)\}\right).$$

Thus, $F_q(u, v) = C(u, v)/\Pi^*(q)$ for all $(u, v) \in \mathcal{R}_q(\mathbf{0})$. By the law of large numbers, when $m_q \rightarrow \infty$, we have

$$\widehat{\kappa}_{m_q}^*(m, 0, q) \xrightarrow{P} \kappa^*(m, 0, q) := \mathbb{E} \left[\min_{i \in G_{1,q}} \frac{2T_0(C(U_{i,q}, V_{i,q})/\Pi^*(q))}{\log U_{i,q} + \log V_{i,q} - 2 \log q} \right]. \quad (3.3.13)$$

At first sight, assuming independence of the pairs $(U_{i,q}, V_{i,q})$, $i \in \mathcal{M}_q$, might be alarming, given that we aim at analyzing (dependent) time series data. However, since for reasons of mitigating the influence of slowly varying function on the estimator, we shall use only those observed pairs that are in a (small) neighbourhood of $\mathbf{0}$. This gives rise to nearly independent data, as we shall see in Section 3.A. There is, of course, a gap between independent variables and nearly independent (e.g., white noise) ones, but our estimator, with its block structure, is fairly robust with respect to dependence (Section 3.4). Hence, we can comfortably assume that the random pairs $(U_{i,q}, V_{i,q})$, $i \in \mathcal{M}_q$, are iid.

The following theorem shows that by choosing $m \geq 1$ and $q > 0$ appropriately, we can make $\kappa^*(m, 0, q)$ as close to the TOMD κ^* as desired, and thus by statement (3.3.13), the estimator $\widehat{\kappa}_{m_q}^*(m, 0, q)$ gets close to the TOMD κ^* .

Theorem 3.3.1 *Let $q \in (0, 1]$ be any, and let the cdf F_q^* satisfy the bound*

$$F_q^*(u, v) \geq uv \text{ for all } u, v \in [0, 1]. \quad (3.3.14)$$

Furthermore, let the functions $\varphi^(u)$ and $\psi^*(u) := u^2/\varphi^*(u)$ be strictly increasing. Then $\kappa^*(m, 0, q)$ can be made as close to κ^* as desired by taking sufficiently large $m \geq 1$ and sufficiently small $q > 0$.*

Proof The proof is long but necessary to present in order to show why and how the estimator works. We note at the outset that the joint cdf F_q^* may not have the uniform on $[0, 1]$ marginal distributions, and so bound (3.3.14) does not mean that F_q^* is PQD. Nevertheless, it is this bound that we need in the proof of Theorem 3.3.1, which we start by expressing $\kappa^*(m, 0, q)$ in terms of the survival function of the random variable

$$\xi_q := \frac{2 \log(C(U_q, V_q)/\Pi^*(q))}{\log(U_q V_q / q^2)}, \quad (3.3.15)$$

where (U_q, V_q) follows the cdf F_q . That is, $\kappa^*(m, 0, q)$ is equal to the integral $\int_0^\infty \mathbb{P}(\xi_q > x)^m dx$, because the cardinality of the set $G_{1,q}$ is m and the pairs $(U_{i,q}, V_{i,q})$, $i \in G_{1,q}$, are iid. With the notation

$$(\widetilde{U}_q, \widetilde{V}_q) := \left(\frac{U_q}{\varphi^*(q)}, \frac{V_q \varphi^*(q)}{q^2} \right),$$

we rewrite ξ_q as

$$\xi_q = \frac{2 \log(C(U_q, V_q)/\Pi^*(q))}{\log(U_q/\varphi^*(q)) + \log(V_q \varphi^*(q)/q^2)} = \frac{2 \log F_q^*(\widetilde{U}_q, \widetilde{V}_q)}{\log \widetilde{U}_q + \log \widetilde{V}_q} \in [0, 2], \quad (3.3.16)$$

where the inclusion into the interval $[0, 2]$ is due to the assumed bound (3.3.14). Hence,

$$\kappa^*(m, 0, q) = \int_0^2 \mathbb{P}(\xi_q > x)^m dx. \quad (3.3.17)$$

Note 3.3.2 *The meaning of equations (3.3.16) is to scale the pair $(U_q, V_q) \in \mathcal{R}_q(\mathbf{0})$ into $(\tilde{U}_q, \tilde{V}_q) \in [0, 1]^2$. This allows to shift our focus from the behaviour of random pairs with respect to $q \downarrow 0$ toward the behaviour of F_q^* and the scaling parameters $\varphi^*(q)$ and $q^2/\varphi^*(q)$.*

Since $(U_q, V_q) \in \mathcal{R}_q(\mathbf{0})$ and the functions φ^* and $z \mapsto z^2/\varphi^*(z)$ are increasing, we have

$$(\tilde{U}_q^*, \tilde{V}_q^*) := \left(\frac{\varphi^*(\sqrt{U_q V_q})}{\varphi^*(q)}, \frac{(\sqrt{U_q V_q})^2/\varphi^*(\sqrt{U_q V_q})}{q^2/\varphi^*(q)} \right) \in [0, 1]^2.$$

Hence,

$$\begin{aligned} F_q^*(\tilde{U}_q^*, \tilde{V}_q^*) &= \frac{1}{\Pi^*(q)} C(\varphi^*(\sqrt{U_q V_q}), U_q V_q / \varphi^*(\sqrt{U_q V_q})) \\ &= \frac{1}{\Pi^*(q)} \sup_{x \in [U_q V_q, 1]} C(x, U_q V_q / x) \\ &\geq \frac{1}{\Pi^*(q)} C(U_q, V_q) \\ &= F_q^*(\tilde{U}_q, \tilde{V}_q). \end{aligned}$$

Since $\tilde{U}_q^* \tilde{V}_q^* = \tilde{U}_q \tilde{V}_q = U_q V_q / q^2 \in (0, 1)$, we therefore arrive at the bound

$$\frac{2 \log F_q^*(\tilde{U}_q^*, \tilde{V}_q^*)}{\log \tilde{U}_q^* + \log \tilde{V}_q^*} \leq \frac{2 \log F_q^*(\tilde{U}_q, \tilde{V}_q)}{\log \tilde{U}_q + \log \tilde{V}_q}. \quad (3.3.18)$$

Note that

$$\begin{aligned} \frac{2 \log F_q^*(\tilde{U}_q^*, \tilde{V}_q^*)}{\log \tilde{U}_q^* + \log \tilde{V}_q^*} &= \frac{2 \log \left(C(\varphi^*(\sqrt{U_q V_q}), (\sqrt{U_q V_q})^2/\varphi^*(\sqrt{U_q V_q})) / \Pi^*(q) \right)}{\log(U_q V_q / q^2)} \\ &= \frac{2 \log \left(\Pi^*(\sqrt{U_q V_q}) / \Pi^*(q) \right)}{\log(U_q V_q / q^2)}. \end{aligned} \quad (3.3.19)$$

Due to equation (3.2.1), we have

$$\frac{\Pi^*(\sqrt{U_q V_q})}{\Pi^*(q)} = \frac{\ell^*(\sqrt{U_q V_q})}{\ell^*(q)} \left(\frac{U_q V_q}{q^2} \right)^{\kappa^*/2}$$

and thus, continuing with equations (3.3.19) and taking into account bound (3.3.18), we obtain

$$\frac{2 \log F_q^*(\tilde{U}_q, \tilde{V}_q)}{\log \tilde{U}_q + \log \tilde{V}_q} \geq \frac{2 \log(\ell^*(\sqrt{U_q V_q})/\ell^*(q))}{\log(U_q V_q / q^2)} + \kappa^*. \quad (3.3.20)$$

With the notation

$$o_q := \frac{2 \log(\ell^*(\sqrt{U_q V_q})/\ell^*(q))}{\log(U_q V_q/q^2)},$$

bound (3.3.20) takes the form

$$\xi_q \geq \max\{0, o_q + \kappa^*\}. \quad (3.3.21)$$

We next prove $o_q \xrightarrow{\mathbb{P}} 0$ when $q \downarrow 0$.

With the notation $W_q = \sqrt{U_q V_q}/q$, the statement is equivalent to

$$o_q = \frac{\log(\ell^*(qW_q)/\ell^*(q))}{\log W_q} \xrightarrow{\mathbb{P}} 0. \quad (3.3.22)$$

Hence, we first fix any $\epsilon > 0$. Then we take any $\delta > 0$ and partition the sample space into three events: $\{W_q < \delta\}$, $\{\delta \leq W_q \leq 1 - \delta\}$ and $\{W_q > 1 - \delta\}$. We have the bound

$$\begin{aligned} & \mathbb{P}\left(\left|\frac{\log(\ell^*(qW_q)/\ell^*(q))}{\log W_q}\right| > \epsilon\right) \\ & \leq \mathbb{P}\left(\sup_{w \in [\delta, 1-\delta]} \left|\frac{\log(\ell^*(qw)/\ell^*(q))}{\log w}\right| > \epsilon\right) + \mathbb{P}(W_q < \delta) + \mathbb{P}(W_q > 1 - \delta). \end{aligned} \quad (3.3.23)$$

Since ℓ^* is slowly varying at 0, we have $\ell^*(qw)/\ell^*(q) \rightarrow 1$ when $q \downarrow 0$ for every $w \in (0, 1]$. The convergence is even uniform in $w \in [\delta, 1]$ for any fixed $\delta > 0$ [e.g., 13, Theorem 1.2.1, p. 6], which implies [29, Lemma 1, p. 310] that $\sup_{w \in [\delta, 1]} |\log(\ell^*(qw)/\ell^*(q))|$ converges to 0 when $q \downarrow 0$. Hence, we conclude from bound (3.3.23) that, for any $\epsilon > 0$ and $\delta > 0$,

$$\limsup_{q \rightarrow 0} \mathbb{P}\left(\left|\frac{\log(\ell^*(qW_q)/\ell^*(q))}{\log W_q}\right| > \epsilon\right) \leq \limsup_{q \rightarrow 0} \mathbb{P}(W_q < \delta) + \limsup_{q \rightarrow 0} \mathbb{P}(W_q > 1 - \delta). \quad (3.3.24)$$

Note that the left-hand side of bound (3.3.24) does not depend on $\delta > 0$. As to the first probability on the right-hand side of bound (3.3.24), we have the following result:

$$\begin{aligned} \mathbb{P}(W_q \leq \delta) &= \mathbb{P}(U_q V_q \leq \delta \varphi^*(q) \delta q^2 / \varphi^*(q)) \\ &\leq \mathbb{P}(U_q \leq \delta \varphi^*(q)) + \mathbb{P}(V_q \leq \delta q^2 / \varphi^*(q)) \\ &= \frac{1}{\Pi^*(q)} \left(C(\delta \varphi^*(q), q^2 / \varphi^*(q)) + C(\varphi^*(q), \delta q^2 / \varphi^*(q)) \right) \\ &\leq \frac{2\Pi^*(q \sqrt{\delta})}{\Pi^*(q)} \rightarrow 2\delta^{\kappa^*/2} \end{aligned} \quad (3.3.25)$$

when $q \downarrow 0$, where we have used equation (3.2.1) and property (3.3.1).

We tackle the second probability on the right-hand side of bound (3.3.24) in a different way, starting as follows:

$$\mathbb{P}(W_q > 1 - \delta) = \mathbb{P}(\widetilde{U}_q \widetilde{V}_q > (1 - \delta)^2) \leq \mathbb{P}(F_q^*(\widetilde{U}_q, \widetilde{V}_q) > (1 - \delta)^2), \quad (3.3.26)$$

where the bound holds because $F_q^*(u, v) \geq uv$ for all $u, v \in [0, 1]$. With

$$K_q^*(t) := \mathbb{P}(F_q^*(\widetilde{U}_q, \widetilde{V}_q) \leq t)$$

we continue bound (3.3.26) and have

$$\mathbb{P}(W_q > 1 - \delta) = 1 - K_q^*((1 - \delta)^2) \leq 1 - (1 - \delta)^2 \quad (3.3.27)$$

for every $q \in (0, 1]$ because

$$K_q^*(t) \geq \mathbb{P}(F_q^*(\tilde{U}_q, 1) \leq t) = \mathbb{P}(G_q(\tilde{U}_q) \leq t) = \mathbb{P}(\tilde{U}_q \leq G_q^{-1}(t)) = G_q \circ G_q^{-1}(t) = t,$$

where $G_q : [0, 1] \rightarrow [0, 1]$ denotes the cdf of \tilde{U}_q given by

$$G_q(u) = \frac{C(\varphi^*(q)u, q^2/\varphi^*(q))}{C(\varphi^*(q), q^2/\varphi^*(q))} = \frac{C(\varphi^*(q)u, q^2/\varphi^*(q))}{\Pi^*(q)}.$$

Hence, in view of bounds (3.3.25) and (3.3.27), the entire right-hand side of bound (3.3.24) vanishes when $\delta \downarrow 0$. This concludes the proof of statement (3.3.22).

Fix any $\epsilon \in (0, \kappa^*)$. From equation (3.3.17), we have the upper bound

$$\kappa^*(m, 0, q) \leq \kappa^* + \epsilon + \int_{\kappa^* + \epsilon}^2 \mathbb{P}(\xi_q > x)^m dx. \quad (3.3.28)$$

For the lower bound, we start as follows:

$$\begin{aligned} \kappa^*(m, 0, q) &= \kappa^* - \epsilon - \int_0^{\kappa^* - \epsilon} 1 - \mathbb{P}(\xi_q > x)^m dx + \int_{\kappa^* + \epsilon}^2 \mathbb{P}(\xi_q > x)^m dx \\ &\geq \kappa^* - \epsilon - (\kappa^* - \epsilon)m\mathbb{P}(\xi_q \leq \kappa^* - \epsilon) + \int_{\kappa^* + \epsilon}^2 \mathbb{P}(\xi_q > x)^m dx. \end{aligned} \quad (3.3.29)$$

Using bound (3.3.21), we obtain

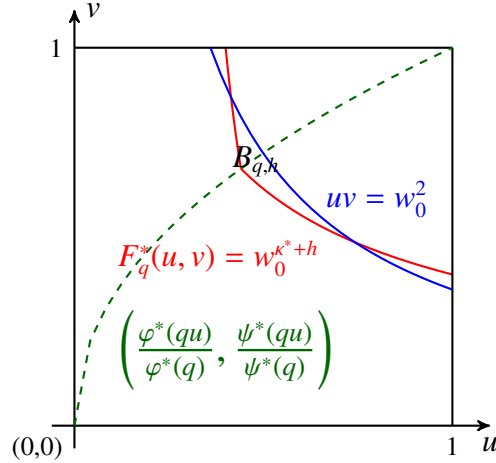
$$\begin{aligned} \mathbb{P}(\xi_q \leq \kappa^* - \epsilon) &\leq \mathbb{P}(\max\{0, o_q + \kappa^*\} \leq \kappa^* - \epsilon) \\ &= \mathbb{P}(o_q \leq -\epsilon, o_q + \kappa^* > 0) + \mathbb{P}(o_q + \kappa^* \leq 0) \\ &= 2\mathbb{P}(|o_q| \geq \epsilon), \end{aligned} \quad (3.3.30)$$

which can be made as small as desired by choosing a small $q > 0$. Hence, due to bounds (3.3.28)–(3.3.30), for any $m \geq 1$ we can choose sufficiently small $\epsilon > 0$ and $q > 0$ such that $\kappa^*(m, 0, q)$ becomes as close to κ^* as desired, provided that the integral $\int_{\kappa^*}^2 \mathbb{P}(\xi_q > x)^m dx$ can be made as small as needed by choosing a sufficiently large m . We prove the latter fact next, without loss of generality assuming that $\kappa^* < 2$, which prevents F_q^* from being the independence copula.

By the Lebesgue dominated convergence theorem, the integral $\int_{\kappa^*}^2 \mathbb{P}(\xi_q > x)^m dx$ converges to 0 when $m \rightarrow \infty$ provided that $\mathbb{P}(\xi_q > x) < 1$ for all $x \in (\kappa^*, 2)$. Hence, we need to show that

$$\mathbb{P}(\xi_q \leq \kappa^* + h) > 0 \quad \text{for every } h \in (0, 2 - \kappa^*). \quad (3.3.31)$$

The proof follows the idea of Sun et al. [19, Theorem 3.2], but there are substantial adjustments, which we present next.

Figure 3.1: The area $B_{q,h}$ and associated curves.

We start with the bound

$$\mathbb{P}(\xi_q \leq \kappa^* + h) = \mathbb{P}\left(\frac{2 \log F_q^*(\tilde{U}_q, \tilde{V}_q)}{\log \tilde{U}_q + \log \tilde{V}_q} \leq \frac{2 \log w_0^{\kappa^* + h}}{\log w_0^2}\right) \geq \mathbb{P}\left((\tilde{U}_q, \tilde{V}_q) \in B_{q,h}\right), \quad (3.3.32)$$

where (see Figure 3.1)

$$B_{q,h} := \left\{ (u, v) \in [0, 1]^2 : uv \leq w_0^2, F_q^*(u, v) > w_0^{\kappa^* + h} \right\}$$

for some $w_0 \in (0, 1)$. With the notation $w_0 = \sqrt{u_q v_q}$, we have

$$u_q = \varphi^*(q w_0) / \varphi^*(q), \quad v_q = w_0^2 \varphi^*(q) / \varphi^*(q w_0),$$

and

$$F_q^*(u_q, v_q) = \frac{\ell^*(q w_0)}{\ell^*(q)} w_0^{\kappa^*} \approx w_0^{\kappa^*} > w_0^{\kappa^* + h}.$$

Hence, $F_q^*(u_q, v_q) > w_0^{\kappa^* + h}$ for sufficiently small $q > 0$ (depending on w_0 and h). Note that

$$F_q^*\left(\frac{\varphi^*(q w_0^{1+h/2\kappa^*})}{\varphi^*(q)}, \frac{w_0^{2+h/\kappa^*} \varphi^*(q)}{\varphi^*(q w_0^{1+h/2\kappa^*})}\right) = \frac{\Pi^*(q w_0^{1+h/2\kappa^*})}{\Pi^*(q)} \approx w_0^{\kappa^* + h/2}.$$

We have

$$F_q^*\left(\frac{\varphi^*(q w_0^{1+h/2\kappa^*})}{\varphi^*(q)}, \frac{w_0^{2+h/\kappa^*} \varphi^*(q)}{\varphi^*(q w_0^{1+h/2\kappa^*})}\right) > w_0^{\kappa^* + h}$$

for sufficiently small $q > 0$ (depending on w_0 and h). By denoting

$$u_{q,h} = \frac{\varphi^*(q w_0^{1+h/2\kappa^*})}{\varphi^*(q)}, \quad v_{q,h} = \frac{w_0^{2+h/\kappa^*} \varphi^*(q)}{\varphi^*(q w_0^{1+h/2\kappa^*})},$$

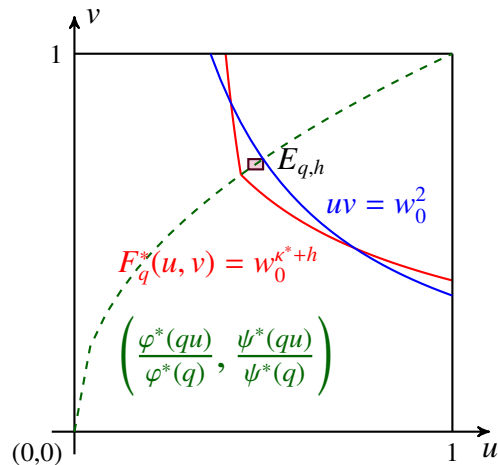


Figure 3.2: The rectangle $E_{q,h} \subset B_{q,h}$ and associated curves.

we have $u_{q,h}v_{q,h} = w_0^{2+h/\kappa^*} < w_0^2$. Thus, $(u_{q,h}, v_{q,h}) \in B_{q,h}$. In particular, $(u_{q,0}, v_{q,0}) = (u_q, v_q)$. Since φ^* and ψ^* are strictly increasing functions, we have $u_{q,h} < u_q$ and $v_{q,h} < v_q$, and so the rectangle (see Figure 3.2)

$$E_{q,h} := (u_{q,h}, u_q] \times (v_{q,h}, v_q]$$

is a non-empty subset of $B_{q,h}$. We have

$$\begin{aligned} \mathbb{P}((\tilde{U}_q, \tilde{V}_q) \in E_{q,h}) &= F_q^*(u_q, v_q) - F_q^*(u_q, v_{q,h}) - F_q^*(u_{q,h}, v_q) + F_q^*(u_{q,h}, v_{q,h}) \\ &= \frac{1}{\Pi^*(q)} \left[C(\varphi^*(qw_0), \psi^*(qw_0)) - C(\varphi^*(qw_0), \psi^*(qw_0^{1+h/2\kappa^*})) \right. \\ &\quad \left. - C(\varphi^*(qw_0^{1+h/2\kappa^*}), \psi^*(qw_0)) + C(\varphi^*(qw_0^{1+h/2\kappa^*}), \psi^*(qw_0^{1+h/2\kappa^*})) \right] \\ &\geq \frac{w_0^{k^*}}{\ell^*(q)} \left[\ell^*(qw_0) - 2\ell^*(qw_0^{1+h/4\kappa^*})w_0^{h/4} + \ell^*(qw_0^{1+h/2\kappa^*})w_0^{h/2} \right] \\ &\rightarrow w_0^{k^*} (1 - w_0^{h/4})^2 > 0 \end{aligned}$$

when $q \downarrow 0$. In summary, $\mathbb{P}((\tilde{U}_q, \tilde{V}_q) \in E_{q,h}) > 0$ for sufficiently small $q > 0$ (depending on w_0 and h). This completes the proof of Theorem 3.3.1.

3.3.4 Illustrative insights into the cdf F_q^*

In addition to F_q^* , $q \in (0, 1]$, we next also discuss – assuming that it exists – the limit

$$F_0^*(u, v) := \lim_{q \downarrow 0} F_q^*(u, v), \quad u, v \in [0, 1],$$

which is an MTD version of the tail order function of Hua and Joe [7, Definition 3].

Example The Marshall-Olkin (M-O) copula $C_{a,b} : [0, 1]^2 \rightarrow [0, 1]$ is defined by [30]

$$C_{a,b}(u, v) = \min(u^{1-a}, uv^{1-b})$$

with parameters $a, b \in [0, 1]$. For every $q \in [0, 1]$, we have [17] $\varphi^*(q) = q^{2b/(a+b)}$ and $\psi^*(q) = q^{2a/(a+b)}$, and so $\Pi^*(u) = u^{\kappa^*}$ with the TOMD $\kappa^* = 2 - 2ab/(a + b)$. Consequently, in view of definition (3.3.5), we have

$$\begin{aligned} F_q^*(u, v) &= \frac{\min\{(q^{2b/(a+b)}u)^{1-a}(q^{2a/(a+b)}v), (q^{2b/(a+b)}u)(q^{2a/(a+b)}v)^{1-b}\}}{q^{2-2ab/(a+b)}} \\ &= \min\{u^{1-a}v, uv^{1-b}\}, \end{aligned}$$

which is equal to $C_{a,b}(u, v)$. Hence, $F_0^* = C_{a,b}$. This implies that both F_q^* and F_0^* satisfy bound (3.3.14), because the M-O copula $C_{a,b}$ is PQD. For additional information on the M-O copula, references, and applications in financial risk management, we refer to, e.g., Asimit et al. [31, 32].

Example The generalized Clayton (GC) copula $C_{\gamma_0, \gamma_1} : [0, 1]^2 \rightarrow [0, 1]$ is defined by

$$C_{\gamma_0, \gamma_1}(u, v) = u^{\gamma_1/\gamma^*} \left(u^{-1/\gamma^*} + v^{-1/\gamma_0} - 1 \right)^{-\gamma_0} \quad (3.3.33)$$

with parameters $\gamma_0 > 0$ and $\gamma_1 \geq 0$, and the notation $\gamma^* := \gamma_0 + \gamma_1$. For details on this copula, its applications to financial risk management, and earlier references, we refer to Su and Furman [33, 34]. In view of formula (3.3.33), we have

$$\begin{aligned} F_q^*(u, v) &= u^{\gamma_1/\gamma^*} \left(\frac{(\varphi^*(q)u)^{-1/\gamma^*} + (\psi^*(q)v)^{-1/\gamma_0} - 1}{\varphi^*(q)^{-1/\gamma^*} + \psi^*(q)^{-1/\gamma_0} - 1} \right)^{-\gamma_0} \\ &= u \left(\frac{\varphi^*(q)^{-1/\gamma^*} + (\psi^*(q)v)^{-1/\gamma_0} u^{1/\gamma^*} - u^{1/\gamma^*}}{\varphi^*(q)^{-1/\gamma^*} + \psi^*(q)^{-1/\gamma_0} - 1} \right)^{-\gamma_0} \\ &\geq u \left(\frac{\varphi^*(q)^{-1/\gamma^*} + (\psi^*(q)v)^{-1/\gamma_0} - 1}{\varphi^*(q)^{-1/\gamma^*} + \psi^*(q)^{-1/\gamma_0} - 1} \right)^{-\gamma_0} \\ &\geq u \left(\frac{(\psi^*(q)v)^{-1/\gamma_0}}{\psi^*(q)^{-1/\gamma_0}} \right)^{-\gamma_0} = uv. \end{aligned}$$

This shows that F_q^* satisfies bound (3.3.14).

Note that $\varphi^*(q)$ satisfies [17, Eq. (6.1)]

$$\varphi^*(q)^{-1/\gamma_0} (\varphi^*(q)^{-1/\gamma^*} - (\gamma_1/\gamma^*)) = (1 - (\gamma_1/\gamma^*)) q^{-2/\gamma_0},$$

which reduces to $\varphi^*(q)^{-1/\gamma^*} - (\gamma_1/\gamma^*) = (1 - (\gamma_1/\gamma^*)) \psi^*(q)^{-1/\gamma_0}$ and gives

$$\begin{aligned} F_q^*(u, v) &= u^{\gamma_1/\gamma^*} \left(\frac{((1 - (\gamma_1/\gamma^*)) \psi^*(q)^{-1/\gamma_0} + (\gamma_1/\gamma^*)) u^{-1/\gamma^*} + (\psi^*(q)v)^{-1/\gamma_0} - 1}{(2 - (\gamma_1/\gamma^*)) \psi^*(q)^{-1/\gamma_0} - (\gamma_0/\gamma^*)} \right)^{-\gamma_0} \\ &\rightarrow u^{\gamma_1/\gamma^*} \left(\frac{[1 - (\gamma_1/\gamma^*)] u^{-1/\gamma^*} + v^{-1/\gamma_0}}{2 - (\gamma_1/\gamma^*)} \right)^{-\gamma_0} =: F_0^*(u, v) \end{aligned}$$

when $q \downarrow 0$.

We next relate F_0^* and the TOMD κ^* via the bound

$$F_0^*(u, v) \leq (uv)^{\kappa^*/2} \quad (3.3.34)$$

for all $u, v \in [0, 1]$, where [17]

$$\kappa^* = 1 + \frac{\gamma_1}{\gamma_1 + 2\gamma_0}. \quad (3.3.35)$$

For any $q \in (0, 1)$, maximizing $F_0^*(x, q^2/x)$ over $x \in [q^2, 1]$ is equivalent to maximizing

$$x \mapsto \frac{\gamma_1}{\gamma^*} \log x - \gamma_0 \log \left(\frac{[1 - (\gamma_1/\gamma^*)]x^{-1/\gamma^*} + (q^2/x)^{-1/\gamma_0}}{2 - (\gamma_1/\gamma^*)} \right).$$

The first-order condition is

$$\frac{\gamma_1}{\gamma^* x} - \frac{\gamma_0 [1 - (\gamma_1/\gamma^*)] (-1/\gamma^*) x^{-1/\gamma^* - 1} + q^{-2/\gamma_0} x^{1/\gamma_0 - 1}}{[1 - (\gamma_1/\gamma^*)] x^{-1/\gamma^*} + (q^2/x)^{-1/\gamma_0}} = 0,$$

which reduces to $x^{-1/\gamma^*} = (q^2/x)^{-1/\gamma_0}$ and gives the solution $x^* = q^{2\gamma^*/(\gamma^* + \gamma_0)}$. Consequently,

$$\begin{aligned} \max_{x \in [q^2, 1]} F_0^*(x, q^2/x) &= F_0^*(q^{2\gamma^*/(\gamma^* + \gamma_0)}, q^{2\gamma_0/(\gamma^* + \gamma_0)}) \\ &= q^{2\gamma_1/(\gamma^* + \gamma_0)} q^{2\gamma_0/(\gamma^* + \gamma_0)} \\ &= q^{1 + \gamma_1/(\gamma_1 + 2\gamma_0)}. \end{aligned}$$

By Furman et al. [17, Eq. (6.2)], this implies bound (3.3.34).

3.4 An illustrative simulation study

The definition of $\widehat{\kappa}_{m,q}^*(m, \theta, q)$ does not require observations (u_i, v_i) , $i \in \mathcal{M}_q$, to arise from independent pairs (U_i, V_i) , although Sun et al. [19] established consistency and other statistical properties under this assumption. Given that we are to apply the estimator on pairs arising from time series, we wish to check the estimator's robustness with respect to dependent data. The underlying time series models that are suitable for financial instruments – such as foreign currency exchange rates, stock market indices, and treasury notes – are complex: they follow, e.g., ARIMA models for the conditional mean and GARCH or some other heteroscedastic models for the conditional variance.

However, since we are concerned with co-movements of extreme losses, which are few and far between inside the original time series, the extreme losses follow models fairly close to white noise. We shall check this conjecture in Section 3.A using a number of portmanteau tests. Hence, in order to check the performance of the TOMD estimator when the iid assumption is (slightly) violated, we conduct a simulated experiment when the observed pairs (u_i, v_i) arise from a time series which is not too far away from being a bivariate white noise [e.g., 35]. Specifically, we next describe a simulation procedure for random pairs whose intra-pair dependence is governed by the generalized Clayton copula and the inter-pair dependence arises from an AR(1) time series model.

To generate $(U_i, V_i) \sim C_{\gamma_0, \gamma_1}$, we start with the conditional cdf of U given $V = v$, which has the expression

$$\mathbb{P}(U \leq u \mid V = v) = \frac{\partial}{\partial v} C_{\gamma_0, \gamma_1}(u, v) = u^{\gamma_1/\gamma^*} \left(u^{-1/\gamma^*} v^{1/\gamma_0} + 1 - v^{1/\gamma_0} \right)^{-\gamma_0 - 1}.$$

With the notation $z = u^{-1/\gamma^*} - 1$, this gives

$$\begin{aligned}\mathbb{P}(U \leq (z+1)^{-\gamma^*} \mid V = v) &= (z+1)^{-\gamma_1} (zv^{1/\gamma_0} + 1)^{-\gamma_0-1} \\ &= \mathbb{P}(Y > z) \mathbb{P}(X > z \mid V = v) \\ &= \mathbb{P}(\min\{X, Y\} > z \mid V = v),\end{aligned}$$

where the random variables $Y \sim \text{Lomax}(\gamma_1, 1)$ and $[X \mid V = v] \sim \text{Lomax}(\gamma_0 + 1, v^{-1/\gamma_0})$ are independent. Hence, to simulate a stationary sequence $(U_i, V_i) \sim C_{\gamma_0, \gamma_1}$, $i \in \mathbb{Z}$, we:

1. generate $Z_i \sim F$ using a time series model;
2. set $V_i := F(Z_i)$;
3. generate independent $X_i \sim \text{Lomax}(\gamma_0 + 1, V_i^{-1/\gamma_0})$ and $Y_i \sim \text{Lomax}(\gamma_1, 1)$;
4. calculate $U_i = (1 + \min\{X_i, Y_i\})^{-\gamma^*}$.

For step (1), we simulate Z_i 's using the strictly stationary and causal AR(1) time series

$$Z_i = \phi Z_{i-1} + \epsilon_i, \quad (\epsilon_i) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1), \quad (3.4.1)$$

where $\phi \in (-1, 1)$ regulates the departure of the sequence (Z_i) from the white noise (ϵ_i) ; if $\phi = 0$, then they coincide. Since we want (Z_i) to carry some dependence, we set $\phi = 0.6$. Hence, the standard deviation is $\sigma = 1/\sqrt{1-\phi^2} = 1.25$, and the autocovariance is $\phi^n/(1-\phi^2)$, $n \geq 0$.

Following steps (1)–(4), we simulate $\{(u_i, v_i) : 1 \leq i \leq n := 500,000\}$ one thousand times, where each pair (u_i, v_i) arises from the generalized Clayton copula C_{γ_0, γ_1} with parameter choices specified in Table 3.4.1. The TOMD κ^* is calculated using formula (3.3.35). Table 3.4.1 also

(γ_0, γ_1)	κ^*	Mean	StDev	A-D	C-vM
(0.1, 0.8)	1.8	1.7920	0.0381	0.7136	0.6853
(0.4, 0.8)	1.5	1.5084	0.0221	0.6982	0.6217
(0.4, 0.2)	1.2	1.2090	0.0127	0.9758	0.9352

Table 3.4.1: Summary of simulation results when $q = 0.05$.

contains various summary statistics under the parameter choices $m = 5$, which is the number of groups $G_{j,q}$, and $q = 0.05$, which is the threshold that serves our working definition of “extreme.” Note that the reported p -values of the Anderson-Darling (A-D) and Cramér-von Mises (C-vM) tests for normality retain the null hypothesis of the simulated estimator values. The fits of these values to the normal distribution are depicted in Figure 3.3. When simulating, we always set θ to $\theta_0 := 10^{-6}$.

For comparison, we use the same $m = 5$ and parameters (γ_0, γ_1) but enlarge the threshold to $q = 0.1$. This increases the number of pairs in the estimated rectangle $\mathcal{R}_q(\mathbf{0})$. Summary statistics are reported in Table 3.4.2, with the A-D and C-vM p -values retaining the null of normality. The fits of the simulated values to the normal distribution are depicted in Figure 3.4.

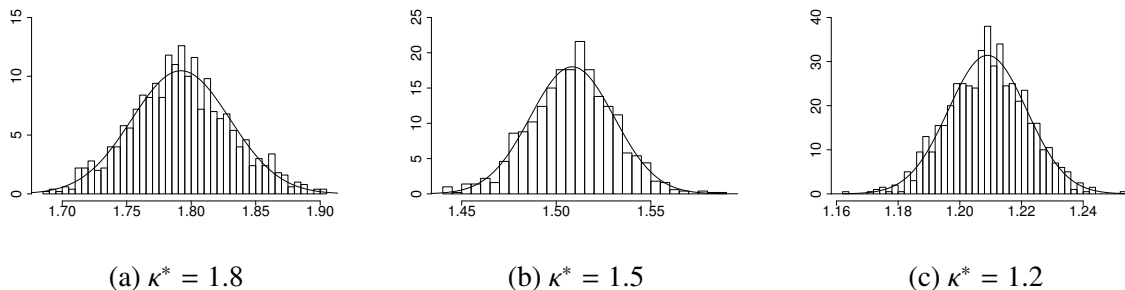


Figure 3.3: Fits of simulated $\widehat{\kappa}_{m_q}^*(5, \theta_0, 0.05)$ to the normal distribution when $q = 0.05$.

(γ_0, γ_1)	κ^*	Mean	StDev	A-D	C-vM
(0.1, 0.8)	1.8	1.8064	0.0197	0.9900	0.9948
(0.4, 0.8)	1.5	1.5149	0.0129	0.9807	0.9514
(0.4, 0.2)	1.2	1.2112	0.0085	0.9322	0.9373

Table 3.4.2: Summary of simulation results when $q = 0.1$.

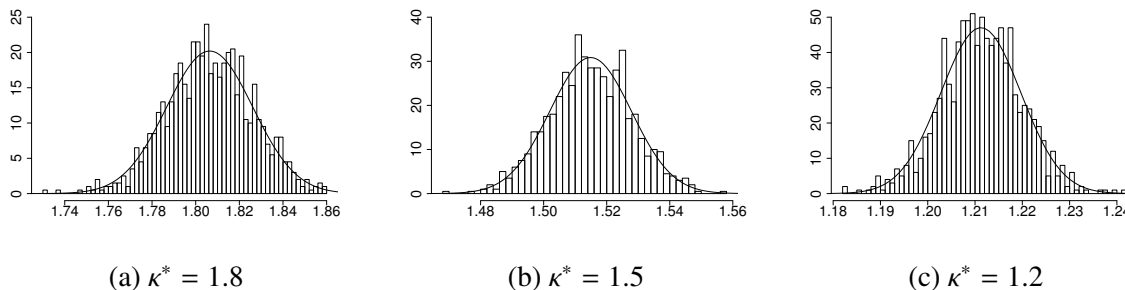


Figure 3.4: Fits of simulated $\widehat{\kappa}_{m_q}^*(5, \theta_0, 0.1)$ to the normal distribution when $q = 0.1$.

3.5 Extreme co-movements of financial instruments

We now explore extreme co-movements of exchange rates of currencies (i.e., CAD/USD, GBP/USD, and JPY/USD), stock market indices (i.e., Dow Jones, S&P 500, and NASDAQ), and diverse financial instruments (i.e., JPY/USD, 10-year Treasury, and NASDAQ) during various periods of time, which have been determined by data availability and/or the date at which we conducted their analyses. Since several data points were missing, we removed the corresponding values from the other time series relevant to our statistical analysis. For the resulting time series (x_t^0) , we then calculated (x_t) defined by

$$x_t = \log(x_t^0) - \log(x_{t-1}^0),$$

which are depicted in the left-hand panels of Figures 3.8–3.10. We have estimated the TOMD κ^* using the procedure described in Section 3.3.1. To compare, we have also estimated the TODD κ using the procedure described in Section 3.3.2. To measure the difference between

the two orders, we have calculated the relative difference in percentages:

$$\text{RD} := \left(\frac{\text{tomd1}}{\text{TODD}} - 1 \right) 100\%.$$

From the theoretical point of view, RD is always either negative or zero, but its empirical version can sometimes become positive, and we shall encounter a few such cases below.

In Figures 3.5–3.7, the upper-triangle panels depict the pairs $(u_i/\varphi^*(q), \varphi^*(q)v_i/q^2)$, $i \in \mathcal{M}_q$, and the lower-triangle panels depict the pairs $(u_i/q, v_i/q)$, $i \in \mathcal{N}_q$, with specially chosen thresholds $q \in (0, 1)$ to be specified and discussed below. In all the examples, we have tested the reasonableness of the PQD assumption for paired extreme pseudo-observations (u_i, v_i) using several hypothesis tests, with findings reported in Section 3.B.

3.5.1 Foreign currency exchange rates

We analyze co-movements of exchange rates of the Canadian and US dollars (CAD/USD), the pound sterling and the US dollar (GBP/USD), and the Japanese yen and the US dollar (JPY/USD) during the period from January 4, 1971, to October 25, 2019 [36]. The differenced log-exchange rates (x_t) are depicted in the three left-hand panels of Figure 3.8. For typographical simplicity, we abbreviate CAD/USD, GBP/USD, JPY/USD into CAD, GBP, JPY, respectively. We couple these exchange rates and analyze the strength of their co-movements in regions (determined by q) of extreme losses. In Figure 3.5, the thresholds $q = 0.075, 0.085,$ and 0.1 have been set for the pairs (JPY, CAD), (JPY, GBP), and (CAD, GBP), respectively. Their TOMD and TODD estimates are reported in Table 3.5.1. Note the relative differences:

JPY	1.3455	0.9026	JPY	1.2967	0.8917
123	CAD	1.5488	112	CAD	1.7759
64	57	GBP	64	53	GBP
(a) TOMD $\widehat{\kappa}_{m_q}^*(5, \theta_0, q)$ and m_q .			(b) TODD $\widehat{\kappa}_{n_q}^{\text{OLS}}$ and n_q .		

Table 3.5.1: The upper-triangle entries of each panel report estimated tail orders, and the lower-triangle entries report the corresponding sample sizes.

$$\begin{aligned} (\text{JPY, CAD}) : \quad \text{RD} &= 3.76\% \\ (\text{JPY, GBP}) : \quad \text{RD} &= 1.22\% \\ (\text{CAD, GBP}) : \quad \text{RD} &= -12.79\% \end{aligned}$$

Although theory says that TOMD is always smaller than TODD, allowing for 5% variability makes the reported positive percentages unsurprising. We therefore conclude that (JPY, CAD) and (JPY, GBP) must have fairly similar maximal and diagonal tail orders, thus implying strong dependence within the pairs. The remaining third pair (CAD, GBP) shows an almost 13% relative decrease in the value of TOMD, thus indicating a notable increase in tail dependence when measured by TOMD if compared to TODD.

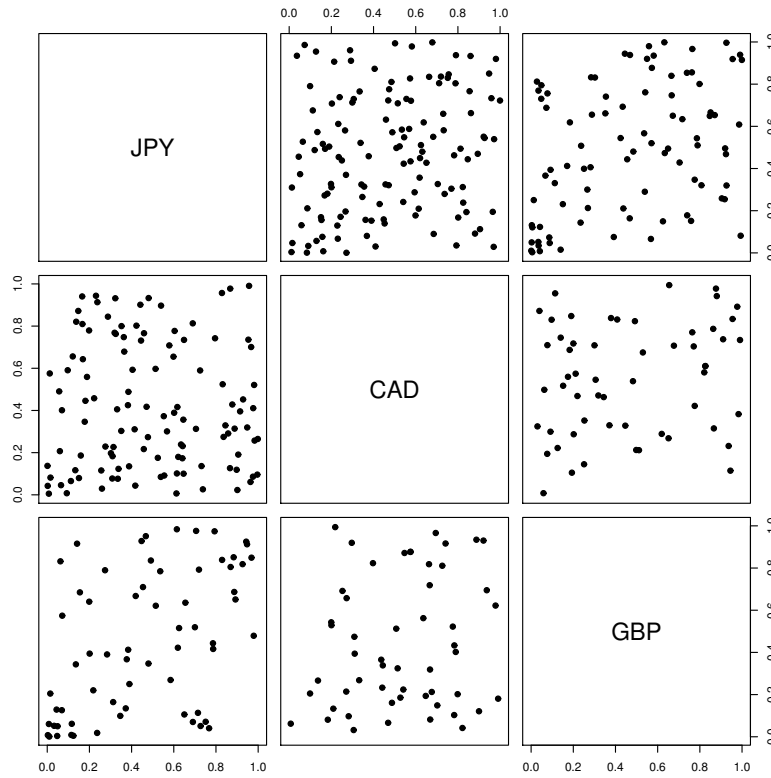


Figure 3.5: Scatterplots of pseudo observations of foreign currency exchange rates.

3.5.2 Stock market indices

We analyze extreme co-movements of the Dow Jones, S&P 500, and NASDAQ during the period from January 4, 1971, to February 28, 2020 [37, 38]. The differenced log-time-series (x_t) are depicted in the three left-hand panels of Figure 3.9. We pair these time series and analyze the strength of their co-movements in regions of extreme losses. In Figure 3.6, the thresholds $q = 0.0075, 0.01, \text{ and } 0.0075$ have been set for the pairs (Dow Jones, S&P 500), (Dow Jones, NASDAQ), and (S&P 500, NASDAQ), respectively. Their TOMD and TODD

Dow Jones	0.8329	0.9816	Dow Jones	1.0788	0.9874
77	S&P 500	0.9422	77	S&P 500	0.9710
68	53	NASDAQ	61	44	NASDAQ

(a) TOMD $\widehat{\kappa}_{m_q}^*(5, \theta_0, q)$ and m_q .

(b) TODD $\widehat{\kappa}_{n_q}^{\text{OLS}}$ and n_q .

Table 3.5.2: The upper-triangle entries of each panel report estimated tail orders, and the lower-triangle entries report the corresponding sample sizes.

estimates are reported in Table 3.5.2. Note the relative differences:

$$\text{(Dow Jones, S\&P 500) : RD} = -22.79\%$$

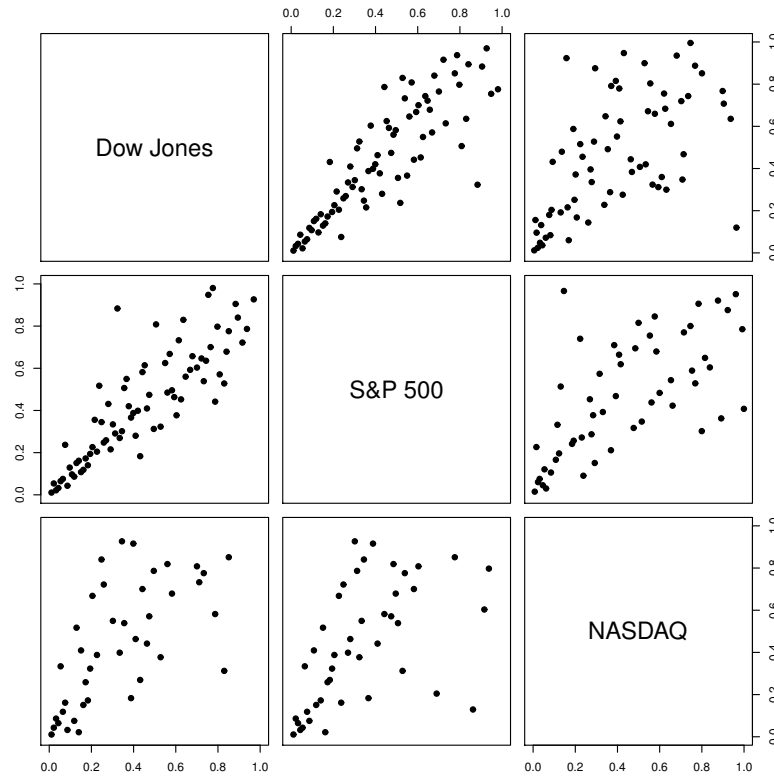


Figure 3.6: Scatterplots of pseudo observations of stock market indices.

(Dow Jones, NASDAQ) : RD = -0.59%

(S&P 500, NASDAQ) : RD = -2.97%

All the values of TOMD are smaller than the corresponding ones of TODD. The pair (Dow Jones, S&P 500) shows an almost 23% decrease in the value of TOMD if compared to the diagonal case, and thus increase in tail dependence when measured by TOMD. Allowing for 5% variability, we conclude that (Dow Jones, NASDAQ) and (S&P 500, NASDAQ) have quite similar TOMD and TODD, thus implying strong dependence within the pairs.

3.5.3 Diverse financial instruments

We analyze pairwise extreme co-movements between daily returns of three different financial instruments: JPY/USD [36], US 10-year Treasury shorthand as US10YT, and NASDAQ [38], which belong to the categories of exchange rates, treasury notes, and stock market indices, respectively. The historical data are from February 5, 1971, to March 3, 2020. The differenced log-time-series (x_t) are depicted in the three left-hand panels of Figure 3.10. We pair these time series and analyze the strength of their co-movements in regions of extreme losses. In Figure 3.7, the thresholds $q = 0.05, 0.05, \text{ and } 0.025$ have been set for the pairs (JPY/USD, US10YT), (JPY/USD, NASDAQ), and (US10YT, NASDAQ), respectively. Their TOMD and

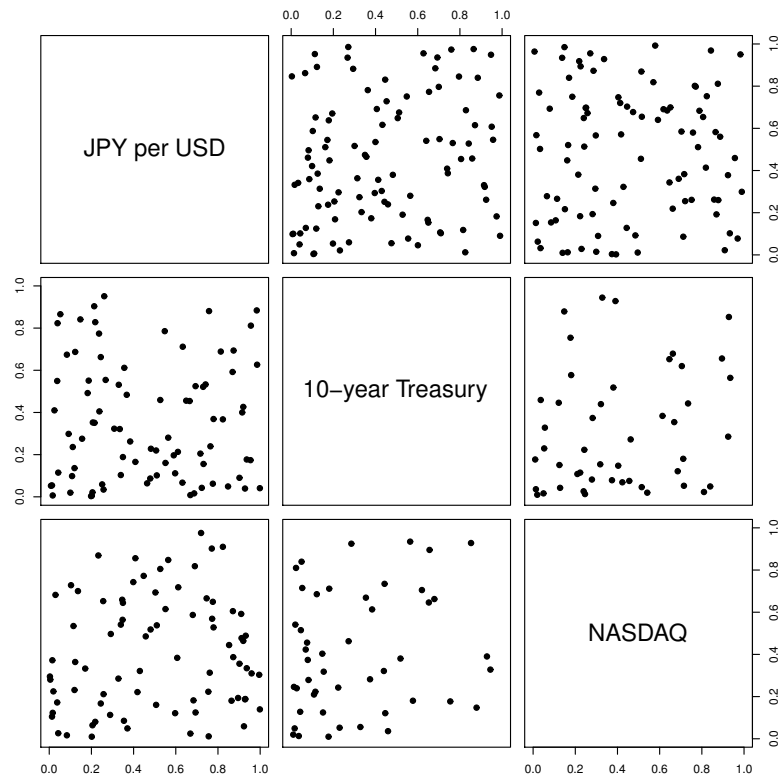


Figure 3.7: The pseudo observations under consideration of diverse financial instruments.

JPY/USD	1.1648	1.4002	JPY/USD	1.3146	1.5276
89	US10YT	0.8516	85	US10YT	1.1292
87	47	NASDAQ	80	47	NASDAQ

(a) TOMD $\widehat{\kappa}_{m_q}^*(5, \theta_0, q)$ and m_q .

(b) TODD $\widehat{\kappa}_{n_q}^{\text{OLS}}$ and n_q .

Table 3.5.3: The upper-triangle entries of each panel report estimated tail orders, and the lower-triangle entries report the corresponding sample sizes.

TODD estimates are reported in Table 3.5.3. Note the relative differences:

$$\begin{aligned}
 (\text{JPY/USD, US10YT}) : & \quad \text{RD} = -11.40\% \\
 (\text{JPY/USD, NASDAQ}) : & \quad \text{RD} = -8.34\% \\
 (\text{US10YT, NASDAQ}) : & \quad \text{RD} = -24.58\%
 \end{aligned}$$

All the values of TOMD are smaller than the corresponding ones of TODD, and all the pairs show considerable decrease in the values of TOMD if compared to the diagonal case.

3.6 Conclusion

In this chapter we have developed a substantial extension of the procedure of Chapter 2 for assessing the maximal strength of co-movements of extreme losses when original data follow dependent dynamical models. We have explored the performance of the modification on simulated bivariate time series. The study has shown that the block-wise construction of the estimator of maximal tail dependence successfully handles time series structures of data sets and, in turn, from them arising extreme co-movements.

We have tested the validity of underlying theoretical assumptions using statistical tests, and also discussed ways for calculating critical values. In addition, we have provided an extensive study of thresholds below which time-series data give rise to what we deem to be extreme losses.

The strength of maximal dependence as well as of the classical diagonal dependence have been explored and compared for a number of financial instruments, such as foreign exchange rates of several major currencies, stock market indices, and treasury notes.

Appendix

3.A Thresholds and pseudo observations

In Figures 3.8–3.10 we have depicted the differenced log-time-series $x_t = \log(x_t^0) - \log(x_{t-1}^0)$ (left-hand panels) and the extreme pseudo-observations (right-hand panels) that arise from the time series data specified in Section 3.5.

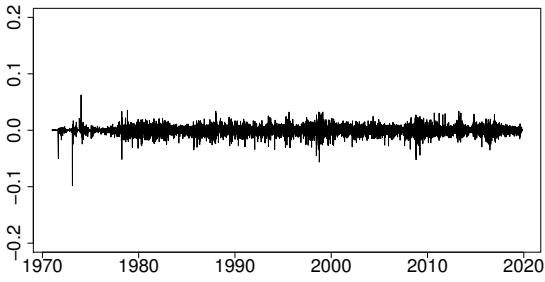
With thresholds $q \in (0, 1)$ reported in Table 3.A.1, the time series give rise to paired extreme

Pairs	Threshold q						
	0.0075	0.01	0.025	0.05	0.075	0.085	0.1
(JPY, CAD)	100(123)
(JPY, GBP)	90(64)	...
(CAD, GBP)	95(57)
(Dow Jones, S&P 500)	60(77)
(Dow Jones, NASDAQ)	...	77(68)
(S&P 500, NASDAQ)	76(53)
(JPY/USD, US10YT)	100(89)
(JPY/USD, NASDAQ)	100(87)
(US10YT, NASDAQ)	88(47)

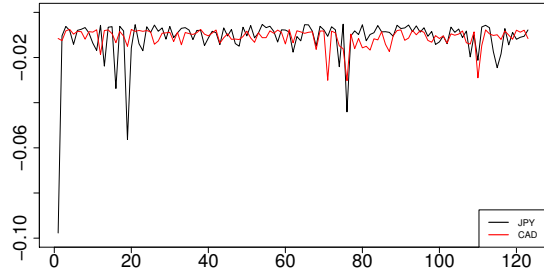
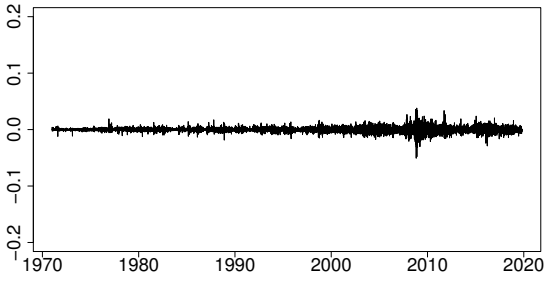
Table 3.A.1: The percentages of p -values retaining the null of white noise at the significance level $\alpha = 0.05$ alongside the sample sizes m_q (in parentheses) for appropriate choices of q .

pseudo-observations that resemble white noise; see the right-hand panels of Figures 3.8–3.10. To substantiate this claim, we have run several portmanteau tests for the null hypothesis

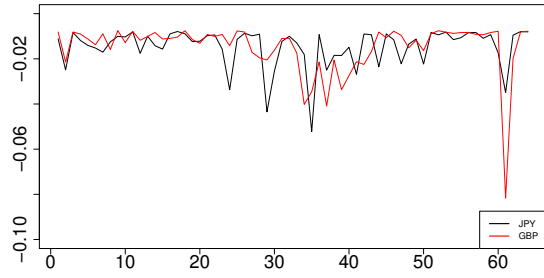
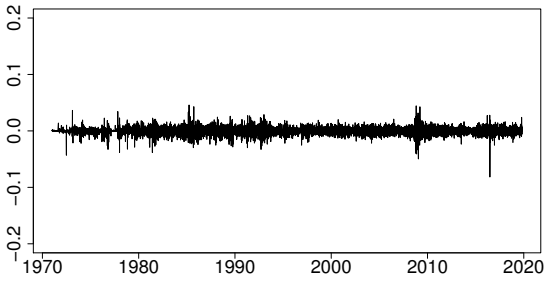
$$\mathcal{H}_0 : \mathbf{\Gamma}_L = \mathbf{0}, \quad L = 1, \dots, 20,$$



(a) JPY/USD

(b) (JPY, CAD), $q = 0.075$, $m_q = 123$.

(c) CAD/USD

(d) (JPY, GBP), $q = 0.085$, $m_q = 64$.

(e) GBP/USD

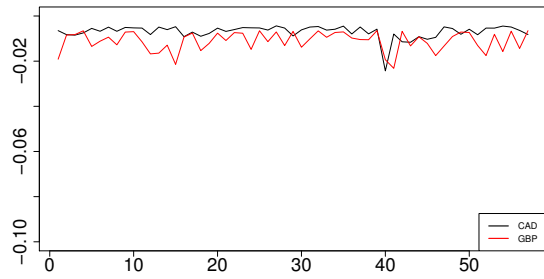
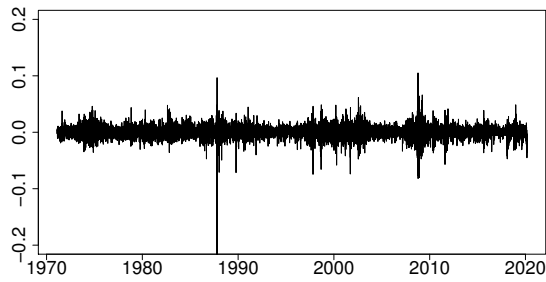
(f) (CAD, GBP), $q = 0.1$, $m_q = 57$.

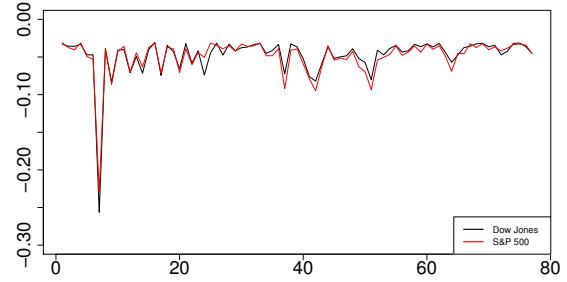
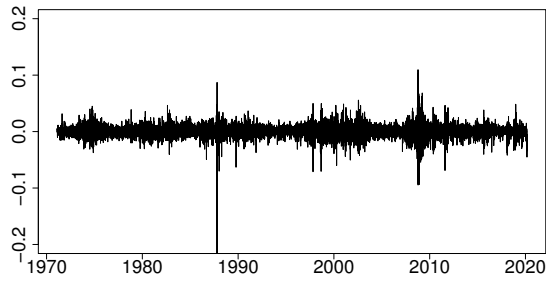
Figure 3.8: Original x_t 's (left-hand panels) and the pairs of extreme pseudo-observations (right-hand panels) for foreign currency exchange from January 4, 1971, to October 25, 2019.

where $\Gamma_L = \text{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t-L})$ and $(\boldsymbol{\varepsilon}_t)_{t=1}^{m_q}$ are the residuals obtained by fitting the original data to the time series model VARMA for sufficiently many lags [39]. The selected portmanteau tests include those of [39], [40], [41], [42], and [43]. The percentages of p -values above the 5% significance level (meaning that the null of white noise is retained) are reported in Table 3.A.1, where we also report the sample sizes m_q .

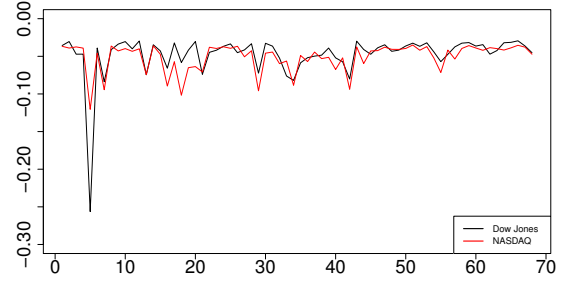
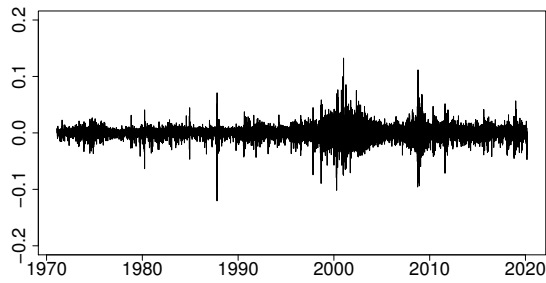
The different choices of $q \in (0, 1)$ warrant an explanation. First, we want to work with as small $q > 0$ as possible, mainly due to two reasons: first, the estimator's deterministic bias becomes small (recall Theorem 3.3.1), and second, the two-dimensional time series of extreme pseudo-observations becomes nearly white-noise. Working close to white noise is important



(a) Dow Jones

(b) (Dow Jones, S&P 500), $q = 0.0075$, $m_q = 77$.

(c) S&P 500

(d) (Dow Jones, NASDAQ), $q = 0.01$, $m_q = 68$.

(e) NASDAQ

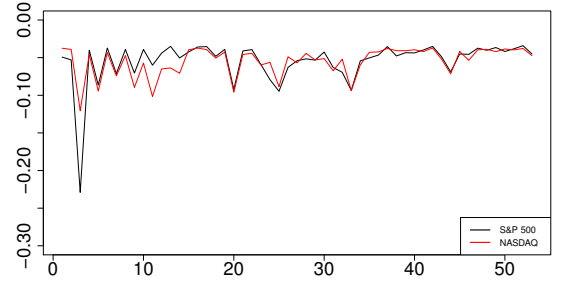
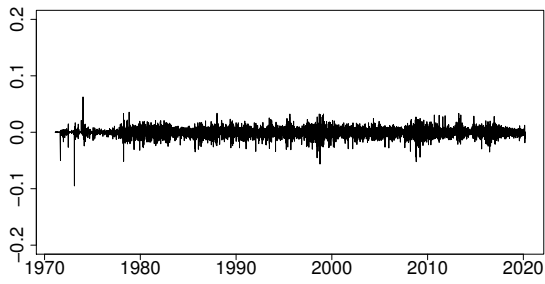
(f) (S&P 500, NASDAQ), $q = 0.0075$, $m_q = 53$.

Figure 3.9: Original x_i 's (left-hand panels) and the pairs of extreme pseudo-observations (right-hand panels) for stock market indices from January 4, 1971, to February 28, 2020.

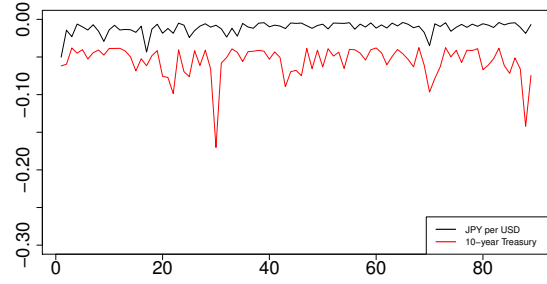
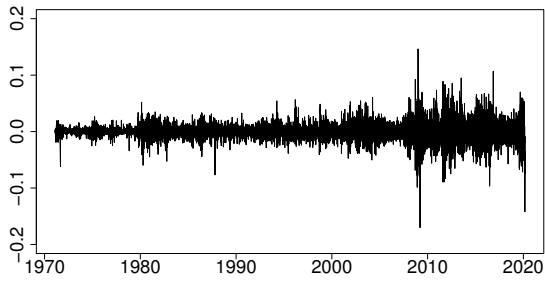
in order to reliably calculate critical values of hypothesis tests for bound (3.3.14), which verify the applicability of Theorem 3.3.1.

3.B Testing the validity of bound (3.3.14)

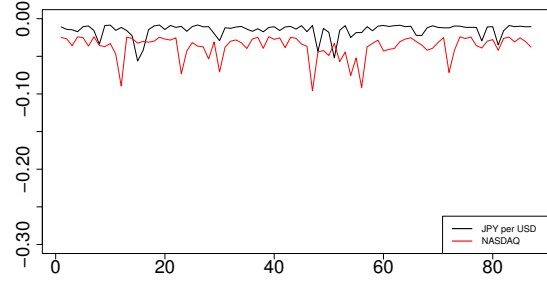
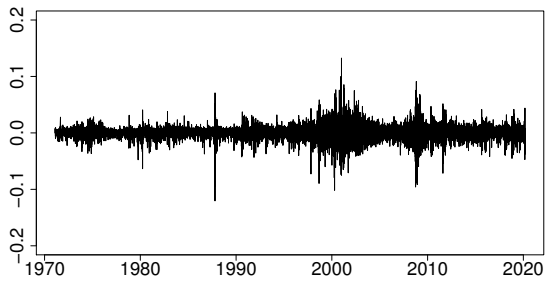
In all the real time-series examples that we explore, we statistically test the reasonableness of bound (3.3.14) for the cdf F_q^* defined in equation (3.3.5). For this, we adapt the Kolmogorov-Smirnov (K-S), Cramér-von Mises (C-vM), Anderson-Darling (A-D) one-sided statistics [cf.



(a) JPY/USD

(b) (JPY/USD, US10YT), $q = 0.05$, $m_q = 89$.

(c) US10YT

(d) (JPY/USD, NASDAQ), $q = 0.05$, $m_q = 87$.

(e) NASDAQ

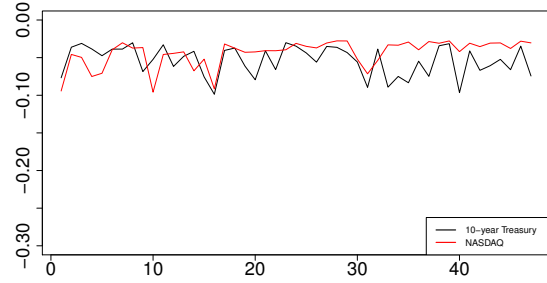
(f) (US10YT, NASDAQ), $q = 0.025$, $m_q = 47$.

Figure 3.10: Original x_t 's (left-hand panels) and the pairs of extreme pseudo-observations (right-hand panels) for diverse financial instruments from February 5, 1971, to March 3, 2020.

44]:

$$\sqrt{m_q} \sup_{(u,v) \in [0,1]^2} \{uv - F_{q, \mathcal{M}_q}^*(u, v)\}_+, \quad (3.B.1)$$

$$m_q \int_{[0,1]^2} (\{uv - F_{q, \mathcal{M}_q}^*(u, v)\}_+)^2 dF_{q, \mathcal{M}_q}^*(u, v), \quad (3.B.2)$$

$$m_q \int_{[0,1]^2} \frac{(\{uv - F_{q, \mathcal{M}_q}^*(u, v)\}_+)^2}{u(1-u)v(1-v)} dF_{q, \mathcal{M}_q}^*(u, v), \quad (3.B.3)$$

respectively, where F_{q, \mathcal{M}_q}^* is defined in equation (3.3.6). Specifically, we use these statistics to test the null hypothesis H_0 of having bound (3.3.14) versus the alternative H_1 of not having the bound: :

$$H_0 : F_q^*(u, v) \geq uv \text{ for all } (u, v) \in [0, 1]^2$$

$$H_1 : F_q^*(u, v) < uv \text{ for some } (u, v) \in [0, 1]^2$$

The critical values of these tests are obtained by sampling from the pairs of pseudo observations. Namely, we calculate the test statistics, repeat the procedure $N = 10,000$ times obtain so many values of the test statistics, and finally calculate the 95th percentiles of the respective test-statistic values. The decision rule is to retain the null hypothesis H_0 if the test statistic is smaller than the critical value, and to reject it otherwise. Specifically, we have adopted the following four-step procedure [cf. 45]:

1. choose the independence copula as the reference, because it is the closest to the alternative;
2. draw N independent samples, each of size n , from the independence copula;
3. calculate K-S, C-vM, and A-D test statistics for each sample, thus obtaining three sets, each of cardinality N , of numerical values;
4. given a significance level α , define the critical values of the three tests as the 95th percentiles of the K-S, C-vM, and A-D values obtained in the previous step.

The obtained results are summarized in Tables 3.B.1–3.B.3, where the abbreviations “Stat,” “Crit,” and “Deci” stand for the test statistic value, the critical value, and the decision, respectively. The decision is to retain the null H_0 when the statistic is smaller than the critical value.

Test	(JPY, CAD)			(JPY, GBP)			(CAD, GBP)		
	$q = 0.075, m_q = 123$			$q = 0.085, m_q = 64$			$q = 0.1, m_q = 57$		
	Stat	Crit	Deci	Stat	Crit	Deci	Stat	Crit	Deci
K-S	0.0189	1.2119	H_0	0.0404	1.1195	H_0	0.5781	1.0989	H_0
C-vM	0.0000	0.2013	H_0	0.0000	0.1844	H_0	0.0330	0.1737	H_0
A-D	0.0008	28.7139	H_0	0.0064	23.9732	H_0	6.0110	24.5422	H_0

Table 3.B.1: Testing H_0 vs H_1 of pseudo observations of foreign currency exchange rates.

In every case, the three tests have retained the null H_0 .

To gain additional insight, in next Section 3.C we also test the null of the equation $F_q^*(u, v) = uv$ for all $(u, v) \in [0, 1]^2$, which is a “boundary” case for H_0 .

3.C Testing the boundary case of bound (3.3.14)

In the examples of Section 3.B, all of which retained the null H_0 for the cdf F_q^* , we now statistically test the reasonableness of the boundary case $F_q^*(u, v) = uv$ for all $(u, v) \in [0, 1]^2$. For

Test	(Dow Jones, S&P 500) $q = 0.0075, m_q = 77$			(Dow Jones, NASDAQ) $q = 0.01, m_q = 68$			(S&P 500, NASDAQ) $q = 0.0075, m_q = 53$		
	Stat	Crit	Deci	Stat	Crit	Deci	Stat	Crit	Deci
K-S	0.0000	1.1474	H_0	0.0000	1.1350	H_0	0.0000	1.0905	H_0
C-vM	0.0000	0.1922	H_0	0.0000	0.1900	H_0	0.0000	0.1786	H_0
A-D	0.0000	26.5523	H_0	0.0000	24.7340	H_0	0.0000	22.9812	H_0

Table 3.B.2: Testing H_0 vs H_1 of pseudo observations of stock market indices.

Test	(JPY/USD, US10YT) $q = 0.05, m_q = 89$			(JPY/USD, NASDAQ) $q = 0.05, m_q = 87$			(US10YT, NASDAQ) $q = 0.025, m_q = 47$		
	Stat	Crit	Deci	Stat	Crit	Deci	Stat	Crit	Deci
K-S	0.0000	1.1715	H_0	0.1913	1.1766	H_0	0.0000	1.0739	H_0
C-vM	0.0000	0.1889	H_0	0.0007	0.2004	H_0	0.0000	0.1794	H_0
A-D	0.0000	25.2966	H_0	0.6637	26.5172	H_0	0.0000	23.8698	H_0

Table 3.B.3: Testing H_0 vs H_1 of pseudo observations of diverse financial instruments.

the task, we employ the Kolmogorov-Smirnov (K-S), Cramér-von Mises (C-vM), Anderson-Darling (A-D) one-sided statistics [cf. 46]:

$$\sqrt{m_q} \sup_{(u,v) \in [0,1]^2} \{F_{q, \mathcal{M}_q}^*(u, v) - uv\}_+, \quad (3.C.1)$$

$$m_q \int_{[0,1]^2} (\{F_{q, \mathcal{M}_q}^*(u, v) - uv\}_+)^2 dF_{q, \mathcal{M}_q}^*(u, v), \quad (3.C.2)$$

$$m_q \int_{[0,1]^2} \frac{(\{F_{q, \mathcal{M}_q}^*(u, v) - uv\}_+)^2}{u(1-u)v(1-v)} dF_{q, \mathcal{M}_q}^*(u, v), \quad (3.C.3)$$

respectively. Specifically, we use statistics (3.C.1)–(3.C.3) to test the hypothesis:

$$H_0^* : F_q^*(u, v) = uv \text{ for all } (u, v) \in [0, 1]^2$$

$$H_1^* : F_q^*(u, v) > uv \text{ for some } (u, v) \in [0, 1]^2$$

Note 3.C.1 We already noted that the cdf F_q^* does not, in general, have uniform marginal distributions, and so the null hypothesis H_0^* does not, in general, mean the null of independence. However, it is not too far away from being such a hypothesis. Indeed, if C is the independence copula, then, as immediately follows from equation (3.3.5), we have $F_q^*(u, v) = uv$ for all $(u, v) \in [0, 1]^2$. Hence, if we reject H_0^* , this means that we also reject the null of independence. Of course, the opposite may not be true: if we retain H_0^* , then this does not, in general, mean that we retain the null of independence. We shall come back to this at the very end of this section, given certain numerical values of TOMD.

Coming now back to our main discussion, we note that the procedures for calculating critical values for any of the three tests for H_0^* vs H_1^* based on statistics (3.C.1)–(3.C.3) are analogous to those we used in Section 3.B in the case of H_0 vs H_1 . The decision to retain the null H_0^* is, of course, taken when the statistic is smaller than the critical value. Our findings are summarized in Tables 3.C.1–3.C.2.

Test	(JPY, CAD)			(JPY, GBP)			(CAD, GBP)		
	$q = 0.075, m_q = 123$			$q = 0.085, m_q = 64$			$q = 0.1, m_q = 57$		
	Stat	Crit	Deci	Stat	Crit	Deci	Stat	Crit	Deci
K-S	1.4859	1.3546	H_1^*	1.4837	1.3386	H_1^*	0.9139	1.3231	H_0^*
C-vM	0.5874	0.2969	H_1^*	0.6968	0.3198	H_1^*	0.0782	0.3299	H_0^*
A-D	25.6607	44.9833	H_0^*	91.2540	50.1569	H_1^*	4.0293	49.6459	H_0^*

Table 3.C.1: Testing H_0^* vs H_1^* of pseudo observations of foreign currency exchange rates.

Test	(Dow Jones, S&P 500)			(Dow Jones, NASDAQ)			(S&P 500, NASDAQ)		
	$q = 0.0075, m_q = 77$			$q = 0.01, m_q = 68$			$q = 0.0075, m_q = 53$		
	Stat	Crit	Deci	Stat	Crit	Deci	Stat	Crit	Deci
K-S	2.4058	1.3441	H_1^*	1.7760	1.3335	H_1^*	1.6858	1.3208	H_1^*
C-vM	2.9460	0.3107	H_1^*	1.3744	0.3269	H_1^*	1.3100	0.3343	H_1^*
A-D	90.9184	47.4610	H_1^*	61.2180	49.3820	H_1^*	51.2729	51.2274	H_1^*

Table 3.C.2: Testing H_0^* vs H_1^* of pseudo observations of stock market indices.

Test	(JPY/USD, US10YT)			(JPY/USD, NASDAQ)			(US10YT, NASDAQ)		
	$q = 0.05, m_q = 89$			$q = 0.05, m_q = 87$			$q = 0.025, m_q = 47$		
	Stat	Crit	Deci	Stat	Crit	Deci	Stat	Crit	Deci
K-S	1.8927	1.3442	H_1^*	0.8564	1.3465	H_0^*	2.1463	1.3158	H_1^*
C-vM	1.0709	0.3092	H_1^*	0.2266	0.3142	H_0^*	1.4086	0.3345	H_1^*
A-D	79.7641	49.0677	H_1^*	33.5631	48.3546	H_0^*	71.3135	52.2009	H_1^*

Table 3.C.3: Testing H_0^* vs H_1^* of pseudo observations of diverse financial instruments.

Hence, independence of the first and second coordinates of the pairs (CAD, GBP) and (JPY/USD, NASDAQ) has not been rejected (recall Note 3.C.1) by any of the three tests, but the estimated TOMDs are equal to 1.5488 and 1.4002, respectively, as seen from Tables 3.5.1 and 3.5.3. These TOMD values suggest that the coordinates of the two aforementioned pairs may actually be fairly dependent.

Bibliography

- [1] M. Pericoli, M. Sbracia, A primer on financial contagion, *Journal of Economic Surveys* 17 (4) (2003) 571–608. 32

- [2] F. Durante, R. Pappadà, N. Torelli, Clustering of financial time series in risky scenarios, *Advances in Data Analysis and Classification* 8 (4) (2014) 359–376. [32](#)
- [3] A. Bücher, S. Jäschke, D. Wied, Nonparametric tests for constant tail dependence with an application to energy and finance, *Journal of Econometrics* 187 (1) (2015) 154–168. [32](#)
- [4] H. White, T.-H. Kim, S. Manganelli, VAR for VaR: Measuring tail dependence using multivariate regression quantiles, *Journal of Econometrics* 187 (1) (2015) 169–188. [32](#), [33](#)
- [5] H. Joe, Parametric families of multivariate distributions with given margins, *Journal of Multivariate Analysis* 46 (2) (1993) 262–282. [32](#)
- [6] S. Coles, J. Heffernan, J. Tawn, Dependence measures for extreme value analyses, *Extremes* 2 (4) (1999) 339–365.
- [7] L. Hua, H. Joe, Tail order and intermediate tail dependence of multivariate copulas, *Journal of Multivariate Analysis* 102 (10) (2011) 1454–1471. [44](#)
- [8] H. Joe, *Dependence Modeling with Copulas*, Chapman and Hall/CRC, 2014. [32](#)
- [9] A. J. McNeil, R. Frey, P. Embrechts, *Quantitative Risk Management*, Princeton University Press, 2015. [32](#)
- [10] G. Draisma, H. Drees, A. Ferreira, L. D. Haan, Bivariate tail estimation: dependence in asymptotic independence, *Bernoulli* 10 (2). [32](#)
- [11] G. Frahm, M. Junker, R. Schmidt, Estimating the tail-dependence coefficient: Properties and pitfalls, *Insurance: Mathematics and Economics* 37 (1) (2005) 80–100.
- [12] P. Krupskii, H. Joe, Nonparametric estimation of multivariate tail probabilities and tail dependence coefficients, *Journal of Multivariate Analysis* 172 (2019) 147–161. [32](#)
- [13] N. H. Bingham, C. M. Goldie, J. L. Teugels, *Regular Variation*, Cambridge University Press, 1987. [32](#), [36](#), [41](#)
- [14] R. B. Nelsen, *An Introduction to Copulas*, Springer New York, 2006. [32](#)
- [15] F. Durante, C. Sempi, *Principles of Copula Theory*, Chapman and Hall/CRC, 2015. [32](#)
- [16] R. Wang, R. Zitikis, An axiomatic foundation for the expected shortfall, *Management Science* 67 (3) (2021) 1413–1429. [32](#)
- [17] E. Furman, J. Su, R. Zitikis, Paths and indices of maximal tail dependence, *ASTIN Bulletin* 45 (3) (2015) 661–678. [32](#), [33](#), [34](#), [45](#), [46](#)
- [18] E. Furman, A. Kuznetsov, J. Su, R. Zitikis, Tail dependence of the Gaussian copula revisited, *Insurance: Mathematics and Economics* 69 (2016) 97–103. [32](#), [34](#)
- [19] N. Sun, C. Yang, R. Zitikis, A statistical methodology for assessing the maximal strength of tail dependence, *ASTIN Bulletin* 50 (3) (2020) 799–825. [33](#), [35](#), [36](#), [37](#), [42](#), [46](#)

- [20] J. Kiefer, J. Wolfowitz, On the deviations of the empiric distribution function of vector chance variables, *Transactions of the American Mathematical Society* 87 (1) (1958) 173–173. 35
- [21] J. Kiefer, et al., On large deviations of the empiric DF of vector chance variables and a law of the iterated logarithm., *Pacific journal of mathematics* 11 (2) (1961) 649–660. 35
- [22] H. Dehling, T. Mikosch, M. Sørensen (Eds.), *Empirical Process Techniques for Dependent Data*, Birkhäuser Boston, 2002. 35
- [23] Y. Davydov, R. Zitikis, On weak convergence of random fields, *Annals of the Institute of Statistical Mathematics* 60 (2) (2008) 345–365. 35
- [24] A. Kontorovich, R. Weiss, Uniform Chernoff and Dvoretzky-Kiefer-Wolfowitz-Type inequalities for Markov Chains and related processes, *Journal of Applied Probability* 51 (4) (2014) 1100–1113. 35
- [25] V. Vovk, R. Wang, Combining p -values via averaging, *Biometrika* 107 (4) (2020) 791–808. 35
- [26] X. Gabaix, R. Ibragimov, Rank - $1/2$: a simple way to improve the ols estimation of tail exponents, *Journal of Business & Economic Statistics* 29 (1) (2011) 24–39. 36, 38
- [27] Basel Committee on Banking Supervision, Minimum Capital Requirements for Market Risk (January, 2016), URL <https://www.bis.org/bcbs/publ/d352.htm>, 2016. 36
- [28] Basel Committee on Banking Supervision, Minimum Capital Requirements for Market Risk (February, 2019), URL <https://www.bis.org/bcbs/publ/d457.htm>, 2019. 36
- [29] R. Bojanic, E. Seneta, Slowly varying functions and asymptotic relations, *Journal of Mathematical Analysis and Applications* 34 (2) (1971) 302–315. 41
- [30] A. W. Marshall, I. Olkin, A multivariate exponential distribution, *Journal of the American Statistical Association* 62 (317) (1967) 30–44. 44
- [31] A. V. Asimit, E. Furman, R. Vernic, On a multivariate Pareto distribution, *Insurance: Mathematics and Economics* 46 (2) (2010) 308–316. 45
- [32] A. V. Asimit, E. Furman, R. Vernic, Statistical inference for a new class of multivariate Pareto distributions, *Communications in Statistics-Simulation and Computation* 45 (2) (2016) 456–471. 45
- [33] J. Su, E. Furman, A form of multivariate Pareto distribution with applications to financial risk measurement, *ASTIN Bulletin: The Journal of the IAA* 47 (1) (2017) 331–357. 45
- [34] J. Su, E. Furman, Multiple risk factor dependence structures: Copulas and related properties, *Insurance: Mathematics and Economics* 74 (2017) 109–121. 45
- [35] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th Edition, Wiley, 2015. 46

- [36] Federal Reserve Board, Foreign Exchange Rates – H.10., URL <https://www.federalreserve.gov/releases/h10/>, board of Governors of the Federal Reserve System, Washington, D.C., 2020. 49, 51
- [37] Wall Street Journal, Market data: Dow Jones Industrial Average, URL <https://www.wsj.com/market-data/quotes/index/DJIA/historical-prices>, 2020. 50
- [38] Yahoo Finance, Market data, URL <https://ca.finance.yahoo.com/>, 2020. 50, 51
- [39] E. Mahdi, A. I. McLeod, Improved multivariate portmanteau test, *Journal of Time Series Analysis* 33 (2) (2012) 211–222. 54
- [40] G. E. P. Box, D. A. Pierce, Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *Journal of the American Statistical Association* 65 (332) (1970) 1509–1526. 54
- [41] G. M. Ljung, G. E. P. Box, On a measure of lack of fit in time series models, *Biometrika* 65 (2) (1978) 297–303. 54
- [42] J. R. M. Hosking, The multivariate Portmanteau statistic, *Journal of the American Statistical Association* 75 (371) (1980) 602–608. 54
- [43] W. K. Li, A. I. McLeod, Distribution of the residual autocorrelations in multivariate arma time series models, *Journal of the Royal Statistical Society: Series B (Methodological)* 43 (2) (1981) 231–239. 54
- [44] I. Gijbels, D. Sznajder, Positive quadrant dependence testing and constrained copula estimation, *Canadian Journal of Statistics* 41 (1) (2013) 36–64. 56
- [45] I. Gijbels, M. Omelka, D. Sznajder, Positive quadrant dependence tests for copulas, *Canadian Journal of Statistics* 38 (4) (2010) 555–581. 57
- [46] C.-F. Tang, D. Wang, H. E. Barmi, J. M. Tebbs, Testing for positive quadrant dependence, *The American Statistician* 75 (1) (2019) 23–30. 58

Chapter 4

Detecting systematic anomalies affecting systems when inputs are stationary time series

4.1 Introduction

Control systems are often exposed to errors, intrusions, and other anomalies whose detection in a timely fashion is of paramount importance. Computer systems monitor and control a myriad of physical processes, and their protection against random errors, deliberate intrusions [e.g., 1–4], false data injections [e.g., 5], and other disruptors is of much interest.

A vast number of methods have been proposed for the purpose. For example, we find methods based on deep learning [e.g., 6], probabilistic arguments [e.g., 7, 8], artificial neural networks [e.g., 9], and Fourier techniques [e.g., 10]. Chen et al. [11] discuss the effects of an early warning mechanism on system’s reliability. For a recently developed LSTM-based intrusion detection system for in-vehicle can bus communications, we refer to [12].

For complementary reviews of anomaly detection, we refer to Chandola et al. [13], Bhuyan et al. [14], Fisch [15]. For general information on various facets of risk and with them associated problems, we refer to, e.g., Aven et al. [16], Zio [17]. For more specialized discussions on the topic, we refer to, e.g., Cheng et al. [18], Liang et al. [5].

The emphasis in the present chapter is on temporal aspects and dependence structures that arise in this area of research. There have been a number of studies tackling these issues from several perspectives. For example, Barahona and Poon [19] present a computational procedure capable of robust and sensitive statistical detection of deterministic and chaotic dynamics in short and noisy time series. Hu et al. [20] explore the role of dependence when assessing quantities such as system-compromise probabilities and the cost of attacks, which are then used to develop optimization strategies. Dasgupta and Li [21] tackle the problem of assessing whether temporal clusters in randomly occurring sequences of events are genuinely random.

Furthermore, Fisch et al. [22] propose what is called the collective and point anomalies (CAPA) method for detecting point anomalies (i.e., outliers in the statistical language) as well as anomalous segments, or collective anomalies. The method is suitable when collective anomalies are characterised by either a change in mean, variance, or both, and it is capable

of distinguishing collective anomalies from point anomalies. This and several other methods have been implemented in an R package by Fisch et al. [23], where we also find the multivariate collective and point anomaly (MVCAPA) method of Fisch et al. [24], the proportion adaptive segment selection (PASS) method of Jeng et al. [25], the Bayesian abnormal region detector (BARD) of Bardwell and Fearnhead [26], and also sequential versions of CAPA and MVCAPA by Fisch et al. [27]. Fisch [15] provides the state of the art on statistical anomaly detection, together with a guide for computational implementation.

The present chapter is devoted to another anomaly-detection method that works irrespective of whether systems are being affected at the input or output stage, or at both stages simultaneously. The important feature that distinguishes our method from the earlier ones is that it can detect persistent anomalies that may not change the regime (e.g., mean, variance, and/or autocorrelations) of data in an abrupt fashion during the period of observation. Hence, those statistical techniques that have been designed to detect outliers and other aberrations become ineffective in such situations.

As in many previous studies, we also consider dependent random inputs and, in turn, dependent outputs. This enables us to use the method in a myriad of applications. We have carefully proven the underlying theoretical results and illustrated the method using stationary time series under various contamination by anomalies scenarios. It is useful to recall at this point that historical data as well as subject-matter knowledge are helpful in deciding how to reduce non-stationary random sequences to stationary ones, and transformations such as differencing and de-periodization can especially be helpful [see, e.g., 28, 29]. We shall rely on such transformations in our real-world illustrative examples in Section 4.3.

The departure from the earlier explored by Gribkova and Zitikis [30] case of independent and identically distributed (iid) inputs to the herein tackled dependent random inputs and thus outputs requires considerable technical innovation and have given rise to notions such as p -reasonable order and temperate dependence, whose connections to classical notions such as phantom distributions have been illuminated. We note that the just mentioned parameter p is related the p -th finite moment of inputs, and thus to the tail heaviness of the input distribution.

The rest of this chapter is organized as follows. In Section 4.2 we introduce and discuss basic notation. In Section 4.3 we analyze two actual examples that illustrate the anomaly-detection method what we develop in subsequent sections. In Section 4.4 we introduce an experiment that further illustrates and guides our technical considerations. In Section 4.5 we lay out a foundation for our anomaly-detection method. In Section 4.6 we illustrate the performance of the method graphically. In Section 4.7 we explain how the method acts in anomaly-free orderly systems, whereas in Section 4.8 we show how the method detects anomalies when they are present. Section 4.9 concludes this chapter with a brief summary of main results and several suggestions for future studies. Although some graphical illustrations are already given in the main body of this chapter, Appendix 4.A contains more extensive illustrations. Technical details such as lemmas and proofs are in Appendix 4.B.

4.2 Setting the stage: basic notation

Throughout this chapter, we assume the existence of a function $h : \mathbb{R}^{1+d} \rightarrow \mathbb{R}$, called transfer function, that connects inputs $X_t \in \mathbb{R}$ and outputs $Y_t \in \mathbb{R}$ via the equation

$$Y_t = h(X_t, \boldsymbol{\varepsilon}_t), \quad (4.2.1)$$

where $\boldsymbol{\varepsilon}_t \in \mathbb{R}^d$ are d -dimensional exogenous random variables, called anomalies. They can of course be equal to $\mathbf{0} := (0, \dots, 0) \in \mathbb{R}^d$, meaning that the system is free of anomalies. In this case the transfer function reduces to $h_0 : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$h_0(x) = h(x, \mathbf{0}),$$

which we call the baseline function. When we wish to emphasize that outputs Y_t arise from this anomaly-free case, we use the notation

$$Y_t^0 = h_0(X_t). \quad (4.2.2)$$

To illustrate, let $d = 2$, in which case we have $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t})$. We may think of $\varepsilon_{1,t}$ as anomalies affecting the inputs X_t before they enter the control system, and $\varepsilon_{2,t}$ as anomalies affecting the (already affected) inputs when they exit the system. Thinking in this fashion, we arrive at the following transfer functions $h : \mathbb{R}^3 \rightarrow \mathbb{R}$, which we use in our numerical experiment later in this chapter:

TF1: $h(x, y, 0) := h_0(x + y)$ when the system is affected by anomalies only at the input stage;

TF2: $h(x, 0, z) := h_0(x) + z$ when the system is affected by anomalies only at the output stage;

TF3: $h(x, y, z) := h_0(x + y) + z$ when the system is affected by anomalies at the input and output stages.

Besides the additive model, there are other models and thus other transfer functions that link inputs with exogenous variables [e.g., 31–34]. Arguments in favour of using one model over another can be found in studies by, e.g., Perote et al. [35], Su [36], Semenikhine et al. [37], Guo et al. [34], and Guo et al. [38].

Model (4.2.1) arises in many areas, including regression analysis, classification, and, generally, in machine learning [e.g., 39]. It also relates our research to the so-called strategy-proof estimation in regression [e.g., 40, 35].

Note 4.2.1 *Visually, model (4.2.1) may give the impression that the outputs depend only on the current value of inputs, but X_t , at least in the case of causal time series, is a linear combination of the contemporary and historical values of the underlying white noise. That is, X_t is the inner product $\langle \boldsymbol{\beta}, \mathbf{Z}_t \rangle$ of a (finite or infinite) sequence $\boldsymbol{\beta} = (\beta_i)_{i \geq 0}$ of parameters and a (finite or infinite) sequence $\mathbf{Z}_t = (Z_{t-i})_{i \geq 0}$ of uncorrelated random variables. We shall elaborate more on this topic in Section 4.9, in the context of potential future work.*

Although the function h might be known to, e.g., the control system's manufacturer, its precise formula may not be known to those working in the area of anomaly detection (e.g., company's IT personnel). The transfer function might even deviate from its original specifications due to, e.g., wear and tear. Furthermore, in the context of, say, economic variables, which we shall encounter in the next section, their relationships might be postulated by academics but in actuality, the true relationships (i.e., the transfer mechanisms from one to another) usually deviate from any model. In addition, the relationships might be, and usually are, affected by exogenous economic and other variables. Hence, to accommodate various scenarios associated with model uncertainty, we shall aim at deriving results for very large classes of transfer functions, that is, under very mild assumptions.

4.3 Two actual illustrations

This section is devoted to two real-world illustrations, which make up the contents of the following two subsections. The illustrations are based on pairs (X_t, Y_t) of economic variables, which are observed only for $1 \leq t \leq n$ for some sample sizes n . The inputs and outputs are dependent. With some luck, one of these variables can be assessed from the values of another variable, although not precisely because the transfer mechanism (i.e., the transfer function h) is not known, except of course in academic models. This, however, is not of concern to us because our primary interest is in finding out whether exogenous economic or other variables are systematically affecting the relationship between X_t and Y_t . That is, we want to answer the following question:

Question 4.3.1 *Are there ε_t 's in model (4.2.1)?*

At this point, we may instinctively start to debate as to the extent of smoothing of the scatterplot $(X_1, Y_1), \dots, (X_n, Y_n)$, assuming that we want to do it: extreme undersmoothing would result in a wiggly function h with no errors ε_t , whereas too much smoothing would result in a nice function h but with large errors ε_t . Hence, the researcher's subjectively chosen level of smoothing determines whether or not there are errors in the model, and how large they are. We therefore do not do any smoothing. Our task is to find out if the hypothetical transfer function from one economic variable to another is affected by exogenous systematic variables, whatever they might be.

4.3.1 Dow Jones and Australian All Ordinaries Indices

The data [41, Example 8.1.1] consist of the closing values of the Dow Jones Index (DJ) and the Australian All Ordinaries Index of Share Prices (AO) recorded at the termination of trading on 251 successive trading days up to August 26, 1994. From the original data, we calculate the percentage relative price changes, known as percentage returns, and plot them in Figure 4.1.

To answer Question 4.3.1, we employ an index I_n , whose mathematical definition will be introduced in Section 4.5. At the moment, what really matters and interests us are the conclusions that we can reach, for which we use the following decision rules:

Decision 1: If the transition from inputs to outputs is accomplished without systematic interference, then, when the sample size n grows, the index I_n stays away from $1/2$.

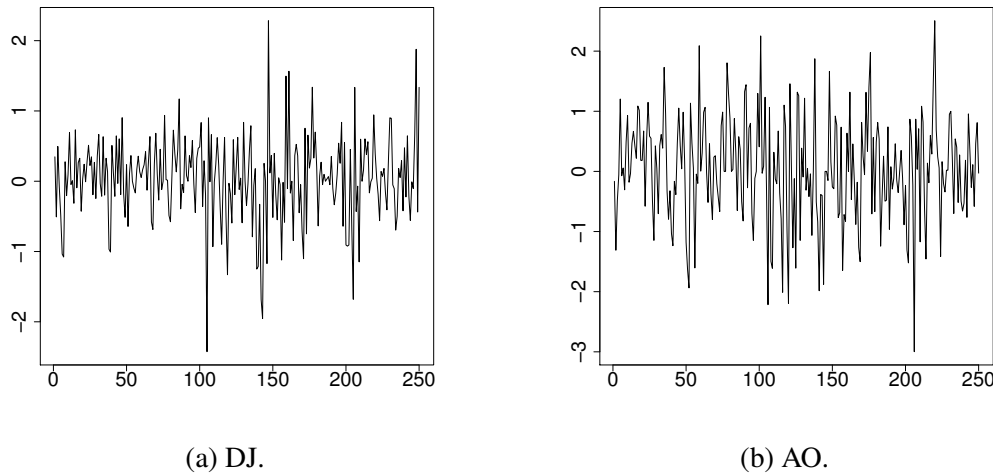


Figure 4.1: The percentage returns of DJ and AO.

Decision 2: If, however, the transition is exposed to systematic interference, then, when the sample size n grows, the index I_n tends to $1/2$.

Due to limited sample sizes n or some other reasons, it may not always be clear whether or not the index I_n tends to $1/2$. In such cases we additionally calculate another index, denoted by $B_{n,2}$, whose mathematical definition will be given in Section 4.5. The meaning of the index $B_{n,2}$ relies on its growth to infinity, and it supplements Decisions 1 and 2 in the following way:

Supplement 1: If the transition from inputs to outputs is accomplished without systematic interference, then, when the sample size n grows, the index $B_{n,2}$ stays asymptotically bounded, that is, $B_{n,2} = O_{\mathbb{P}}(1)$ in mathematical terms.

Supplement 2: If, however, the transition is exposed to systematic interference, then, when the sample size n grows, the index $B_{n,2}$ tends to infinity.

Equipped with these indices I_n and $B_{n,2}$, we can now look at the closing values of DJ and AO. The first question that arises is which of the two variables should be the “input.” The answer is naturally related to causality, but to avoid any prejudicial statement and thus controversy, we do our analysis both ways: first we take DJ as the input and thus AO as the output, and then interchange their roles. The two cases with their respective indices I_n and $B_{n,2}$ are visualized in Figure 4.2.

The graphs suggest that there is exogenous interference when transferring DJ to AO, and also the other way around, although there is a little dip below $1/2$ on the right-hand side of Figure 4.2c, which may not be of importance given its small value. The index $B_{n,2}$ sends the same message as I_n . Hence, we comfortably conclude the existence of interference, although more data might overturn the conclusion.

Note that in Figure 4.2 we always start graphing the panels at $n = 20$. This is so because for small values of n , the index I_n fluctuates wildly between 0 and 1, as it should, which will be clearly seen from the mathematical definition of I_n . Hence, by starting at $n = 20$, we are able to better depict the behaviour of I_n near $1/2$, which is what really matters for our anomaly-detection method.

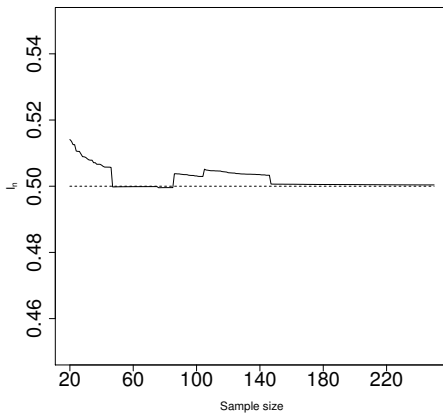
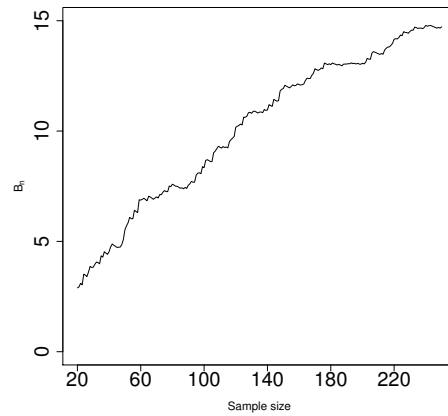
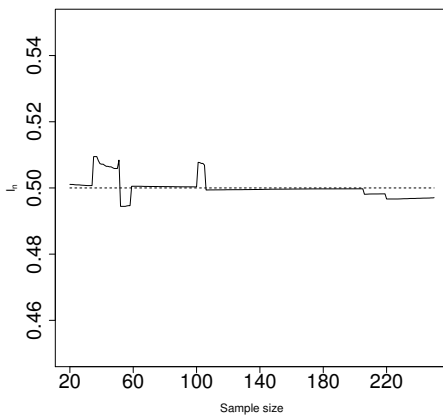
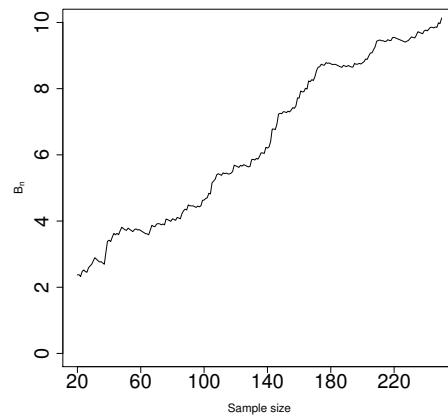
(a) I_n when $(X, Y) = (DJ, AO)$.(b) $B_{n,2}$ when $(X, Y) = (DJ, AO)$.(c) I_n when $(X, Y) = (AO, DJ)$.(d) $B_{n,2}$ when $(X, Y) = (AO, DJ)$.

Figure 4.2: The indices I_n and $B_{n,2}$ corresponding to DJ and AO with respect to the sample sizes $n = 20, \dots, 149$.

4.3.2 Sales with a leading indicator

The data [41, Example 8.1.2] consist of 150-day sales with a leading indicator, plotted in Figure 4.3. The data are non-stationary, and so we difference it at lag 1. The transformed data are plotted in Figure 4.4. We set the differenced leading indicator as the input and the differenced sales as the output. There are two reasons for this choice: first, it makes economic sense, and second, the differenced leading indicator exhibits stationarity whereas differenced sales seem to hint at some periodicity. Having thus made these choices, we next calculate the indices I_n and $B_{n,2}$, which are depicted in Figure 4.5. The index I_n does not tend to $1/2$ and the index $B_{n,2}$ stops rising at about $n = 120$. These observations suggest the lack of exogenous interference when transiting from the inputs to the outputs, that is, from the leading variable to the sales.

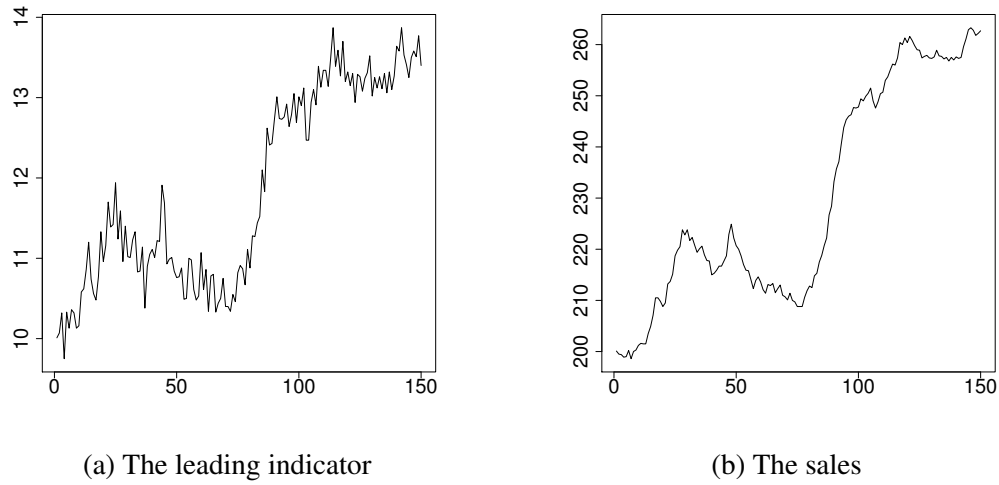


Figure 4.3: The original 150-day data of the leading indicator and the sales.

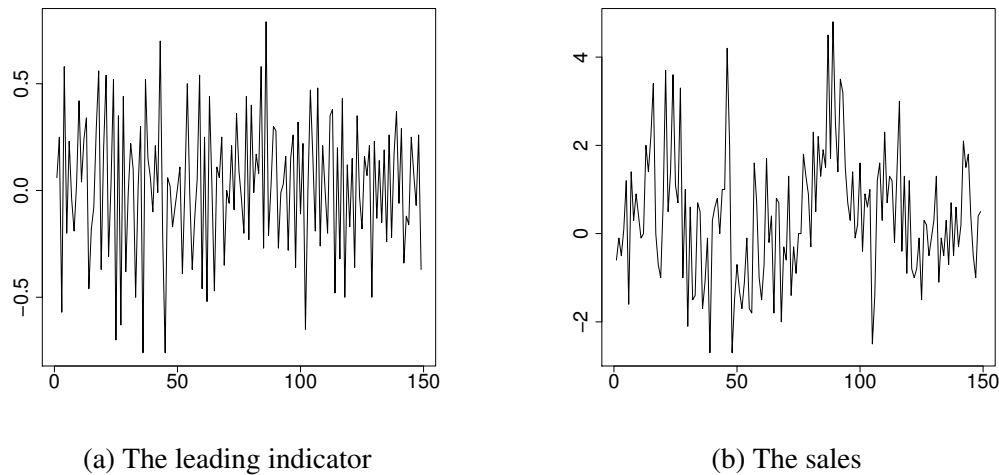


Figure 4.4: The 1-lag differences of the leading indicator and the sales.

4.4 Introducing a controlled experiment

To explore how the anomaly-detection method works, we have designed an experiment based on a simple (from the statistical modeling perspective) control system, which is the automatic voltage regulator (AVR) that has been an active research area with a considerable number of innovative designs and algorithms proposed in the literature. For details, we refer to the recent contributions by, e.g., Çelik and Durgut [42], Gozde [43], and extensive references therein.

In its simplest form, the AVR intakes voltages X_t and outputs more stable voltages Y_t within a pre-specified service range $[a, b]$. When it is known that the system is free of anomalies, the

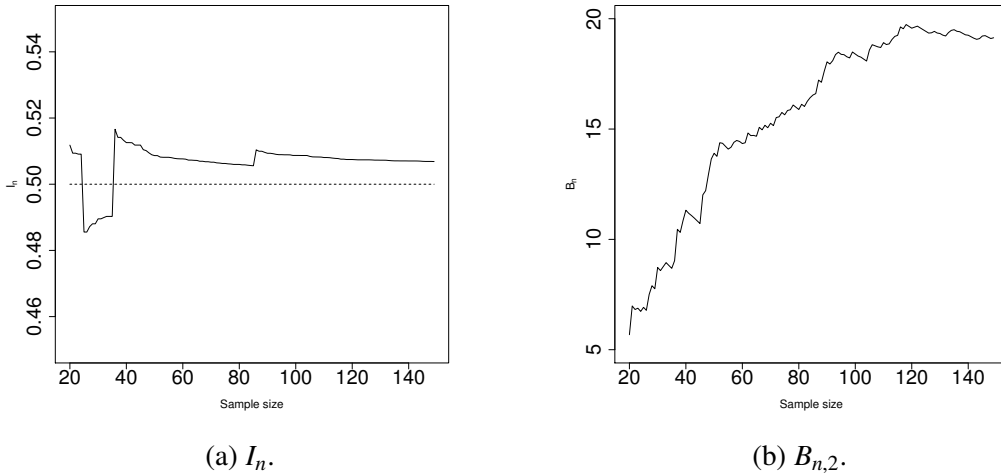


Figure 4.5: The indices I_n and $B_{n,2}$ of the 150-day sales data with respect to $n = 20, \dots, 149$.

outputs are

$$Y_t^0 = (X_t \wedge b - a)_+ + a = \begin{cases} a & \text{when } X_t < a, \\ X_t & \text{when } a \leq X_t \leq b, \\ b & \text{when } X_t > b. \end{cases}$$

Hence, using the “clamped” baseline function (see Figure 4.6)

$$h_c(x) = (x \wedge b - a)_+ + a, \tag{4.4.1}$$

the outputs are $Y_t^0 = h_c(X_t)$. Note that the clamped function h_c is Lipschitz continuous but not

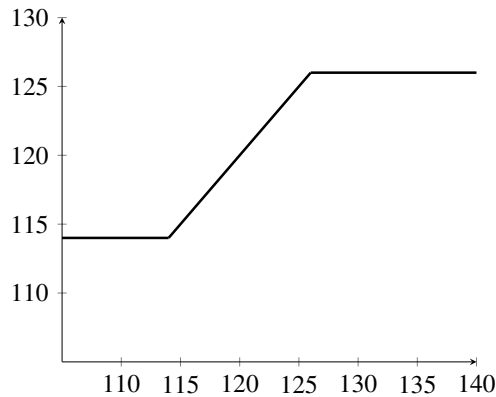


Figure 4.6: The transfer function h_c with $a = 114$ and $b = 126$ corresponding to the automatic voltage regulator with the transfer window 120 ± 6 volts (i.e., $\pm 5\%$).

continuously differentiable, and we shall keep this example in mind when deriving results in Sections 4.7 and 4.8 so that to avoid making assumptions that would exclude functions such as h_c .

Note 4.4.1 *Transfer functions similar to the one in Figure 4.6 appear naturally in various reinsurance treaties, where direct insurers and reinsurers calculate their risk transfers using formulas resembling (4.4.1) with pre-determined deductibles and policy limits as parameters, very much like a and b in equation (4.4.1). Determining whether or not anomalies (e.g., processing errors) are affecting such transfers is of interest to all parties involved.*

In the numerical experiment in Section 4.6, we shall use the clamped function h_c as the baseline function h_0 , and then use the transfer functions TF1–TF3 (Section 4.2) as h . The anomaly-free inputs X_t in the experiment are assumed to follow the ARMA(1, 1) time series, with the input anomalies δ_t and the output anomalies ϵ_t being independent within and among them, and coming from a certain parametric distribution. Note that such anomalies can be interpreted as genuinely unintentional; they may arise from, e.g., systematic measurement-errors due to faulty equipment. In Section 4.7 we shall develop results for the anomaly-free outputs $Y_t^0 = h_0(X_t)$. In Section 4.8 we shall do the same for the anomaly affected case, that is, when $\epsilon_t = (\delta_t, \epsilon_t) \in \mathbb{R}^2$ and thus $Y_t = h(X_t, \delta_t, \epsilon_t)$.

Note 4.4.2 *When the inputs X_t are iid random variables, which is a very special case of the present chapter, anomaly detection in systems with $\delta_t = 0$ has been studied by Gribkova and Zitikis [30], with $\epsilon_t = 0$ by Gribkova and Zitikis [44], and with arbitrary anomalies (δ_t, ϵ_t) by Gribkova and Zitikis [45]. In the present chapter we extend those iid-based results to scenarios when inputs are governed by stationary time-series models, which is a highly important feature from the practical point of view. To achieve these goals, a considerable technical work has to be done, which we present in Appendix 4.B.*

We are now ready to familiarize with the anomaly-detection method, and in particular with mathematical definitions of (dis)orderly systems and of the indices I_n and $B_{n,2}$, as well as of the more general index $B_{n,p}$.

4.5 Anomaly detection: a foundation

Let X_1, \dots, X_n be the observable part of a stationary sequence of inputs X_t , whose marginal cumulative distribution functions (cdf's) are the same; we denote them by F . With these observable inputs, there are associated outputs Y_1, \dots, Y_n , and so we are dealing with the random input-output pairs $(X_1, Y_1), \dots, (X_n, Y_n)$. Based on them, we wish to determine whether the system transferring the inputs into the outputs is functioning as intended or is systematically affected by anomalies. To successfully tackle this problem, we first need to rigorously define (dis)orderly systems.

Let the cdf F be continuous, which allows us without loss of generality to state that all the inputs X_1, \dots, X_n are different. Hence, their order statistics

$$X_{1:n} < X_{2:n} < \dots < X_{n:n}$$

are strictly increasing. This facilitates unambiguous definition of the concomitants of the outputs Y_1, \dots, Y_n , which are denoted by $Y_{1:n}, \dots, Y_{n:n}$ and defined by the equation

$$Y_{t:n} = \sum_{s=1}^n Y_s \mathbb{1}\{X_s = X_{t:n}\},$$

where $\mathbb{1}$ is the indicator: it is equal to 1 when the condition $X_s = X_{t:n}$ is satisfied and 0 otherwise. We are now in the position to define (dis)orderly systems.

Definition We say that the outputs and thus the system are *in p -reasonable order* with respect to the inputs for some $p > 0$ if

$$B_{n,p} := \frac{1}{n^{1/p}} \sum_{t=2}^n |Y_{t,n} - Y_{t-1,n}| = O_{\mathbb{P}}(1)$$

when $n \rightarrow \infty$. If, however, $B_{n,p} \rightarrow_{\mathbb{P}} \infty$, then we say that the outputs and thus the system are *out of p -reasonable order* with respect to the inputs.

Although this definition is a technicality that is necessary for our anomaly-detection method, it is also natural from the practical point of view. Indeed, detection of anomalies in disorderly systems can hardly be a task worth pursuing. As to the parameter p , its role in Definition 4.5 is to control tail heaviness of the outputs, and we shall later see that this is achieved by controlling tail heaviness of the inputs. Roughly speaking, we can view p as the order of finite moments. Note that when the outputs are in p -reasonable order, the outputs are in r -reasonable order for all $r \leq p$. On the other hand, if the outputs are out of p -reasonable order, the outputs are out of r -reasonable order for all $r \geq p$. Hence, we can say that for any given system, there is a threshold p delineating the sets of in-order and out-of-order outputs.

To successfully detect anomalies affecting a system, we of course need to know that the brand new system was in orderly state. For a rigorous definition of the latter notion, we slightly adjust Definition 4.5 as follows.

Definition The anomaly-free outputs and thus the anomaly-free system are *in p -reasonable order* with respect to the inputs for some $p > 0$ if

$$B_{n,p}^0 := \frac{1}{n^{1/p}} \sum_{t=2}^n |Y_{t,n}^0 - Y_{t-1,n}^0| = O_{\mathbb{P}}(1)$$

when $n \rightarrow \infty$. (For obvious reasons, we do not consider systems that are out of order when they are free of anomalies.)

To illustrate the anomaly-free case, that is, when all ε_t 's are equal to $\mathbf{0}$, if all the inputs happen to be equal to the same constant, say c , then the system is in p -reasonable order for every $p > 0$, because $B_{n,p}^0 = 0$. More generally, next Theorem 4.5.1 will show that if the transfer function h is sufficiently smooth (e.g., Lipschitz continuous), then the system is in p -reasonable order for some $p > 0$ even when the inputs are random, although not too heavy tailed. We need to introduce additional notation before we can formulate the theorem.

Let a_X and b_X be the endpoints of the support of the cdf F , that is,

$$\begin{aligned} a_X &= \sup\{x \in \mathbb{R} : F(x) = 0\}, \\ b_X &= \inf\{x \in \mathbb{R} : F(x) = 1\}. \end{aligned}$$

These endpoints can of course be infinite, but they never coincide because the cdf F is assumed to be continuous. Therefore, the open interval (a_X, b_X) is never empty.

Next, we recall that h_0 is called absolutely continuous if there is a function h_0^* , called the Radon-Nikodym derivative of h_0 , that satisfies the equation

$$h_0(v) - h_0(u) = \int_u^v h_0^*(x) dx$$

for all $u \leq v$. Now we are ready to formulate the theorem that describes the circumstances under which the anomaly-free system is orderly.

Theorem 4.5.1 *The anomaly-free outputs are in p -reasonable order with respect to the inputs for some $p \geq 1$ if there is $\alpha \in [1, p]$ such that $\mathbb{E}(|X_1|^{p/\alpha}) < \infty$ and one of the following conditions holds:*

- (i) *If $\alpha = 1$, then the baseline function h_0 is Lipschitz continuous, that is, there is a constant $K \geq 0$ such that, for all $x, y \in [a_X, b_X]$,*

$$|h_0(x) - h_0(y)| \leq K|x - y|.$$

- (ii) *If $\alpha > 1$, then the baseline function h_0 is absolutely continuous on the interval $[a_X, b_X]$ and its Radon-Nikodym derivative h_0^* satisfies*

$$\int_{-\infty}^{\infty} |h_0^*(x)|^{\alpha/(\alpha-1)} dx < \infty.$$

Next are two facts (to be proven later) upon which we base our anomaly-detection method:

Fact 1: If the anomaly-free outputs are in p -reasonable order with respect to the inputs, then, under some fairly weak assumptions on the inputs and the transfer function h (details in Section 4.7), the index

$$I_n^0 := \frac{\sum_{i=2}^n (Y_{i,n}^0 - Y_{i-1,n}^0)_+}{\sum_{i=2}^n |Y_{i,n}^0 - Y_{i-1,n}^0|} \quad (4.5.1)$$

converges, when $n \rightarrow \infty$, to a limit other than $1/2$.

Fact 2: If, due to anomalies, the outputs are out of p -reasonable order with respect to the inputs, then (details in Section 4.8) the index

$$I_n := \frac{\sum_{i=2}^n (Y_{i,n} - Y_{i-1,n})_+}{\sum_{i=2}^n |Y_{i,n} - Y_{i-1,n}|} \quad (4.5.2)$$

converges to $1/2$ when $n \rightarrow \infty$.

Establishing these two facts rigorously is a complex and lengthy exercise, which we do in Sections 4.7 and 4.8, as well as in Appendix 4.B. To show that the task is worth the effort, in the next section we show how the anomaly-detection method actually works in the case of the AVR-based experiment that we introduced in Section 4.4.

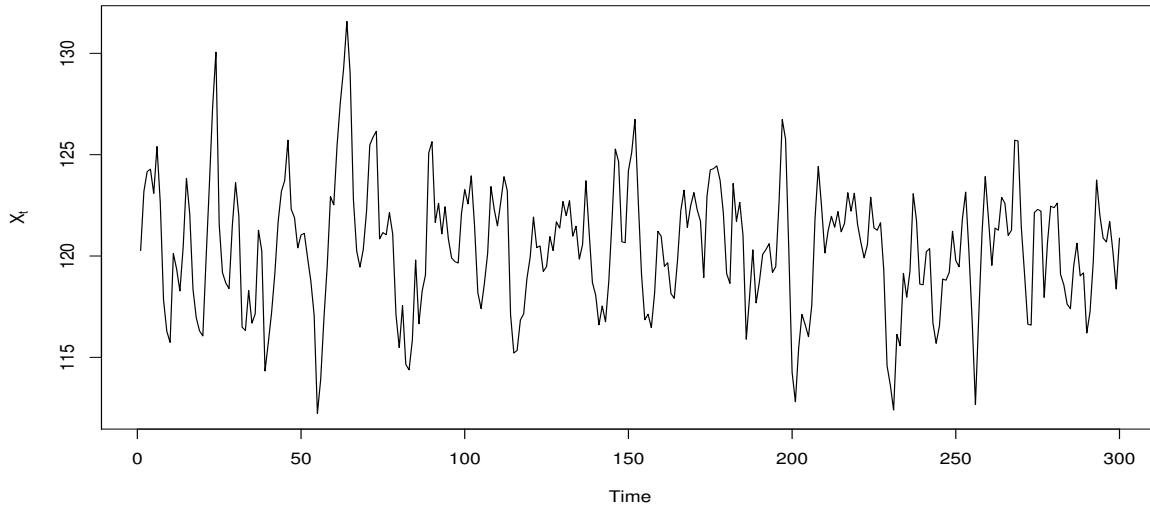


Figure 4.7: ARMA(1, 1) inputs $(X_t)_{t \in \mathbb{Z}}$ as specified by model (4.6.1).

4.6 The experiment: parameter choices and results

To illustrate the anomaly-detection method, and in particular Facts 1 and 2 formulated in the previous section, we use the AVR-based experiment with the following parameter choices.

First, the anomaly-free inputs X_t (see Figure 4.7) follow the ARMA(1, 1) time series model

$$(X_t - 120) = 0.6(X_{t-1} - 120) + \eta_t + 0.4\eta_{t-1}, \quad (4.6.1)$$

where the white noise sequence η_t consists of iid, mean zero, normal $\mathcal{N}(0, \sigma_\eta^2)$ random variables with the variances

$$\sigma_\eta^2 = \frac{3^2(1 - 0.6^2)}{1 + 2(0.6)(0.4) + 0.4^2} = \frac{5.76}{1.64} \approx 3.512195.$$

Under these specifications [e.g., 28, Eq. (3.4.7), p. 79] the input time series X_t has the marginal normal distribution with mean 120 and variance 9, that is,

$$X_t \sim \mathcal{N}(120, 9)$$

for every $t \in \mathbb{Z}$.

Next, the input anomalies δ_t are iid Lomax($\alpha, 1$) with shape parameter $\alpha > 0$ [e.g., 46, Section 2.3.11, pp. 23–24], and the output anomalies ϵ_t are also iid Lomax($\alpha, 1$). Both the input and output anomalies are independent of each other, and they are also independent of the inputs X_t . (Such anomalies can be interpreted as genuinely unintentional.) Hence, the anomalies are independent, non-negative, random variables with the means

$$\mathbb{E}(\delta_t) = \mathbb{E}(\epsilon_t) = \frac{1}{\alpha - 1}$$

and the variances

$$\text{Var}(\delta_t) = \text{Var}(\epsilon_t) = \begin{cases} \frac{\alpha}{(\alpha-1)^2(\alpha-2)} & \text{when } \alpha > 2, \\ \infty & \text{when } 1 < \alpha \leq 2. \end{cases}$$

We set the following values for the shape parameter α :

- $\alpha = 11$, which gives $\mathbb{E}(\delta_t) = \mathbb{E}(\epsilon_t) = 0.1$ and $\text{Var}(\delta_t) = \text{Var}(\epsilon_t) = 0.0122$, thus making, in average, the anomalies look small if compared to the nominal voltage 120;
- $\alpha = 1.2$, which gives $\mathbb{E}(\delta_t) = \mathbb{E}(\epsilon_t) = 5$ and infinite variances, thus making, in average, the anomalies look moderate in size if compared to the nominal voltage 120.

We shall see that in both cases the method detects the anomalies with remarkable easiness, although the required sample size when $\alpha = 11$ needs to be, naturally, larger than when $\alpha = 1.2$ in order to reach the same conclusion. A few clarifying notes follow.

Note 4.6.1 *The terms “small” and “moderate” that we used to describe anomalies with average values 0.1 and 5, respectively, are our terms and may not coincide with what the reader might think about such anomalies. Nevertheless, it seems to us that the terms “small” and “moderate” correlate well with the accepted notions of “strict” and “satisfactory” AVR service ranges, which are 120 ± 3 and 120 ± 6 , respectively.*

Note 4.6.2 *Among the two choices of α made above, one leads to a finite variance and another to infinite. These two distinct scenarios are of practical interest. Indeed, based on empirical evidence, there has been a considerable discussion in the literature as to what distribution tails (and related dependence structures) could be suitable for modelling, e.g., data traffic and cyber risks. For details and further references on the topic, we refer to, e.g., Heath et al. [47], Maillart and Sornette [48], Edwards et al. [49].*

To proceed with the set-up of our AVR-based experiment, we next introduce three service ranges, among which “satisfactory” and “strict” are commonly used terms in practice, and “precise” is an artefact.

Satisfactory:

$$[a, b] = [114, 126],$$

which is 120 ± 6 (i.e., $\pm 5\%$) and is considered a standard supply range in, e.g., Canada (recall Figure 4.6), with 120 being the nominal voltage.

Strict:

$$[a, b] = [117, 123],$$

which is 120 ± 3 (i.e., $\pm 2.5\%$).

Precise:

$$[a, b] = [120, 120] = \{120\}.$$

The anomaly-detection method in the case of the clamped transfer function h_c corresponding to these three service ranges is illustrated in the next three subsections, with a more extensive set of illustrative graphs provided in Appendix 4.A. For space considerations, we only consider the case $p = 2$. We start with the precise service range.

4.6.1 Precise service range

When the AVR service range is precise, which is an artefact created only for illustrative purposes, the clamped function is constant, that is,

$$h_c(v) = 120.$$

Since $h_0 = h_c$ in this experiment, all the three transfer functions TF1–TF3 in the anomaly-free case yield

$$B_{n,2}^0 = 0,$$

thus implying that the system is orderly. The index I_n^0 is undefined, as it is the ratio $0/0$. These notes also apply to the anomaly-affected case $h(x, y, 0)$, as it is equal to $h_c(x + y)$, which is 120, a constant. In the remaining two anomaly-affected cases $h(x, 0, z)$ and $h(x, y, z)$, which are identical and given by the equations

$$h(x, 0, z) = 120 + z = h(x, y, z),$$

the output anomalies affect the system. Figure 4.8 depicts I_n and $B_{n,2}$ for various sample sizes n . Note that the index I_n initially fluctuates but quickly starts to tend to $1/2$, whereas $B_{n,2}$ is increasing with respect to the sample size. These two observations suggest that anomalies are affecting the system, which is indeed the case given the experimental design.

4.6.2 Strict service range

When the AVR service range is strict, the clamped function is

$$h_c(v) = (\min\{123, v\} - 117)_+ + 117 = \begin{cases} 117 & \text{when } v < 117, \\ v & \text{when } 117 \leq v \leq 123, \\ 123 & \text{when } v > 123. \end{cases}$$

Figure 4.9 depicts the anomaly-free indices I_n^0 and $B_{n,2}^0$ for various sample sizes n . Looking at the graphs, we safely infer that $B_{n,p}^0 = O_{\mathbb{P}}(1)$, which implies that the anomaly-free system is orderly with respect to the inputs, and we also see that I_n^0 does not converge to $1/2$, which confirms that the system is free of anomalies.

When, however, the system is affected by anomalies at the input and/or output stages, the behaviour of I_n and $B_{n,2}$ changes drastically. We see from Figures 4.11 (when $\alpha = 1.2$) and 4.12 (when $\alpha = 11$) that the index I_n tends to $1/2$. Naturally, it tends to $1/2$ faster when $\alpha = 1.2$ than when $\alpha = 11$, simply because the anomalies in the latter case are less noticeable. For both α values, the index $B_{n,2}$ has the tendency to grow. These observations suggest that the system is being affected by anomalies, which is indeed the case.

4.6.3 Satisfactory service range

When the AVR service range is satisfactory, the clamped function is

$$h_c(v) = (\min\{126, v\} - 114)_+ + 114 = \begin{cases} 114 & \text{when } v < 114, \\ v & \text{when } 114 \leq v \leq 126, \\ 126 & \text{when } v > 126. \end{cases}$$

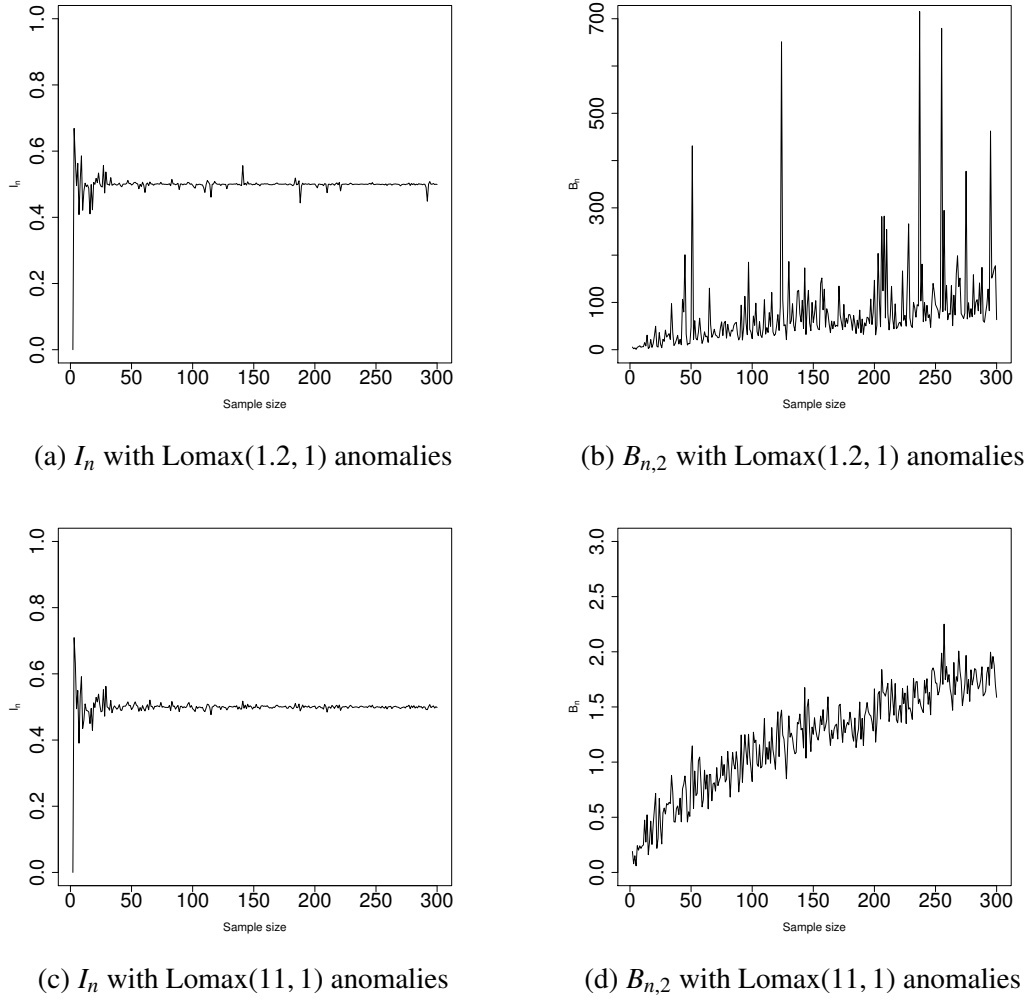


Figure 4.8: The anomaly-affected indices I_n and $B_{n,2}$ for the precise service range with respect to the sample sizes $2 \leq n \leq 300$ for ARMA(1, 1) inputs.

Figure 4.10 resembles Figure 4.9, and Figures 4.13–4.14 convey essentially the same information as Figures 4.11–4.12.

4.7 Anomaly-free systems

In this section we specify conditions under which the index I_n^0 given by equation (4.5.1) converges to a limit other than $1/2$. We note at the outset that since the positive part z_+ of any real number $z \in \mathbb{R}$ can be expressed as $(|z| + z)/2$, the index I_n^0 can be re-written as

$$I_n^0 = \frac{1}{2} \left(1 + \frac{Y_{n,n}^0 - Y_{1,n}^0}{\sum_{i=2}^n |Y_{i,n}^0 - Y_{i-1,n}^0|} \right). \tag{4.7.1}$$

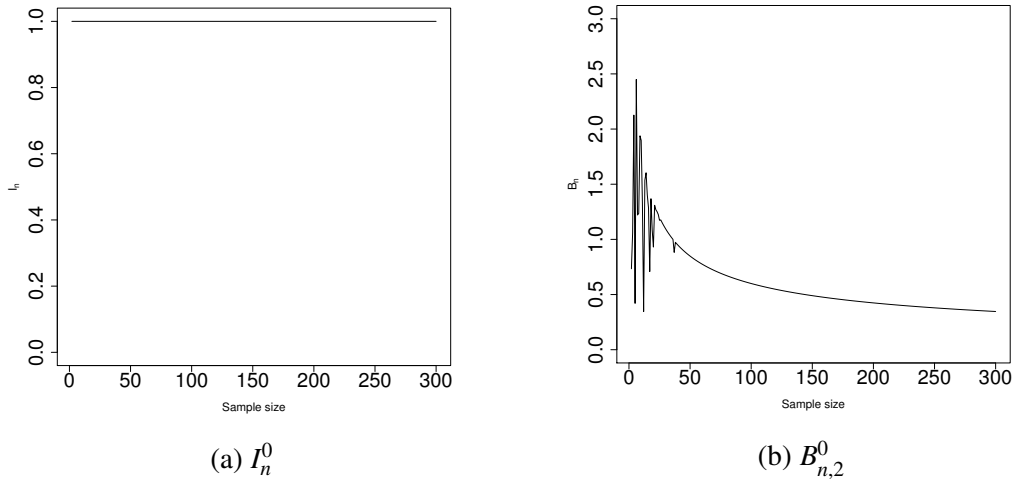


Figure 4.9: The anomaly-free indices I_n^0 and $B_{n,2}^0$ for the strict service range with respect to the sample sizes $2 \leq n \leq 300$ for ARMA(1, 1) inputs.

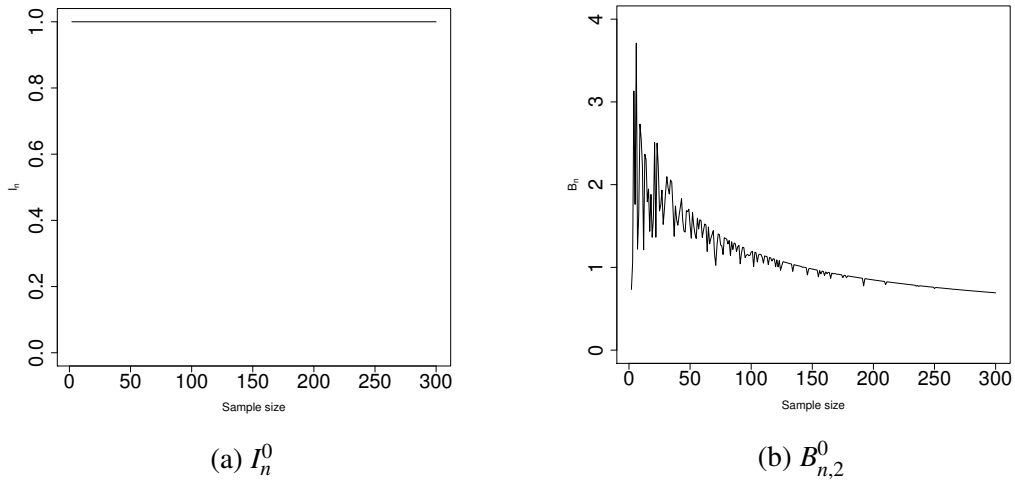


Figure 4.10: The anomaly-free indices I_n^0 and $B_{n,2}^0$ for the satisfactory service range with respect to the sample sizes $2 \leq n \leq 300$ for ARMA(1, 1) inputs.

Hence, our goal becomes to understand when and where the numerator and the denominator of the ratio

$$\Lambda_n := \frac{Y_{n,n}^0 - Y_{1,n}^0}{\sum_{i=2}^n |Y_{i,n}^0 - Y_{i-1,n}^0|} \tag{4.7.2}$$

converge. These are the topics of the following two subsections.

4.7.1 Asymptotics of the Λ_n numerator

We start with a definition.

Definition We say that the inputs X_t having the same continuous marginal cdf F are *temperately dependent* if, for every $x \in (a_X, b_X)$ and when $n \rightarrow \infty$,

$$\mathbb{P}(X_{1:n} \geq x) \rightarrow 0 \quad \text{and} \quad \mathbb{P}(X_{n:n} \leq x) \rightarrow 0. \quad (4.7.3)$$

A few clarifying notes follow. First, the open interval (a_X, b_X) is not empty because the cdf F is continuous. Second, our use of the term *temperately dependent* is natural because property (4.7.3) is simultaneously related to 1-minimally and n -maximally dependent random variables in the terminology used by Gascuel and Caraux [50]. To clarify, consider two extreme cases:

- If X_t is a sequence of iid random variables, then $\mathbb{P}(X_{1:n} \geq x) = (1 - F(x-))^n$ and $\mathbb{P}(X_{n:n} \leq x) = F(x)^n$, where $x \mapsto F(x-)$ is the left-continuous version of F . Since for all $x \in (a_X, b_X)$, both $F(x)$ and $F(x-)$ are in the interval $(0, 1)$, property (4.7.3) holds.
- If X_t is a sequence of super-dependent random variables, that is, if there is a random variable X such that $X_t = X$ for all $t \in \mathbb{Z}$, then $X_{1:n} = X = X_{n:n}$ and so neither of the two probabilities in property (4.7.3) converges to 0.

Finally, we note that the concept of temperate dependence is closely related to the existence of phantom distributions, which originate from the work of O'Brien [51]. For further details, examples, and extensive references on this topic, we refer to Jakubowski [52], Bradley [53], Doukhan et al. [54]. Phantom distributions for non-stationary random sequences have been tackled by Jakubowski [55].

Theorem 4.7.1 *Let the inputs X_t be strictly stationary and temperately dependent. If the baseline function h_0 is absolutely continuous on the interval $[a_X, b_X]$ and its Radon-Nikodym derivative h_0^* is integrable on $[a_X, b_X]$, then*

$$Y_{n,n}^0 - Y_{1,n}^0 \xrightarrow{\mathbb{P}} h_0(b_X) - h_0(a_X) \quad (4.7.4)$$

when $n \rightarrow \infty$.

Recall that a strictly stationary time series X_t is α -mixing (i.e., strongly mixing) if

$$\alpha_X(t) := \sup |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \rightarrow 0$$

when $t \rightarrow \infty$, with the supremum taken over all $A \in \mathcal{F}_{-\infty}^0$ and $B \in \mathcal{F}_t^\infty$, where the two σ -algebras are defined as follows:

$$\mathcal{F}_{-\infty}^0 = \sigma(X_u, u \leq 0) \quad \text{and} \quad \mathcal{F}_k^\infty = \sigma(X_v, v \geq k).$$

We refer to Lin and Lu [56], Bradley [53], Rio [57] for details and references on various notions of mixing. In the context of the present chapter, a particularly important random sequence is the strictly stationary ARMA(p, q) time series X_t . We refer to Mokkadem [58] who has shown, among other things, that a strictly stationary ARMA time series is β -mixing (i.e., completely regular) and thus α -mixing (i.e., strongly mixing).

Theorem 4.7.2 *If the inputs X_t are strictly stationary and α -mixing, then they are temperately dependent.*

The proofs of Theorems 4.7.1 and 4.7.2 are in Appendix 4.B.

4.7.2 Asymptotics of the Λ_n denominator

We start with a definition, which is a weak (i.e., in probability) form of the classical Glivenko-Cantelli theorem.

Definition We say that the inputs X_t having the same marginal cdf F satisfy the Glivenko-Cantelli property if

$$\|F_n - F\| := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{\mathbb{P}} 0, \quad (4.7.5)$$

where F_n is the empirical cdf based on the random variables X_1, \dots, X_n .

The classical Glivenko-Cantelli theorem says that statement (4.7.5) (with convergence in probability replaced by almost surely) holds for iid random sequences. Establishing the Glivenko-Cantelli property for dependent sequences has been a challenging but fruitful task. In particular, results by Cai and Roussas [59, Corollary 2.1, p. 49] and Rio [57, Proposition 7.1, p. 114] tell us that if a strictly stationary sequence X_t is α -mixing and there exists a constant $\nu > 0$ such that

$$\alpha_X(t) = O(t^{-\nu}) \quad (4.7.6)$$

when $t \rightarrow \infty$, then the Glivenko-Cantelli property holds.

Consider now the stationary ARMA(p, q) time series X_t that follows the dynamical model

$$\sum_{i=0}^p \phi_i X_{t-i} = \sum_{j=0}^q \theta_j \eta_{t-j}, \quad t \in \mathbb{Z},$$

with $\phi_0 = 1$ and some parameters $\phi_i, \theta_j \in \mathbb{R}$ such that the absolute values of all the roots of the characteristic polynomial $z \mapsto \sum_{i=0}^p \phi_i z^i$ are (strictly) greater than 1. Hence, the time series is causal.

Note 4.7.3 *There is a clash of notation between the p in ARMA(p, q) and the p in the earlier introduced definition of p -reasonable order. The two p 's are unrelated, and we do not expect them to cause any confusion. We have simply run out of different notation, especially given the deeply rooted traditions in the literature, such as those in time series analysis.*

Mokkadem [58, Theorem 1] has proved that if the white noise sequence η_t is iid with absolutely continuous (with respect to the Lebesgue measure) marginal distributions, then the time series X_t is geometrically completely regular. That is, there exists $\rho \in (0, 1)$ such that

$$\beta_X(t) = O(\rho^t) \quad (4.7.7)$$

when $t \rightarrow \infty$, where $\beta_X(t)$ is the complete regularity coefficient [60] defined by

$$\beta_X(t) = \mathbb{E} \left(\sup | \mathbb{P}(B | \mathcal{F}_{-\infty}^0) - \mathbb{P}(B) | \right),$$

where the supremum is taken over all $B \in \mathcal{F}_t^\infty$. As noted by [58], the bound

$$\alpha_X(t) \leq \beta_X(t)$$

holds, and thus statement (4.7.7) implies (4.7.6) for any $\nu > 0$, which in turn establishes the Glivenko-Cantelli property for the sequence X_t .

We are now in the position to formulate the main result of this subsection concerning the denominator on the right-hand side of equation (4.7.1).

Theorem 4.7.4 *Let the inputs X_t be strictly stationary, temperately dependent, satisfy the Glivenko-Cantelli property, and have finite p -th moments $\mathbb{E}(|X_t|^p) < \infty$ for some $p > 2$. Let the cdf F and its quantile function F^{-1} be continuous. Finally, assume that the baseline function h_0 is absolutely continuous on $[a_X, b_X]$ and such that its Radon-Nikodym derivative h_0^* is continuous on a finite interval $[a, b] \subseteq [a_X, b_X]$ and vanishes outside $[a, b]$. Then*

$$\sum_{t=2}^n |Y_{t,n}^0 - Y_{t-1,n}^0| \rightarrow_{\mathbb{P}} \int_a^b |h_0^*(x)| dx \quad (4.7.8)$$

when $n \rightarrow \infty$.

The restriction of the support of the Radon-Nikodym derivative h_0^* to only a finite interval $[a, b]$ is due to two reasons:

- 1) those real-life examples (automatic voltage regulators, insurance layers, etc.) that initiated our current research are based on finite transfer windows;
- 2) dealing with finite intervals $[a, b]$ considerably simplifies mathematical technicalities, which is an appealing feature, especially because we do not have a solid practical justification that would warrant further technical complexities.

4.7.3 Back to the index I_n^0

The following corollary to Theorems 4.7.1 and 4.7.4 is the main result of entire Section 4.7. Since the conditions of Theorem 4.7.1 make up only a subset of the conditions of Theorem 4.7.4, we thus impose the latter set of conditions when formulating the corollary.

Corollary 4.7.5 *Under the conditions of Theorem 4.7.4, we have*

$$I_n^0 \rightarrow_{\mathbb{P}} I_\infty^0 := \frac{\int_a^b (h_0^*(x))_+ dx}{\int_a^b |h_0^*(x)| dx} \quad (4.7.9)$$

when $n \rightarrow \infty$.

Note the representation (recall equation (4.7.1))

$$I_\infty^0 = \frac{1}{2} \left(1 + \Lambda(h_0) \right), \quad (4.7.10)$$

where

$$\Lambda(h_0) = \frac{h_0(b) - h_0(a)}{\int_a^b |h_0^*(x)| dx}.$$

We shall see in the next section that for anomaly-affected systems, the empirical index I_n converges to 0.5. To distinguish this case from the limit I_∞^0 in the currently discussed anomaly-free case, we need to assume $\Lambda(h_0) \neq 0$, which is tantamount to assuming $h_0(b) \neq h_0(a)$, because the numerator of $\Lambda(h_0)$ is positive. This is natural from the practical point of view as it excludes those transfer functions (which are usually non-decreasing) whose values at the end-points of

the transfer window $[a, b]$ coincide. Moreover, given model uncertainty, we need to ensure that h_0 belongs to a class of functions for which $\Lambda(h_0)$ is sufficiently distant from 0 so that in the presence of statistical uncertainty we could still – with high confidence – be able to see whether the empirical index I_n converges to 0.5 or some other number.

Note 4.7.6 *The ratio on the right-hand side of statement (4.7.9) arises as a normalized distance in a functional space [61], which after a discretization gives rise to the index I_n^0 and thus, in turn, to the index I_n [62]. For a generalization of these indices to multi-argument functions with further applications, we refer to Davydov et al. [63]. For related mathematical considerations, we refer to Polyak [64].*

4.8 Anomaly-affected orderly systems

If the system is out of p -reasonable order, then the index I_n tends to $1/2$, as shown in the next theorem.

Theorem 4.8.1 *Let the outputs Y_t be identically distributed random variables with finite p -th moments $\mathbb{E}(|Y_t|^p) < \infty$ for some $p \geq 1$. If the outputs are out of p -reasonable order with respect to the inputs, then*

$$I_n \xrightarrow{\mathbb{P}} \frac{1}{2}$$

when $n \rightarrow \infty$.

To apply Theorem 4.8.1 for detecting non-degenerate anomalies ε_t , we need to assume that when all ε_t 's are equal to $\mathbf{0}$, then the system is in p -reasonable order. Hence, our task in this section is this: Assuming that the system with anomaly-free outputs $Y_i^0 = h_0(X_i) = h(X_i, \mathbf{0})$ is in p -reasonable order for some $p > 0$, we need to show that the system becomes out of p -reasonable order when the outputs Y_t are equal to $h(X_t, \varepsilon_t)$ with non-degenerate at $\mathbf{0}$ random anomalies ε_t . In other words, assuming $B_{n,p}^0 = O_{\mathbb{P}}(1)$ when $n \rightarrow \infty$, we need to show that $B_{n,p} \rightarrow_{\mathbb{P}} \infty$ when ε_t 's are non-degenerate at $\mathbf{0}$.

To avoid overloading arguments with mathematical complexities, from now on we set $d = 2$ and work with the three transfer functions TF1–TF3 (Section 4.2). Consequently, in the anomaly-free case we have

$$B_{n,p}^0 = \frac{1}{n^{1/p}} \sum_{t=2}^n |h_0(X_{t:n}) - h_0(X_{t-1:n})|. \quad (4.8.1)$$

We refer to Theorem 4.5.1 for a description of those inputs X_t and the baseline function h_0 for which the anomaly-free system is in p -reasonable order. Our next theorem deals with the case when the input anomalies δ_t are absent.

Theorem 4.8.2 *Let $\delta_t = 0$ for all $t \in \mathbb{Z}$, and let the anomaly-free outputs Y_t^0 be in p -reasonable order with respect to the inputs for some $p > 0$. The outputs Y_t are out of p -reasonable order with respect to the inputs if and only if the output anomalies ε_t are out of p -reasonable order with respect to the inputs.*

To illustrate Theorem 4.8.2, consider the case when the output anomalies ϵ_t and the inputs X_t are independent. Assume also that the output anomalies ϵ_t are iid and have finite first moments. (Such anomalies can be interpreted as genuinely unintentional.) In this case, the joint distribution of the concomitants $(\epsilon_{1,n}, \dots, \epsilon_{n,n})$ is the same as the joint distribution of the anomalies $(\epsilon_1, \dots, \epsilon_n)$ themselves. Consequently, the output anomalies ϵ_t are out of p -reasonable order with respect to the inputs if and only if

$$\frac{1}{n^{1/p}} \sum_{t=2}^n |\epsilon_t - \epsilon_{t-1}| \rightarrow_{\mathbb{P}} \infty \quad (4.8.2)$$

when $n \rightarrow \infty$. Statement (4.8.2) holds (see Lemma 4.B.5) whenever the distribution of ϵ_1 is non-degenerate (at any one point). Hence, we have the following corollary to Theorem 4.8.2.

Corollary 4.8.3 *Let $\delta_t = 0$ for all $t \in \mathbb{Z}$, and let the anomaly-free outputs Y_t^0 be in p -reasonable order with respect to the inputs. The outputs Y_t are out of p -reasonable order with respect to the inputs for every $p > 1$ whenever the output anomalies ϵ_t are iid, non-degenerate, and independent of the inputs.*

For a special but important case of Corollary 4.8.3, recall that by Theorem 4.5.1, the anomaly-free outputs Y_t^0 are in p -reasonable order with respect to the inputs when the baseline function h_0 is Lipschitz continuous. We shall encounter the latter assumption in the following two theorems.

First, we tackle the case when the output anomalies ϵ_t are not present.

Theorem 4.8.4 *Let $\epsilon_t = 0$ for all $t \in \mathbb{Z}$, and let the baseline function h_0 be Lipschitz continuous. Furthermore, let the inputs X_t be strictly stationary, α -mixing, and have finite p -th moments $\mathbb{E}(|X_t|^p) < \infty$ for some $p > 1$. Then the outputs Y_t are out of p -reasonable order with respect to the inputs whenever the following conditions hold:*

- (i) *the input anomalies δ_t are iid and independent of the inputs X_t ;*
- (ii) $\mathbb{E}(|h_0(X_1 + \delta_2) - h_0(X_1 + \delta_1)|) > 0$.

A sufficient condition for assumption (ii) can be obtained via Lemma 4.B.5 and the elementary bound

$$\mathbb{E}(|h_0(X_1 + \delta_2) - h_0(X_1 + \delta_1)|) \geq \mathbb{E}(|g_0(\delta_2) - g_0(\delta_1)|),$$

where

$$g_0(y) = \mathbb{E}(h_0(X_1 + y)). \quad (4.8.3)$$

That is, assumption (ii) is satisfied whenever the distribution of $g_0(\delta_1)$ is non-degenerate.

Finally, we tackle the case when the two anomalies δ_t and ϵ_t are non-degenerate.

Theorem 4.8.5 *Let the baseline function h_0 be Lipschitz continuous. Furthermore, let the inputs X_t be strictly stationary, α -mixing, and have finite p -th moments $\mathbb{E}(|X_t|^p) < \infty$ for some $p > 1$. Then the outputs Y_t are out of p -reasonable order with respect to the inputs whenever the following conditions hold:*

- (i) *the input anomalies δ_t are iid, the output anomalies ϵ_t are also iid, they are independent of each other, and are also independent of the inputs X_t ;*
- (ii) $\mathbb{E}\left(|h_0(X_1 + \delta_2) + \epsilon_2 - h_0(X_1 + \delta_1) - \epsilon_1|\right) > 0$.

Assumption (ii) is satisfied when the random variable $g_0(\delta_1) + \epsilon_1$ is non-degenerate, where g_0 is the same function as in equation (4.8.3).

4.9 A summary and potential extensions

In this chapter we have explored a method for detecting systematic anomalies affecting systems when genuine anomaly-free inputs belong to a large class of stationary time series, or can be reduced to such. The anomalies may mimic (from the distributional point of view) the genuine inputs so closely that the contaminated system may not exhibit any visual aberrations, yet they can be detected using the herein proposed method. Supporting probabilistic and statistical results have been rigorously derived, and conditions under which they hold carefully specified. This rigour facilitates confidence when interpreting results and thus when making decisions.

To illustrate how the method works in practice, we have illustrated it using actual time series and also included a numerical experiment under various model and anomaly specifications. The results have shown that the method is robust and is able to detect even tiny systematic anomalies, although, naturally, under longer periods of observation. The method covers light- and heavy-tailed inputs, thus showing its versatility in applications, including those that are associated with data traffic and cyber risks.

Several interesting topics for future study naturally arise from the present chapter, throughout which we have so far concentrated on the model

$$Y_t = h(X_t, \epsilon_t) \quad (4.9.1)$$

with one-dimensional inputs $X_t \in \mathbb{R}$ and outputs $Y_t \in \mathbb{R}$, and d -dimensional anomalies $\epsilon_t \in \mathbb{R}^d$. With the time series structure of inputs, contemporary and historical observations enter into the model via the equation $X_t = \langle \beta, \mathbf{Z}_t \rangle$. This point of view together with naturally occurring multidimensional predictors in regression, classification, and, generally, in machine learning lead us to the model

$$Y_t = h(\mathbf{X}_t, \epsilon_t) \quad (4.9.2)$$

with k -dimensional predictors $\mathbf{X}_t \in \mathbb{R}^k$ for some $k \in \mathbb{N} \cup \{+\infty\}$. More generally, problems associated with anomaly detection in parallel computer systems, electrical grid, and wireless communication architectures such as SISO, SIMO, etc. [e.g., 65, 66] lead us to the model

$$\mathbf{Y}_t = h(\mathbf{X}_t, \epsilon_t) \quad (4.9.3)$$

with q -valued ($q \in \mathbb{N}$) transfer function $h : \mathbb{R}^{k+d} \rightarrow \mathbb{R}^q$ and thus q -dimensional outputs \mathbf{Y}_t .

The transition from the univariate inputs X_t to the multivariate ones \mathbf{X}_t gives rise to serious mathematical challenges, particularly because of the lack of total ordering in multi-dimensional Euclidean spaces. We feel that the coordinate-wise ordering might lead to a useful anomaly-detection method, but at this moment it looks ad hoc, lacking geometric interpretation and thus

intuitive appeal. The optimization problems tackled by Davydov and Zitikis [61], Davydov et al. [63] might give a clue as to what path to take. Alternatively, studies by Koshevoy [67], Mosler [68] on ordering multi-dimensional elements could give rise to an effective solution.

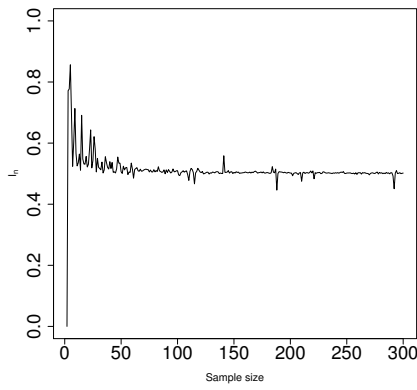
The multivariate nature of outputs \mathbf{Y}_t also creates serious statistical-testing and decision-making problems, but we feel that with some effort, such problems can be tackled with the help of e-values studied by Vovk and Wang [69] and the multiple testing procedures developed by Wang and Ramdas [70]. The e-values are expectation-based versions of the classical p-values. They are simpler to use, thus facilitating multiple hypothesis testing and, in turn, decision making. Hence, the e-values can give rise to impressively powerful and convenient statistical tools in the context of systematic-anomaly detection in, e.g., parallel computer systems, electrical grid, wireless communication architectures, and so on.

Finally, we conclude with the note that anomaly detection problems involve adversarial aspects (e.g., adversarial signal processing, adversarial hypothesis testing) which involve adversaries (intruders) who change their strategies over time. For a glimpse of such research areas, we refer to [71], [72], [73], [74]. Naturally, machine learning techniques play a pivotal role in these areas.

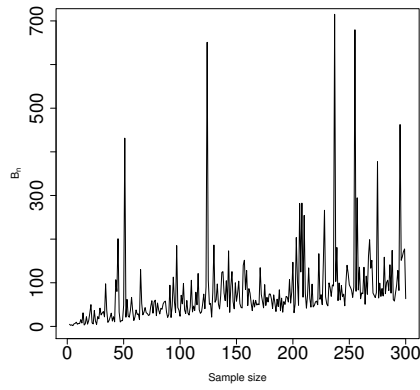
Appendix

4.A Graphical illustrations

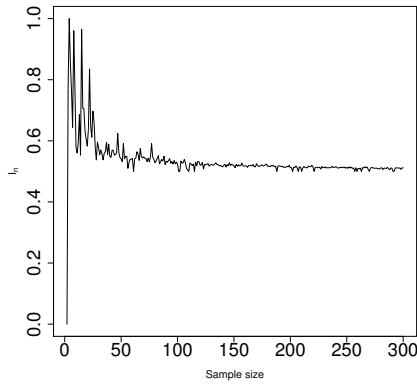
In this appendix we illustrate the behaviour of I_n and $B_{n,2}$ when genuine, anomaly-free inputs follow the ARMA(1, 1) time series and the system is affected by iid Lomax(α , 1) anomalies at the input and/or output stages. In the figures that follow, the system is always affected by anomalies. Hence, the index I_n always tends to 1/2 whereas $B_{n,2}$ grows together with the sample size n . Note also that convergence of I_n to 1/2 is slower when anomaly averages are smaller, meaning that anomalies are less noticeable. This suggests, naturally, that larger sample sizes are needed to reach desired confidence when making decisions.



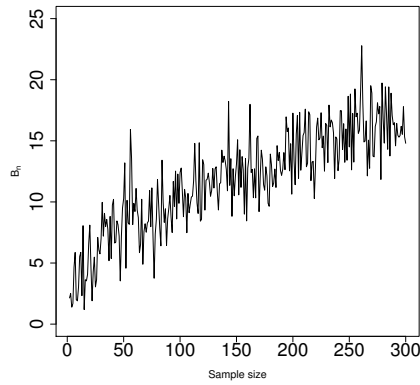
(a) I_n for $h(X_t, 0, \epsilon_t)$.



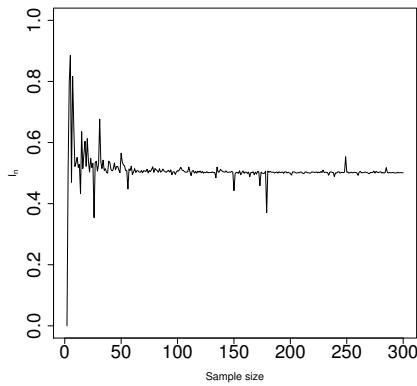
(b) $B_{n,2}$ for $h(X_t, 0, \epsilon_t)$.



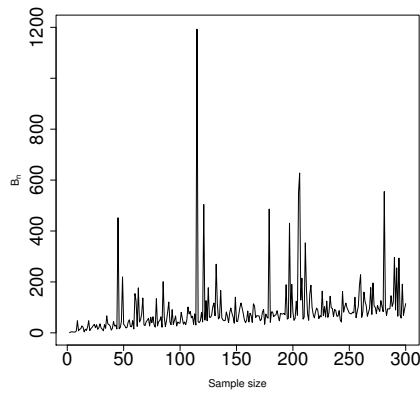
(c) I_n for $h(X_t, \delta_t, 0)$.



(d) $B_{n,2}$ for $h(X_t, \delta_t, 0)$.



(e) I_n for $h(X_t, \delta_t, \epsilon_t)$.



(f) $B_{n,2}$ for $h(X_t, \delta_t, \epsilon_t)$.

Figure 4.11: The anomaly-affected indices I_n and $B_{n,2}$ for the strict service range with respect to $2 \leq n \leq 300$ for ARMA(1, 1) inputs and iid Lomax(1.2, 1) anomalies.

4.B Technical details

To prove Theorem 4.5.1, we need a lemma, which we shall also use when proving other results.

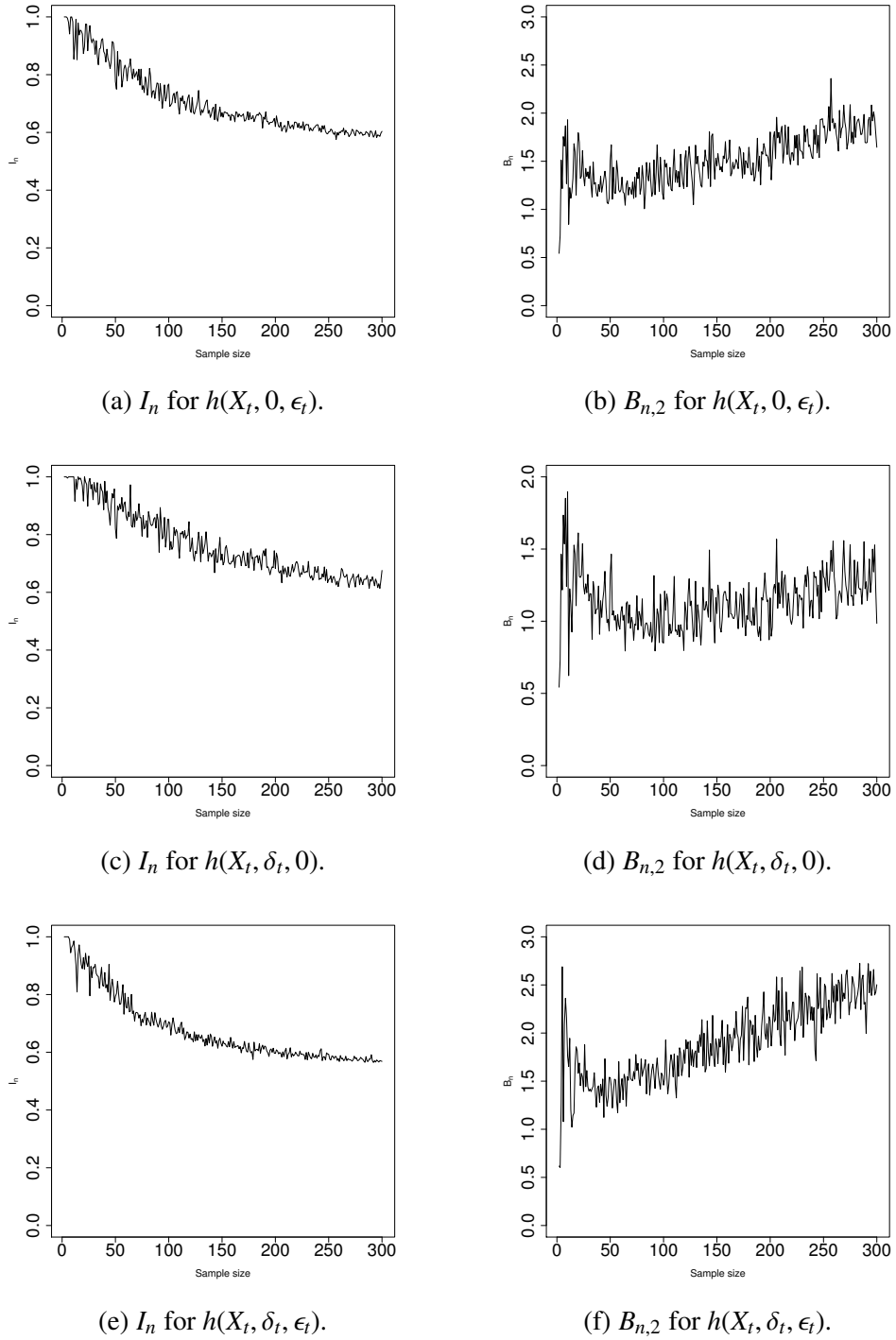
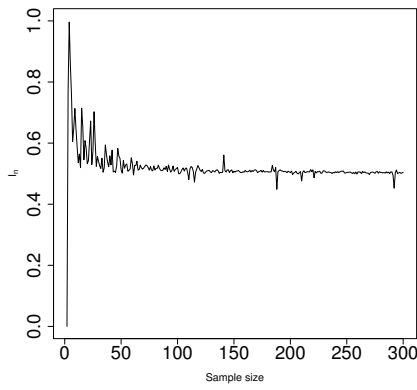


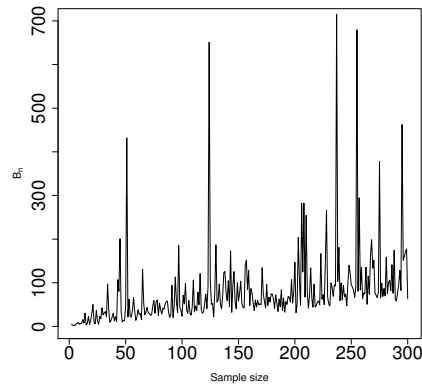
Figure 4.12: The anomaly-affected indices I_n and $B_{n,2}$ for the strict service range with respect to $2 \leq n \leq 300$ for ARMA(1, 1) inputs and iid Lomax(11, 1) anomalies.

Lemma 4.B.1 *Let $\xi_t, t \in \mathbb{Z}$, be identically distributed random variables such that $\mathbb{E}(|\xi_t|^p) < \infty$ for some $p \geq 1$. Then*

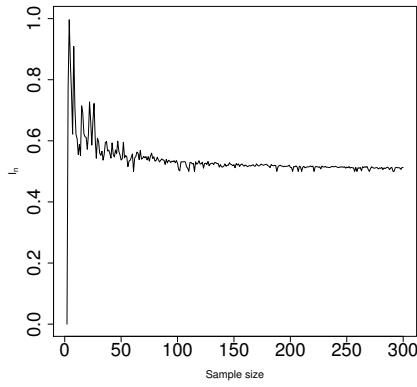
$$n^{-1/p}(\mathbb{E}(\xi_{n:n}) - \mathbb{E}(\xi_{1:n})) = O(1) \tag{4.B.1}$$



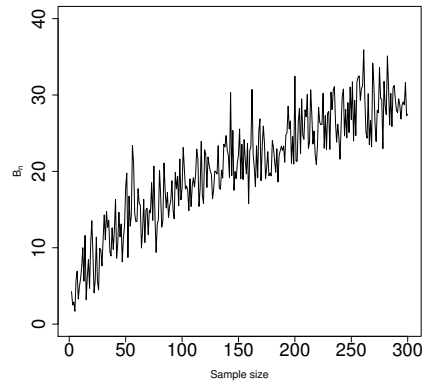
(a) I_n for $h(X_t, 0, \epsilon_t)$.



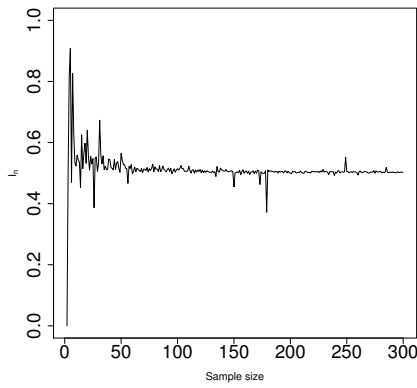
(b) $B_{n,2}$ for $h(X_t, 0, \epsilon_t)$.



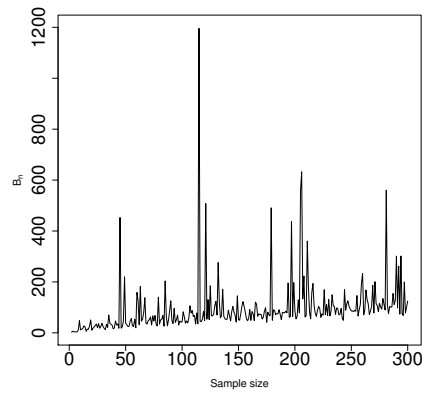
(c) I_n for $h(X_t, \delta_t, 0)$.



(d) $B_{n,2}$ for $h(X_t, \delta_t, 0)$.



(e) I_n for $h(X_t, \delta_t, \epsilon_t)$.



(f) $B_{n,2}$ for $h(X_t, \delta_t, \epsilon_t)$.

Figure 4.13: The anomaly-affected indices I_n and $B_{n,2}$ for the satisfactory service range with respect to $2 \leq n \leq 300$ for ARMA(1, 1) inputs and iid Lomax(1.2, 1) anomalies.

when $n \rightarrow \infty$, and thus $n^{-1/p}(\xi_{n:n} - \xi_{1:n}) = O_{\mathbb{P}}(1)$, where $\xi_{1:n} \leq \xi_{2:n} \leq \dots \leq \xi_{n:n}$ are the order statistics of $\xi_1, \xi_2, \dots, \xi_n$.

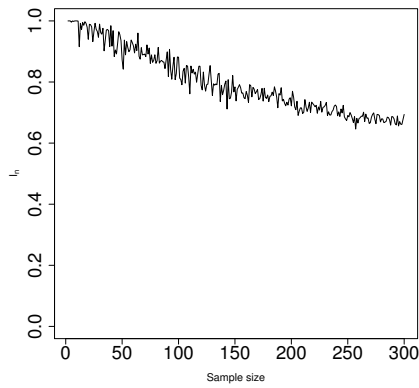
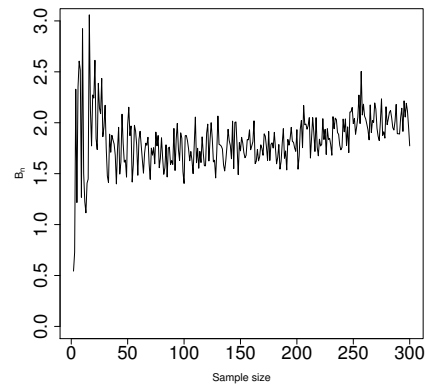
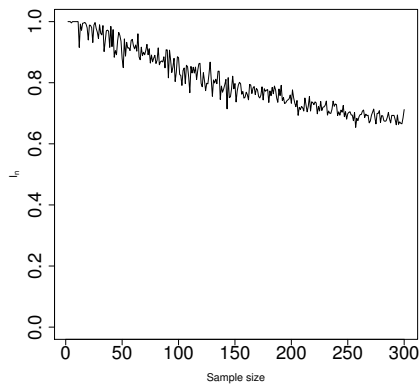
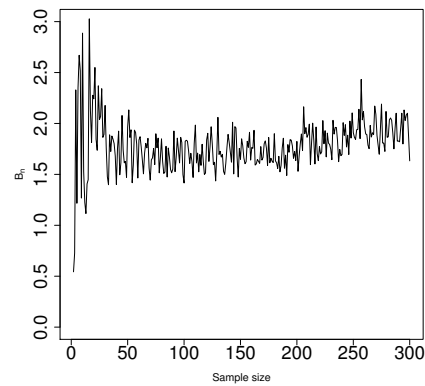
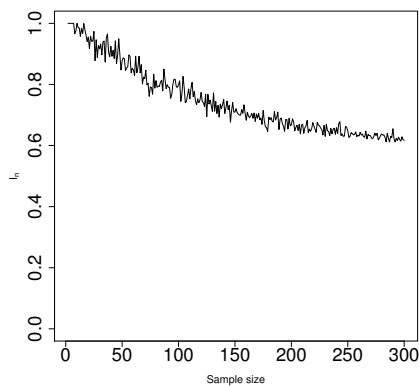
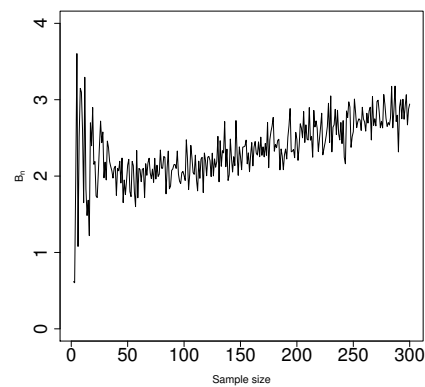
(a) I_n for $h(X_t, 0, \epsilon_t)$.(b) $B_{n,2}$ for $h(X_t, 0, \epsilon_t)$.(c) I_n for $h(X_t, \delta_t, 0)$.(d) $B_{n,2}$ for $h(X_t, \delta_t, 0)$.(e) I_n for $h(X_t, \delta_t, \epsilon_t)$.(f) $B_{n,2}$ for $h(X_t, \delta_t, \epsilon_t)$.

Figure 4.14: The anomaly-affected indices I_n and $B_{n,2}$ for the satisfactory service range with respect to $2 \leq n \leq 300$ for ARMA(1, 1) inputs and iid Lomax(11, 1) anomalies.

Proof Using bounds (7) of Gascuel and Caraux [50] and then applying Hölder's inequality, we

have

$$\begin{aligned}
\mathbb{E}(\xi_{n:n}) - \mathbb{E}(\xi_{1:n}) &\leq n \int_{1-1/n}^1 F_\xi^{-1}(y) dy - n \int_0^{1/n} F_\xi^{-1}(y) dy \\
&\leq n \int_{1-1/n}^1 |F_\xi^{-1}(y)| dy + n \int_0^{1/n} |F_\xi^{-1}(y)| dy \\
&\leq cn \left(\int_0^1 |F_\xi^{-1}(y)|^p dy \right)^{1/p} n^{-1/q} \\
&\leq cn^{1/p} \left(\mathbb{E}(|\xi_1|^p) \right)^{1/p},
\end{aligned}$$

where F_ξ denotes the cdf of ξ_1 , and $q \in [1, \infty]$ is such that $p^{-1} + q^{-1} = 1$. (When $p = 1$, we set $q = \infty$.) This proves statement (4.B.1). To prove the concluding part of the lemma, we choose any constant $\lambda > 0$ and write the bounds

$$\mathbb{P}(n^{-1/p}(\xi_{n:n} - \xi_{1:n}) > \lambda) \leq \frac{n^{-1/p}}{\lambda} \left(\mathbb{E}(\xi_{n:n}) - \mathbb{E}(\xi_{1:n}) \right) \leq \frac{c}{\lambda}$$

with a finite constant $c < \infty$ that does not depend on n and λ . This finishes the entire proof of Lemma 4.B.1.

Gascuel and Caraux [50, bounds (8)] allow different distributions of ξ_t 's, and only simple though space consuming modifications of Theorem 4.8.1 and its proof (to be later given in this appendix) are required to accommodate this case. This is significant because it implies high robustness of convergence of I_n to 1/2 with respect to possibly varied (i.e., non-stationary) marginal distributions of the outputs Y_t .

Proof of Theorem 4.5.1 We first prove part (i). Using Lipschitz continuity of h_0 , we have

$$\begin{aligned}
\frac{1}{n^{1/p}} \sum_{t=2}^n |Y_{t,n}^0 - Y_{t-1,n}^0| &\leq \frac{K}{n^{1/p}} \sum_{t=2}^n (X_{t:n} - X_{t-1:n}) \\
&= \frac{K}{n^{1/p}} (X_{n:n} - X_{1:n}).
\end{aligned} \tag{4.B.2}$$

Lemma 4.B.1 with X_t 's instead of ξ_t 's gives us the statement $n^{-1/p}(X_{n:n} - X_{1:n}) = O_{\mathbb{P}}(1)$ and completes the proof of part (i).

To prove part (ii), we start with the bound

$$\begin{aligned}
\frac{1}{n^{1/p}} \sum_{t=2}^n |Y_{t,n}^0 - Y_{t-1,n}^0| &= \frac{1}{n^{1/p}} \sum_{t=2}^n |h_0(X_{t:n}) - h_0(X_{t-1:n})| \\
&= \frac{1}{n^{1/p}} \sum_{t=2}^n \left| \int_{X_{t-1:n}}^{X_{t:n}} h_0^*(x) dx \right| \\
&\leq \frac{1}{n^{1/p}} \int_{X_{1:n}}^{X_{n:n}} |h_0^*(x)| dx.
\end{aligned} \tag{4.B.3}$$

Applying Hölder's inequality on the right-hand side of bound (4.B.3) with β such that $\alpha^{-1} + \beta^{-1} = 1$, we have

$$\begin{aligned} \frac{1}{n^{1/p}} \sum_{t=2}^n |Y_{t,n}^0 - Y_{t-1,n}^0| &\leq \frac{1}{n^{1/p}} (X_{n:n} - X_{1:n})^{1/\alpha} \left(\int_{X_{1:n}}^{X_{n:n}} |h_0^*(x)|^\beta dx \right)^{1/\beta} \\ &\leq \frac{c}{n^{1/p}} (X_{n:n} - X_{1:n})^{1/\alpha}. \end{aligned} \quad (4.B.4)$$

To show that the right-hand side of bound (4.B.4) is of order $O_{\mathbb{P}}(1)$, we fix any $\lambda > 0$ and write the bound

$$\mathbb{P}\left(\frac{1}{n^{1/p}} (X_{n:n} - X_{1:n})^{1/\alpha} > \lambda\right) \leq \frac{1}{\lambda^\alpha} n^{-\alpha/p} (\mathbb{E}(\xi_{n:n}) - \mathbb{E}(\xi_{1:n})). \quad (4.B.5)$$

Lemma 4.B.1 with X_t 's instead of ξ_t 's and with p/α instead of p shows that the right-hand side of bound (4.B.5) can be made as small as desired by choosing a sufficiently large λ and for all sufficiently large n . This completes the proof of part (ii), and that of Theorem 4.5.1 as well.

Proof of Theorem 4.7.1 Since the baseline function h_0 is absolutely continuous on $[a_X, b_X]$, there is an integrable on $[a_X, b_X]$ function h_0^* such that $h_0(v) - h_0(u) = \int_u^v h_0^*(x) dx$ for all $u, v \in [a_X, b_X]$ such that $u \leq v$. Hence,

$$\begin{aligned} Y_{n,n}^0 - Y_{1,n}^0 - (h_0(b_X) - h_0(a_X)) &= h_0(X_{n:n}) - h_0(X_{1:n}) - (h_0(b_X) - h_0(a_X)) \\ &= - \int_{X_{n:n}}^{b_X} h_0^*(x) dx - \int_{a_X}^{X_{1:n}} h_0^*(x) dx \\ &= - \int_{a_X}^{b_X} \mathbb{1}\{x \geq X_{n:n}\} h_0^*(x) dx - \int_{a_X}^{b_X} \mathbb{1}\{x < X_{1:n}\} h_0^*(x) dx. \end{aligned}$$

Consequently, for every $\lambda > 0$, using Markov's inequality we have

$$\begin{aligned} \mathbb{P}\left(|Y_{n,n}^0 - Y_{1,n}^0 - (h_0(b_X) - h_0(a_X))| > \lambda\right) &\leq \frac{1}{\lambda} \mathbb{E}\left(|Y_{n,n}^0 - Y_{1,n}^0 - (h_0(b_X) - h_0(a_X))|\right) \\ &\leq \frac{1}{\lambda} \int_{a_X}^{b_X} \mathbb{P}(X_{n:n} \leq x) |h_0^*(x)| dx \\ &\quad + \frac{1}{\lambda} \int_{a_X}^{b_X} \mathbb{P}(X_{1:n} > x) |h_0^*(x)| dx. \end{aligned} \quad (4.B.6)$$

Since the inputs X_t are temperately dependent and $\int_{a_X}^{b_X} |h_0^*(x)| dx < \infty$, the Lebesgue dominated convergence theorem implies that the two integrals on the right-hand side of bound (4.B.6) converge to 0 when $n \rightarrow \infty$. This completes the proof of Theorem 4.7.1.

Proof of Theorem 4.7.2 Fix any $t \in \mathbb{N}$ and let $n \geq t$. We have

$$\begin{aligned} F_{X_{n:n}}(x) &= \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) \\ &\leq \mathbb{P}(X_t \leq x, X_{2t} \leq x, \dots, X_{\lfloor n/t \rfloor t} \leq x) \\ &\leq \mathbb{P}(X_t \leq x) \mathbb{P}(X_{2t} \leq x, \dots, X_{\lfloor n/t \rfloor t} \leq x) + \alpha_X(t) \\ &= F(x) \mathbb{P}(X_{2t} \leq x, \dots, X_{\lfloor n/t \rfloor t} \leq x) + \alpha_X(t) \end{aligned}$$

$$\begin{aligned}
&\leq F(x)^2 \mathbb{P}(X_{3t} \leq x, \dots, X_{\lfloor n/t \rfloor t} \leq x) + \alpha_X(t)(1 + F(x)) \\
&\leq \dots \\
&\leq F(x)^{\lfloor n/t \rfloor} + \alpha_X(t)(1 + F(x) + \dots + F(x)^{\lfloor n/t \rfloor - 1}) \\
&= F(x)^{\lfloor n/t \rfloor} + \frac{\alpha_X(t)(1 - F(x)^{\lfloor n/t \rfloor})}{1 - F(x)}.
\end{aligned}$$

When $x < b_X$, we have $F(x) < 1$ and so

$$\limsup_{n \rightarrow \infty} F_{X_{n:n}}(x) \leq \frac{\alpha_X(t)}{1 - F(x)}.$$

Letting $t \rightarrow \infty$, we have

$$\limsup_{n \rightarrow \infty} F_{X_{n:n}}(x) = 0$$

and so $F_{X_{n:n}}(x) \rightarrow 0$ when $n \rightarrow \infty$. This establishes the second part of property (4.7.3).

When $x > a_X$, we set $\xi_t := -X_t$ for all $t \in \mathbb{Z}$. By the previous case, we know that $F_{\xi_{n:n}}(z) \rightarrow 0$ for all $z < b_\xi$. Since $X_{1:n} = -\xi_{n:n}$ and $a_X = -b_\xi$, we have $\mathbb{P}(X_{1:n} \geq -z) \rightarrow 0$ for all $-z > a_X$. This establishes the first part of property (4.7.3) and concludes the entire proof of Theorem 4.7.2.

To prove Theorem 4.7.4, we need a lemma.

Lemma 4.B.2 *Let the inputs X_t be strictly stationary, temperately dependent, and satisfy the Glivenko-Cantelli property. If the cdf F and the corresponding quantile function F^{-1} are continuous, then for any finite subinterval $[a, b]$ of $[a_X, b_X]$, we have*

$$\max_{1 \leq t \leq n+1} (Z_{t,n} - Z_{t-1,n}) \rightarrow_{\mathbb{P}} 0 \quad (4.B.7)$$

when $n \rightarrow \infty$, where $Z_{0,n} := a$, $Z_{n+1,n} := b$, and, for all $t = 1, \dots, n$,

$$Z_{t,n} := h_c(X_{t:n}) = \begin{cases} a & \text{when } X_{t:n} < a, \\ X_{t:n} & \text{when } a \leq X_{t:n} \leq b, \\ b & \text{when } X_{t:n} > b. \end{cases}$$

Proof Since $a_X \leq a$ and the inputs X_t are temperately dependent, we have $X_{1:n} \rightarrow_{\mathbb{P}} a_X$ and so $Z_{1,n} - a \rightarrow_{\mathbb{P}} 0$. Likewise, since $b \leq b_X$, we have $X_{n:n} \rightarrow_{\mathbb{P}} b_X$ and so $b - Z_{n,n} \rightarrow_{\mathbb{P}} 0$. Consequently, statement (4.B.7) holds provided that

$$\max_{2 \leq t \leq n} (Z_{t,n} - Z_{t-1,n}) \rightarrow_{\mathbb{P}} 0. \quad (4.B.8)$$

Note that we only need to consider those t 's for which $X_{t:n} > a$ and $X_{t-1:n} \leq b$. These two restrictions are equivalent to $F_n^{-1}(t/n) > a$ and $F_n^{-1}((t-1)/n) \leq b$, respectively. Note that $F_n^{-1}(t/n) > a$ is equivalent to $t/n > F_n(a)$, and $F_n^{-1}((t-1)/n) \leq b$ is equivalent to $(t-1)/n \leq F_n(b)$. Due to the Glivenko-Cantelli property, we therefore conclude that for any (small) $\delta > 0$ and for all sufficiently large n , all those t 's for which the bounds $X_{t:n} > a$ and $X_{t-1:n} \leq b$ hold are such that $(1 - \delta)F(a)n \leq t \leq (1 + \delta)F(a)n$. For typographical simplicity, we rewrite the latter

bounds as $\alpha n \leq t \leq \beta n$, where $\alpha := (1 - \delta)F(a)$ and $\beta := (1 + \delta)F(a)$. Hence, statement (4.B.8) follows if

$$\max_{\alpha n \leq t \leq \beta n} (Z_{t,n} - Z_{t-1,n}) \rightarrow_{\mathbb{P}} 0. \quad (4.B.9)$$

Since $Z_{t,n} - Z_{t-1,n} \leq X_{t:n} - X_{t-1:n}$ for all $t = 2, \dots, n$, statement (4.B.9) follows if

$$\max_{\alpha n \leq t \leq \beta n} (X_{t:n} - X_{t-1:n}) \rightarrow_{\mathbb{P}} 0. \quad (4.B.10)$$

To prove the latter statement, we write

$$\begin{aligned} \max_{\alpha n \leq t \leq \beta n} (X_{t:n} - X_{t-1:n}) &= \max_{\alpha n \leq t \leq \beta n} (F_n^{-1}(t/n) - F_n^{-1}((t-1)/n)) \\ &\leq \max_{\alpha n \leq t \leq \beta n} (F^{-1}(t/n) - F^{-1}((t-1)/n)) \\ &\quad + \max_{\alpha n \leq t \leq \beta n} |F_n^{-1}(t/n) - F^{-1}(t/n)| \\ &\quad + \max_{\alpha n \leq t \leq \beta n} |F_n^{-1}((t-1)/n) - F^{-1}((t-1)/n)|. \end{aligned} \quad (4.B.11)$$

The first maximum on the right-hand side of bound (4.B.11) converges to 0 because F^{-1} is continuous on $(0, 1)$ and thus uniformly continuous on every closed subinterval of $(0, 1)$. As to the second and third maxima on the right-hand side of bound (4.B.11), they converge to 0 in probability because

$$\Gamma_n := \sup_{t \in [t_0, t_1]} |F_n^{-1}(t) - F^{-1}(t)| \rightarrow_{\mathbb{P}} 0 \quad (4.B.12)$$

for every closed interval $[t_0, t_1] \subset (0, 1)$, because the Glivenko-Cantelli property holds. To show that the just noted implication is true, we proceed as follows.

Statement (4.B.12) means that, for any fixed $\gamma > 0$, the probability of the event $\Gamma_n \leq \gamma$ converges to 1 when $n \rightarrow \infty$. This event has at least the same, if not larger, probability as the event

$$F^{-1}(t) - \gamma < F_n^{-1}(t) \leq F^{-1}(t) + \gamma \quad \text{for all } t \in [t_0, t_1], \quad (4.B.13)$$

which is equivalent to

$$F_n(F^{-1}(t) - \gamma) < t \leq F_n(F^{-1}(t) + \gamma) \quad \text{for all } t \in [t_0, t_1].$$

The latter event has at least the same, if not larger, probability as the event

$$F(F^{-1}(t) - \gamma) + \|F_n - F\| < t \leq F(F^{-1}(t) + \gamma) - \|F_n - F\| \quad \text{for all } t \in [t_0, t_1]. \quad (4.B.14)$$

Since $t = F(F^{-1}(t))$, event (4.B.14) has at least the same, if not larger, probability as the event

$$-\Delta_1(\gamma) + \|F_n - F\| < 0 \leq \Delta_2(\gamma) - \|F_n - F\|, \quad (4.B.15)$$

where

$$\Delta_1(\gamma) := \inf_{t \in [t_0, t_1]} (F(F^{-1}(t)) - F(F^{-1}(t) - \gamma))$$

and

$$\Delta_2(\gamma) := \inf_{t \in [t_0, t_1]} (F(F^{-1}(t) + \gamma) - F^{-1}(F^{-1}(t))).$$

Since the cdf F is strictly increasing (because we have assumed that F^{-1} is continuous), the quantities $\Delta_1(\gamma)$ and $\Delta_2(\gamma)$ are (strictly) positive for every $\gamma > 0$. We therefore conclude that statement (4.B.15) holds with as large a probability as desired, provided that n is sufficiently large. This, in turn, implies that event (4.B.13) can be made as close to 1 as desired, provided that n is sufficiently large. The proof of Lemma 4.B.2 is finished.

Proof of Theorem 4.7.4 Since the baseline function h_0 is absolutely continuous on $[a_X, b_X]$ and its Radon-Nikodym derivative h_0^* vanishes outside the interval $[a, b]$, we have

$$\begin{aligned} |Y_{t,n}^0 - Y_{t-1,n}^0| &= \left| \int_{X_{t,n}}^{X_{t-1,n}} h_0^*(x) dx \right| \\ &= \left| \int_{[X_{t-1,n}, X_{t,n}] \cap [a,b]} h_0^*(x) dx \right| = |h_0^*(\xi_{t,n})|(Z_{t,n} - Z_{t-1,n}), \end{aligned}$$

where, due to the mean-value theorem, the right-most equation holds for some $\xi_{t,n} \in [Z_{t-1,n}, Z_{t,n}]$ with $Z_{t,n}$'s defined in Lemma 4.B.2. Consequently,

$$\begin{aligned} \Theta_n &:= \sum_{t=2}^n |Y_{t,n}^0 - Y_{t-1,n}^0| - \int_a^b |h_0^*(x)| dx \\ &= \sum_{t=2}^n |h_0^*(\xi_{t,n})|(Z_{t,n} - Z_{t-1,n}) - \int_a^b |h_0^*(x)| dx. \end{aligned}$$

Obviously, $Z_{t,n} \in [a, b]$ for all $t = 2, \dots, n$. We also have $Z_{0,n} = a$ and $Z_{n+1,n} = b$. By Lemma 4.B.2,

$$\max_{1 \leq t \leq n+1} (Z_{t,n} - Z_{t-1,n}) \rightarrow_{\mathbb{P}} 0. \quad (4.B.16)$$

Furthermore,

$$\begin{aligned} \Theta_n &= \sum_{t=1}^{n+1} |h_0^*(\xi_{t,n})|(Z_{t,n} - Z_{t-1,n}) - \int_a^b |h_0^*(x)| dx - \sum_{t \in \{1, n+1\}} |h_0^*(\xi_{t,n})|(Z_{t,n} - Z_{t-1,n}) \\ &= \sum_{t=1}^{n+1} |h_0^*(\xi_{t,n})|(Z_{t,n} - Z_{t-1,n}) - \int_a^b |h_0^*(x)| dx + o_{\mathbb{P}}(1), \end{aligned} \quad (4.B.17)$$

where the last equation holds when $n \rightarrow \infty$ because the function h_0^* is bounded, and $Z_{1,n} - a \rightarrow_{\mathbb{P}} 0$ and $b - Z_{n,n} \rightarrow_{\mathbb{P}} 0$ when $n \rightarrow \infty$, which we verified at the beginning of the proof of Lemma 4.B.2. Hence, equation (4.B.17) holds, and in order to prove $\Theta_n \rightarrow_{\mathbb{P}} 0$, we need to show

$$\Theta_n^* := \sum_{t=1}^{n+1} |h_0^*(\xi_{t,n})|(Z_{t,n} - Z_{t-1,n}) - \int_a^b |h_0^*(x)| dx \rightarrow_{\mathbb{P}} 0. \quad (4.B.18)$$

In other words, we need to show that, for every $\gamma > 0$,

$$\mathbb{P}\left(|\Theta_n^*| \geq \gamma\right) \rightarrow 0 \quad (4.B.19)$$

when $n \rightarrow \infty$. For this, we first rewrite statement (4.B.16) explicitly: for every $\lambda > 0$,

$$\mathbb{P}\left(\max_{1 \leq t \leq n+1} (Z_{t,n} - Z_{t-1,n}) \geq \lambda\right) \rightarrow 0 \quad (4.B.20)$$

when $n \rightarrow \infty$. Hence, statement (4.B.19) follows if, for any $\gamma > 0$, we can find $\lambda > 0$ such that

$$\mathbb{P}\left(|\Theta_n^*| \geq \gamma, \max_{1 \leq t \leq n+1} (Z_{t,n} - Z_{t-1,n}) < \lambda\right) \rightarrow 0 \quad (4.B.21)$$

when $n \rightarrow \infty$. We now recall the very basic definition of Riemann integral, according to which, for any $\gamma > 0$, we can find $\lambda > 0$ such that

$$\left| \sum_{t=1}^{n+1} |h_0^*(\zeta_{t,n})|(z_{t,n} - z_{t-1,n}) - \int_a^b |h_0^*(x)| dx \right| < \gamma$$

whenever

$$\max_{1 \leq t \leq n+1} (z_{t,n} - z_{t-1,n}) < \lambda, \quad (4.B.22)$$

where $z_{0,n} := a$, $z_{n+1,n} := b$, $z_{t-1,n} \leq z_{t,n}$ for $t = 1, \dots, n+1$, and $\zeta_{t,n} \in [z_{t-1,n}, z_{t,n}]$. Hence, with the same $\lambda > 0$ as in statement (4.B.22), probability (4.B.21) is equal to 0. This establishes statement (4.B.19) and finishes the proof of Theorem 4.7.4.

Proof of Corollary 4.7.5 By Theorems 4.7.1 and 4.7.4, we have

$$I_n^0 \rightarrow_{\mathbb{P}} \frac{1}{2} \left(1 + \frac{h_0(b_X) - h_0(a_X)}{\int_a^b |h_0^*(x)| dx} \right)$$

when $n \rightarrow \infty$. Furthermore, we have

$$h_0(b_X) - h_0(a_X) = h_0(b) - h_0(a) = \int_a^b h_0^*(x) dx$$

because the Radon-Nikodym derivative h_0^* of h_0 vanishes outside the interval $[a, b]$ and the positive part z_+ of every real number $z \in \mathbb{R}$ can be written as $(|z| + z)/2$. This concludes the proof of Corollary 4.7.5.

Proof of Theorem 4.8.1 Since $z_+ = (|z| + z)/2$ for every real number $z \in \mathbb{R}$, we have

$$I_n = \frac{1}{2} \left(1 + \frac{Y_{n,n} - Y_{1,n}}{\sum_{t=2}^n |Y_{t,n} - Y_{t-1,n}|} \right).$$

Lemma 4.B.1 with Y_t 's instead of ξ_t 's says that $n^{-1/p}(Y_{n:n} - Y_{1:n}) = O_{\mathbb{P}}(1)$. Since the system is out of p -reasonable order, we have $n^{-1/p} \sum_{t=2}^n |Y_{t,n} - Y_{t-1,n}| \rightarrow_{\mathbb{P}} \infty$ when $n \rightarrow \infty$ and thus $I_n \rightarrow_{\mathbb{P}} 1/2$. This concludes the proof of Theorem 4.8.1.

To prove Theorem 4.8.2, we need a formula for $B_{n,p}$ analogous to equation (4.8.1).

Lemma 4.B.3 *The concomitants $Y_{1,n}, \dots, Y_{n,n}$ of the outputs $Y_t = h(X_t, \boldsymbol{\varepsilon}_t)$, $t = 1, \dots, n$, with respect to the inputs X_1, \dots, X_n are given by*

$$Y_{t,n} = h(X_{t,n}, \boldsymbol{\varepsilon}_{t,n}),$$

where $\boldsymbol{\varepsilon}_{1,n}, \dots, \boldsymbol{\varepsilon}_{n,n}$ are the concomitants of the anomalies $\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n$ with respect to X_1, \dots, X_n , that is,

$$\boldsymbol{\varepsilon}_{t,n} = \sum_{s=1}^n \boldsymbol{\varepsilon}_s \mathbb{1}\{X_s = X_{t:n}\}.$$

Consequently,

$$B_{n,p} = \frac{1}{n^{1/p}} \sum_{t=2}^n |h(X_{t:n}, \boldsymbol{\varepsilon}_{t,n}) - h(X_{t-1:n}, \boldsymbol{\varepsilon}_{t-1,n})|. \quad (4.B.23)$$

Proof Since the cdf F of each input X_t is continuous, we can assume without loss of generality that all the inputs X_1, \dots, X_n are unequal. Hence, we can write the equation

$$\boldsymbol{\varepsilon}_t = \sum_{s=1}^n \boldsymbol{\varepsilon}_s \mathbb{1}\{X_s = X_t\}.$$

This implies that the concomitants of the outputs Y_1, \dots, Y_n with respect to the inputs X_1, \dots, X_n can be expressed as follows:

$$\begin{aligned} Y_{t,n} &= \sum_{s=1}^n Y_s \mathbb{1}\{X_s = X_{t:n}\} \\ &= \sum_{s=1}^n h(X_s, \boldsymbol{\varepsilon}_s) \mathbb{1}\{X_s = X_{t:n}\} \\ &= \sum_{s=1}^n h(X_{s:n}, \boldsymbol{\varepsilon}_{s,n}) \mathbb{1}\{X_{s:n} = X_{t:n}\} \\ &= h(X_{t:n}, \boldsymbol{\varepsilon}_{t,n}). \end{aligned}$$

This establishes equation (4.B.23) and concludes the proof of Lemma 4.B.3.

Note 4.B.4 Kim and David [75, Section 4] use the notation $\boldsymbol{\varepsilon}_{[t]}$ instead of $\boldsymbol{\varepsilon}_{t,n}$, in which case the equation $Y_{t,n} = h(X_{t:n}, \boldsymbol{\varepsilon}_{t,n})$ turns into $Y_{t,n} = h(X_{t:n}, \boldsymbol{\varepsilon}_{[t]})$. We prefer the notation $\boldsymbol{\varepsilon}_{t,n}$ as it reminds us that the anomaly concomitants depend on the sample size n .

Proof of Theorem 4.8.2 Since the anomaly-free outputs Y_t^0 are in p -reasonable order with respect to the inputs X_t for some $p > 0$, we have $B_{n,p}^0 = O_{\mathbb{P}}(1)$. By Lemma 4.B.3, we have

$$\begin{aligned} B_{n,p} &= \frac{1}{n^{1/p}} \sum_{t=2}^n |h(X_{t:n}, \mathbf{0}, \boldsymbol{\varepsilon}_{t,n}) - h(X_{t-1:n}, \mathbf{0}, \boldsymbol{\varepsilon}_{t-1,n})| \\ &= \frac{1}{n^{1/p}} \sum_{t=2}^n |h_0(X_{t:n}) + \boldsymbol{\varepsilon}_{t,n} - h_0(X_{t-1:n}) - \boldsymbol{\varepsilon}_{t-1,n}| \\ &\geq \frac{1}{n^{1/p}} \sum_{t=2}^n |\boldsymbol{\varepsilon}_{t,n} - \boldsymbol{\varepsilon}_{t-1,n}| - \frac{1}{n^{1/p}} \sum_{t=2}^n |h_0(X_{t:n}) - h_0(X_{t-1:n})| \\ &= \frac{1}{n^{1/p}} \sum_{t=2}^n |\boldsymbol{\varepsilon}_{t,n} - \boldsymbol{\varepsilon}_{t-1,n}| + O_{\mathbb{P}}(1). \end{aligned}$$

Hence, if the output anomalies ϵ_t are out of p -reasonable order with respect to the inputs, meaning that

$$\frac{1}{n^{1/p}} \sum_{t=2}^n |\epsilon_{t,n} - \epsilon_{t-1,n}| \rightarrow_{\mathbb{P}} \infty,$$

then $B_{n,p} \rightarrow_{\mathbb{P}} \infty$ when $n \rightarrow \infty$.

Conversely, if $B_{n,p} \rightarrow_{\mathbb{P}} \infty$ when $n \rightarrow \infty$, then the bound

$$\begin{aligned} B_{n,p} &\leq \frac{1}{n^{1/p}} \sum_{t=2}^n |\epsilon_{t,n} - \epsilon_{t-1,n}| + \frac{1}{n^{1/p}} \sum_{t=2}^n |h_0(X_{t,n}) - h_0(X_{t-1,n})| \\ &= \frac{1}{n^{1/p}} \sum_{t=2}^n |\epsilon_{t,n} - \epsilon_{t-1,n}| + O_{\mathbb{P}}(1) \end{aligned}$$

implies that the output anomalies ϵ_t are out of p -reasonable order with respect to the inputs. This concludes the proof of Theorem 4.8.2.

Lemma 4.B.5 *If random variables ξ_t are iid and have finite first moments, then*

$$\frac{1}{n^{1/p}} \sum_{t=2}^n |\xi_t - \xi_{t-1}| \rightarrow_{\mathbb{P}} \infty \quad (4.B.24)$$

whenever the distribution of ξ_1 is non-degenerate.

Proof Since the summands $|\xi_t - \xi_{t-1}|$, $t = 2, 3, \dots$, are 1-dependent, splitting the sum into the sums with respect to even and odd t 's yields statement (4.B.24) for every $p > 1$ if the moment $\mathbb{E}(|\xi_2 - \xi_1|)$ is (strictly) positive. Since ξ_2 and ξ_1 are iid, the aforementioned moment is positive whenever the distribution of ξ_1 is non-degenerate.

Proof of Theorem 4.8.4 Since the baseline function h_0 is Lipschitz continuous, Theorem 4.5.1 implies that the anomaly-free outputs Y_t^0 are in p -reasonable order with respect to the inputs X_t . Consequently, $B_{n,p}^0 = O_{\mathbb{P}}(1)$. By Lemma 4.B.3, we have

$$B_{n,p} = \frac{1}{n^{1/p}} \sum_{t=2}^n |h(X_{t,n}, \delta_{t,n}, 0) - h(X_{t-1,n}, \delta_{t-1,n}, 0)|.$$

Using Lipschitz continuity of h_0 and also Lemma 4.B.1 with X_t instead of ξ_t , we have

$$\begin{aligned} B_{n,p} &= \frac{1}{n^{1/p}} \sum_{t=2}^n |h_0(X_{t,n} + \delta_{t,n}) - h_0(X_{t,n} + \delta_{t-1,n}) + h_0(X_{t,n} + \delta_{t-1,n}) - h_0(X_{t-1,n} + \delta_{t-1,n})| \\ &\geq \frac{1}{n^{1/p}} \sum_{t=2}^n |h_0(X_{t,n} + \delta_{t,n}) - h_0(X_{t,n} + \delta_{t-1,n})| - \frac{K}{n^{1/p}} \sum_{t=2}^n (X_{t,n} - X_{t-1,n}) \\ &= \frac{1}{n^{1/p}} \sum_{t=2}^n |h_0(X_{t,n} + \delta_{t,n}) - h_0(X_{t,n} + \delta_{t-1,n})| - \frac{K}{n^{1/p}} (X_{n,n} - X_{1,n}) \end{aligned}$$

$$= \frac{1}{n^{1/p}} \sum_{t=2}^n |h_0(X_{t:n} + \delta_{t,n}) - h_0(X_{t:n} + \delta_{t-1,n})| + O_{\mathbb{P}}(1).$$

Hence, with the notation

$$\xi_t = |h_0(X_t + \delta_t) - h_0(X_t + \delta_{t-1})|,$$

we are left to check the statement

$$\frac{1}{n^{1/p}} \sum_{t=2}^n \xi_t \xrightarrow{\mathbb{P}} \infty. \quad (4.B.25)$$

That is, we need to show that for every $\lambda < \infty$, we have

$$\mathbb{P}\left(\frac{1}{n^{1/p}} \sum_{t=2}^n \xi_t \leq \lambda\right) \rightarrow 0$$

when $n \rightarrow \infty$. Since the moment $\mu := \mathbb{E}(\xi_1) = \mathbb{E}(\xi_t)$ is strictly positive by assumption (ii), for all sufficiently large n we have

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n^{1/p}} \sum_{t=2}^n \xi_t \leq \lambda\right) &\leq \mathbb{P}\left(-\left|\frac{1}{n^{1/p}} \sum_{t=2}^n (\xi_t - \mu)\right| + n^{1-1/p}\mu \leq \lambda\right) \\ &= \mathbb{P}\left(\left|\frac{1}{n^{1/p}} \sum_{t=2}^n (\xi_t - \mu)\right| \geq n^{1-1/p}\mu - \lambda\right) \\ &\leq \frac{1}{(n^{1-1/p}\mu - \lambda)^2} \mathbb{E}\left(\left|\frac{1}{n^{1/p}} \sum_{t=2}^n (\xi_t - \mu)\right|^2\right) \\ &\leq \frac{c}{n^2} \mathbb{E}\left(\left|\sum_{t=2}^n (\xi_t - \mu)\right|^2\right). \end{aligned} \quad (4.B.26)$$

The use of the second moment for bounding the probability was prudent because ξ_t 's are bounded by a constant, which follows because the baseline function h_0 is bounded. Hence, our task becomes to prove

$$\frac{1}{n^2} \mathbb{E}\left(\left|\sum_{t=2}^n (\xi_t - \mu)\right|^2\right) \rightarrow 0. \quad (4.B.27)$$

Since the sequence ξ_t is strictly stationary, by Rio [57, Corollary 1.2, p. 10] we have

$$\begin{aligned} \mathbb{E}\left(\left|\sum_{t=2}^n (\xi_t - \mu)\right|^2\right) &= \sum_{t=2}^n \text{Var}(\xi_t) + 2 \sum_{2 \leq s < t \leq n} \text{Cov}(\xi_s, \xi_t) \\ &= (n-1)\text{Var}(\xi_0) + 2 \sum_{t=1}^{n-2} (n-1-t)\text{Cov}(\xi_0, \xi_t) \\ &\leq c(n-1)\alpha_{\xi}(0) + c(n-2)\alpha_{\xi}(1) + c \sum_{t=2}^{n-2} (n-1-t)\alpha_{\xi}(t), \end{aligned} \quad (4.B.28)$$

where c is a finite constant that depends on h_0 . Next, for every $t \geq 2$, we have

$$\begin{aligned} \alpha_\xi(t) &= \sup \left\{ \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right| : A \in \sigma(\xi_u, u \leq 0), B \in \sigma(\xi_v, v \geq t) \right\} \\ &\leq \sup \left\{ \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right| : A \in \sigma(X_u, \delta_u, \delta_{u-1}, u \leq 0), B \in \sigma(X_v, \delta_v, \delta_{v-1}, v \geq t) \right\} \\ &\leq \sup \left\{ \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right| : A \in \sigma(X_u, u \leq 0), B \in \sigma(X_v, v \geq t) \right\}, \end{aligned} \quad (4.B.29)$$

where the last inequality holds because

$$\begin{aligned} \left| \mathbb{P}(A \cap C \cap B \cap D) - \mathbb{P}(A \cap C)\mathbb{P}(B \cap D) \right| &= \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right| \mathbb{P}(C \cap D) \\ &\leq \left| \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) \right| \end{aligned}$$

for all $A \in \sigma(X_u, u \leq 0)$, $B \in \sigma(X_v, v \geq t)$, and all $C \in \sigma(\delta_u, u \leq 0)$, $D \in \sigma(\delta_v, v \geq t-1)$, upon recalling that A and B are independent of C and D , and also C and D are independent of each other as long as $t \geq 2$. Note that the right-hand side of bound (4.B.29) is equal to $\alpha_X(t)$, and so we have the bound $\alpha_\xi(t) \leq \alpha_X(t)$ for all $t \geq 2$. This result together with bound (4.B.28) imply

$$\frac{1}{n^2} \mathbb{E} \left(\left| \sum_{t=2}^n (\xi_t - \mu) \right|^2 \right) \leq \frac{c}{n} \alpha_X(0) + \frac{c}{n} \alpha_X(1) + \frac{c}{n^2} \sum_{t=2}^{n-2} (n-1-t) \alpha_X(t)$$

with a finite constant c . The right-hand side of the latter bound converges to 0 when $n \rightarrow \infty$, provided that $\alpha_X(t)$ converges to 0 when $t \rightarrow \infty$, which is true because the inputs X_t are α -mixing. This concludes the proof of Theorem 4.8.4.

Proof of Theorem 4.8.5 Since the baseline function h_0 is Lipschitz continuous, Theorem 4.5.1 implies that the anomaly-free outputs Y_t^0 are in p -reasonable order with respect to the inputs X_t . Consequently, $B_{n,p}^0 = O_{\mathbb{P}}(1)$. By Lemma 4.B.3, we have

$$B_{n,p} = \frac{1}{n^{1/p}} \sum_{t=2}^n |h(X_{t:n}, \delta_{t,n}, \epsilon_{t,n}) - h(X_{t-1:n}, \delta_{t-1,n}, \epsilon_{t-1,n})|.$$

Proceeding analogously as in the proof of Theorem 4.8.4, we have

$$B_{n,p} \geq \frac{1}{n^{1/p}} \sum_{t=2}^n \zeta_t + O_{\mathbb{P}}(1)$$

when $n \rightarrow \infty$, where

$$\zeta_t = |h_0(X_t + \delta_t) + \epsilon_t - h_0(X_t + \delta_{t-1}) - \epsilon_{t-1}|.$$

The rest is analogous to the proof of Theorem 4.8.4 starting with statement (4.B.25), and we thus skip the details. This finishes the proof of Theorem 4.8.5.

Bibliography

- [1] D. E. Denning, An intrusion-detection model, *IEEE Transactions on Software Engineering* SE-13 (2) (1987) 222–232. [63](#)
- [2] H. Debar, M. Dacier, A. Wespi, Towards a taxonomy of intrusion-detection systems, *Computer Networks* 31 (8) (1999) 805–822.
- [3] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, S. Sastry, Attacks against process control systems, in: *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security - ASIACCS '11*, ACM Press, 2011.
- [4] N. A. Premathilaka, A. C. Aponso, N. Krishnarajah, Review on state of art intrusion detection systems designed for the cloud computing paradigm, in: *2013 47th International Carnahan Conference on Security Technology (ICCST)*, IEEE, 2013. [63](#)
- [5] G. Liang, J. Zhao, F. Luo, S. R. Weller, Z. Y. Dong, A review of false data injection attacks against modern power systems, *IEEE Transactions on Smart Grid* 8 (4) (2017) 1630–1638. [63](#)
- [6] Y. He, G. J. Mendis, J. Wei, Real-time detection of false data injection attacks in smart grid: a deep learning-based intelligent mechanism, *IEEE Transactions on Smart Grid* 8 (5) (2017) 2505–2516. [63](#)
- [7] Y. Huang, J. Tang, Y. Cheng, H. Li, K. A. Campbell, Z. Han, Real-time detection of false data injection in smart grid networks: an adaptive cusum method and analysis, *IEEE Systems Journal* 10 (2) (2016) 532–543. [63](#)
- [8] T. Onoda, Probabilistic models-based intrusion detection using sequence characteristics in control system communication, *Neural Computing and Applications* 27 (5) (2016) 1119–1127. [63](#)
- [9] S. Potluri, C. Diedrich, G. K. R. Sangala, Identifying false data injection attacks in industrial control systems using artificial neural networks, in: *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, IEEE, 2017. [63](#)
- [10] Y. Zhang, Optimized detection algorithm of complex intrusion interference signal in mobile wireless network, *Journal of Discrete Mathematical Sciences and Cryptography* 21 (3) (2018) 771–779. [63](#)
- [11] D. Chen, M. Xu, W. Shi, Defending a cyber system with early warning mechanism, *Reliability Engineering & System Safety* 169 (2018) 224–234. [63](#)
- [12] M. D. Hossain, H. Inoue, H. Ochiai, D. Fall, Y. Kadobayashi, LSTM-based intrusion detection system for in-vehicle can bus communications, *IEEE Access* 8 (2020) 185489–185502. [63](#)
- [13] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection, *ACM Computing Surveys* 41 (3) (2009) 1–58. [63](#)

- [14] M. H. Bhuyan, D. K. Bhattacharyya, J. K. Kalita, Network anomaly detection: methods, systems and tools, *IEEE Communications Surveys & Tutorials* 16 (1) (2014) 303–336. [63](#)
- [15] A. Fisch, *Novel Methods for Anomaly Detection*, Lancaster University, Ph.D. Dissertation, 2020. [63](#), [64](#)
- [16] T. Aven, P. Baraldi, R. Flage, E. Zio, *Uncertainty in Risk Assessment: the Representation and Treatment of Uncertainties by Probabilistic and Non-probabilistic Methods*, John Wiley & Sons, 2014. [63](#)
- [17] E. Zio, The future of risk assessment, *Reliability Engineering & System Safety* 177 (2018) 176–190. [63](#)
- [18] L. Cheng, F. Liu, D. D. Yao, Enterprise data breach: causes, challenges, prevention, and future directions, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 7 (5) (2017) e1211. [63](#)
- [19] M. Barahona, C.-S. Poon, Detection of nonlinear dynamics in short, noisy time series, *Nature* 381 (6579) (1996) 215–217. [63](#)
- [20] X. Hu, M. Xu, S. Xu, P. Zhao, Multiple cyber attacks against a target with observation errors and dependent outcomes: Characterization and optimization, *Reliability Engineering & System Safety* 159 (2017) 119–133. [63](#)
- [21] A. Dasgupta, B. Li, Detection and Analysis of Spikes in a Random Sequence, *Methodology and Computing in Applied Probability* 20 (4) (2018) 1429–1451. [63](#)
- [22] A. T. M. Fisch, I. A. Eckley, P. Fearnhead, A linear time method for the detection of point and collective anomalies, Technical Report, arXiv:1806.01947 . [63](#)
- [23] A. Fisch, D. Grose, I. A. Eckley, P. Fearnhead, L. Bardwell, Anomaly: detection of anomalous structure in time series data, Technical Report, arXiv:2010.09353 . [64](#)
- [24] A. T. M. Fisch, I. A. Eckley, P. Fearnhead, Subset multivariate collective and point anomaly detection, Technical Report, arXiv:1909.01691 . [64](#)
- [25] X. J. Jeng, T. T. Cai, H. Li, Simultaneous discovery of rare and common segment variants, *Biometrika* 100 (1) (2013) 157–172. [64](#)
- [26] L. Bardwell, P. Fearnhead, Bayesian detection of abnormal segments in multiple time series, *Bayesian Analysis* 12 (1). [64](#)
- [27] A. T. M. Fisch, L. Bardwell, I. A. Eckley, Real time anomaly detection and categorisation, Technical Report, arXiv:2009.06670 . [64](#)
- [28] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung, *Time Series Analysis*, John Wiley & Sons, 2015. [64](#), [74](#)

- [29] P. J. Brockwell, R. A. Davis, *Time Series: Theory and Methods*, Springer New York, 1991. 64
- [30] N. Gribkova, R. Zitikis, Detecting intrusions in control systems : A rule of thumb, its justification and illustrations, *Journal of Statistics and Management Systems* 23 (8) (2020) 1285–1304. 64, 71
- [31] I. Finkelshtain, O. Kella, M. Scarsini, On risk aversion with two risks, *Journal of Mathematical Economics* 31 (2) (1999) 239–250. 65
- [32] G. Franke, H. Schlesinger, R. C. Stapleton, Multiplicative Background Risk, *Management Science* 52 (1) (2006) 146–153.
- [33] G. Franke, H. Schlesinger, R. C. Stapleton, Risk taking with additive and multiplicative background risks, *Journal of Economic Theory* 146 (4) (2011) 1547–1568.
- [34] X. Guo, A. Wagener, W.-K. Wong, L. Zhu, The two-moment decision model with additive risks, *Risk Management* 20 (1) (2018) 77–94. 65
- [35] J. Perote, J. P.-P. na, M. Vorsatz, Strategic behavior in regressions: an experimental study, *Theory and Decision* 79 (3) (2015) 517–546. 65
- [36] J. Su, *Multiple Risk Factors Dependence Structures with Applications to Actuarial Risk Management*, York University, ph.D. Dissertation, 2016. 65
- [37] V. Semenikhine, E. Furman, J. Su, On a multiplicative multivariate gamma distribution with applications in insurance, *Risks* 6 (3) (2018) 79. 65
- [38] X. Guo, R. H. Chan, W.-K. Wong, L. Zhu, Mean–variance, mean–VaR, and mean–CVaR models for portfolio selection with background risk, *Risk Management* 21 (2) (2019) 73–98. 65
- [39] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer New York, 2009. 65
- [40] J. Perote, J. Perote-Peña, Strategy-proof estimators for simple regression, *Mathematical Social Sciences* 47 (2) (2004) 153–176. 65
- [41] P. J. Brockwell, R. A. Davis, *Introduction to Time Series and Forecasting*, Springer International Publishing, 2016. 66, 68
- [42] E. Çelik, R. Durgut, Performance enhancement of automatic voltage regulator by modified cost function and symbiotic organisms search algorithm, *Engineering Science and Technology, an International Journal* 21 (5) (2018) 1104–1111. 69
- [43] H. Gozde, Robust 2DOF state-feedback PI-controller based on meta-heuristic optimization for automatic voltage regulation system, *ISA Transactions* 98 (2020) 26–36. 69
- [44] N. Gribkova, R. Zitikis, A user-friendly algorithm for detecting the influence of background risks on a model, *Risks* 6 (3) (2018) 100. 71

- [45] N. Gribkova, R. Zitikis, Statistical detection and classification of background risks affecting inputs and outputs, *METRON* 77 (1) (2019) 1–18. [71](#)
- [46] C.-D. Lai, M. Xie, *Stochastic Ageing and Dependence for Reliability*, Springer New York, 2006. [74](#)
- [47] D. Heath, S. Resnick, G. Samorodnitsky, Heavy tails and long range dependence in on/off processes and associated fluid models, *Mathematics of Operations Research* 23 (1) (1998) 145–165. [75](#)
- [48] T. Maillart, D. Sornette, Heavy-tailed distribution of cyber-risks, *The European Physical Journal B* 75 (3) (2010) 357–364. [75](#)
- [49] B. Edwards, S. Hofmeyr, S. Forrest, Hype and heavy tails: A closer look at data breaches, *Journal of Cybersecurity* 2 (1) (2016) 3–14. [75](#)
- [50] O. Gascuel, G. Caraux, Bounds on expectations of order statistics via extremal dependences, *Statistics & Probability Letters* 15 (2) (1992) 143–148. [79](#), [89](#), [90](#)
- [51] G. L. O’Brien, Extreme values for stationary and markov sequences, *The Annals of Probability* 15 (1). [79](#)
- [52] A. Jakubowski, *Asymptotic Independent Representations for Sums and Order Statistics of Stationary Sequences*, Uniwersytet Mikołaja Kopernika, Toruń, Poland, 1991. [79](#)
- [53] R. Bradley, *Introduction to Strong Mixing Conditions*, Kendrick Press, Heber City, Utah, 2007. [79](#)
- [54] P. Doukhan, A. Jakubowski, G. Lang, Phantom distribution functions for some stationary sequences, *Extremes* 18 (4) (2015) 697–725. [79](#)
- [55] A. Jakubowski, An asymptotic independent representation in limit theorems for maxima of nonstationary random sequences, *The Annals of Probability* 21 (2). [79](#)
- [56] Z. Lin, C. Lu, *Limit Theory for Mixing Dependent Random Variables*, Springer Netherlands, Dordrecht, 1996. [79](#)
- [57] E. Rio, *Asymptotic theory of weakly dependent random processes*, Springer, 2017. [79](#), [80](#), [98](#)
- [58] A. Mokkadem, Mixing properties of ARMA processes, *Stochastic Processes and their Applications* 29 (2) (1988) 309–315. [79](#), [80](#)
- [59] Z. Cai, G. G. Roussas, Uniform strong estimation under α -mixing, with rates, *Statistics & Probability Letters* 15 (1) (1992) 47–55. [80](#)
- [60] Y. A. Davydov, Mixing Conditions for Markov Chains, *Theory of Probability and Its Applications* 18 (2) (1973) 312–328. [80](#)

- [61] Y. Davydov, R. Zitikis, Quantifying non-monotonicity of functions and the lack of positivity in signed measures, *Modern Stochastics: Theory and Applications* 4 (3) (2017) 219–231. 82, 85
- [62] L. Chen, Y. Davydov, N. Gribkova, R. Zitikis, Estimating the index of increase via balancing deterministic and random data, *Mathematical Methods of Statistics* 27 (2) (2018) 83–102. 82
- [63] Y. Davydov, E. Moldavskaya, R. Zitikis, Searching for and quantifying nonconvexity regions of functions, *Lithuanian Mathematical Journal* 59 (4) (2019) 507–518. 82, 85
- [64] B. T. Polyak, Gradient methods for solving equations and inequalities, *USSR Computational Mathematics and Mathematical Physics* 4 (6) (1964) 17–32. 82
- [65] D. Tse, P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005. 84
- [66] R. S. Kshetrimayum, *Fundamentals of MIMO Wireless Communications*, Cambridge University Press, 2017. 84
- [67] G. Koshevoy, The Lorenz zonotope and multivariate majorizations, *Social Choice and Welfare* 15 (1) (1997) 1–14. 85
- [68] K. Mosler, *Multivariate Dispersion, Central Regions, and Depth*, Springer New York, 2002. 85
- [69] V. Vovk, R. Wang, E-values: calibration, combination, and applications, *Annals of Statistics To appear*. 85
- [70] R. Wang, A. Ramdas, False discovery rate control with e-values, Technical Report, arXiv:2009.02824 . 85
- [71] M. Barni, F. Pérez-González, Coping with the enemy: Advances in adversary-aware signal processing, in: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 8682–8686, 2013. 85
- [72] B. Biggio, F. Roli, Wild patterns: Ten years after the rise of adversarial machine learning, *Pattern Recognition* 84 (2018) 317–331. 85
- [73] M. Barni, B. Tondi, Adversarial source identification game with corrupted training, *IEEE Transactions on Information Theory* 64 (5) (2018) 3894–3915. 85
- [74] E. Alhajjar, P. Maxwell, N. Bastian, Adversarial machine learning in network intrusion detection systems, *Expert Systems with Applications* 186 (2021) 115782. 85
- [75] S. Kim, H. David, On the dependence structure of order statistics and concomitants of order statistics, *Journal of statistical planning and inference* 24 (3) (1990) 363–368. 96

Chapter 5

Future work

5.1 About the TOMD κ^*

In this research, we propose the average block-minima estimator for the TOMD κ^* using observed data. This estimator successfully avoids the complexity of involving information of the underlying paths of maximal dependence. However, there are still possibilities to improve the average block-minima estimator at the current stage, for example, the block size m of the average block-minima estimator is currently fixed. It is very likely that by selecting $m = m_n \in \{1, \dots, n\}$ given the available sample size n the average block-minima estimator may converge faster, which is actually very important in the application of this estimator because converging faster means fewer observations are needed to achieve a reasonable estimation result given that majority of the observations will be excluded even before applying the average block-minima estimation methods when estimating a TOMD in practice.

Based on our numerical examples we could roughly conclude that the average block-minima estimator works well even for observations with stationary time series structure. However, as the readers may notice, the construction of the average block-minima estimator consists of many steps while we do not give theoretical analysis for each of the steps, which include

1. How to determine the risk level $q \in (0, 1)$ in order to make the temporal dependence structure of the extracted data within $\mathcal{R}_{q,n}(\mathbf{0})$ close to white noise;
2. The properties of the optimized $\varphi_n^*(q)$ given the risk level $q \in (0, 1)$ and the total sample size n ;
3. Whether the asymptotic behavior of the average block-minima estimator $\hat{\kappa}_{m,q,n}(m, \theta, q)$ holds if all values $(\varphi^*(q), \Pi^*(q))$ in proof of Theorem 2.3.5 are substituted by their empirical versions $(\varphi^*(q)_n, \Pi_n^*(q))$;
4. The asymptotic normality of the average block-minima estimator, which seems to hold according to our simulation studies while has not been validated yet.

Besides, there are some other aspects for estimating the TOMD which may be of our interest, for example, when searching for efficient and reliable estimators, explore different block

designs and their combinations using, e.g., generalized means, as illustrated by Vovk and Wang [1] in a statistically different but philosophically closely related context. Moreover, there are technique issues in this research such as exploring the roles of the functions T_θ and $T_0 = \log$ when defining estimators.

5.2 About the anomaly detection index I

In this research, we explore the performance of the index proposed by Gribkova and Zitikis [2] when the input-output system has stationary series inputs. Theoretically, we prove that the index I_n still converges to 0.5 if the system is out of order. Unlike statistical outliers, system anomalies such as cyber intrusions are very difficult to observe at a first glance. Hence it is not surprising that the index I_n is constructed based on continuously monitoring of the outputs. However, the earlier the anomalies such as cyber intrusions are detected, the less losses will occur. Thus, more theoretical properties of the index I_n need to be explored in order to set up a standard to identify whether anomalies exist in earlier stage. For example, a hypothesis-testing type standard may be established if the asymptotic distribution of I_n or B_n could be determined.

Another interesting aspect related to I_n is that if we could develop such kind of indices for more complex systems. Currently our research object is the index I_n for univariate inputs and univariate outputs. However, complex systems such as parallel computing systems usually have multivariate inputs and outputs with anomalies occurring in various forms. In this case, the major challenge might be about how the order of the inputs may be defined since there is no simple order statistics for multivariate random variables.

Bibliography

- [1] V. Vovk, R. Wang, Combining p -values via averaging, *Biometrika* 107 (4) (2020) 791–808. [106](#)
- [2] N. Gribkova, R. Zitikis, A user-friendly algorithm for detecting the influence of background risks on a model, *Risks* 6 (3) (2018) 100. [106](#)

Curriculum Vitae

Name: Ning Sun

Post-Secondary Education and Degrees: Western University
London, ON, Canada
2018 - 2022 Ph.D. in Statistics

Wuhan University
Wuhan, Hubei, China
2016 - 2019 Ph.D. in Economics

Western University
London, ON, Canada
2011 - 2012 M.Sc. in Statistics

Beihang University
Beijing, China
2009 - 2011 M.A. in Financial Engineering

Beihang University
Beijing, China
2005 - 2009 B.A. in Financial Engineering

Scholarship: Western Graduate Research Scholarships.
Western University
2019 - 2021

OGS.
Western University
2019 - 2021

Publications:

1. N. Sun, C. Yang, and R. Zitikis. Tail maximal dependence in bivariate models: estimation and applications. Under review by *Mathematical Methods of Statistics*.

2. N. Sun, C. Yang, and R. Zitikis. Detecting systematic anomalies affecting systems when inputs are stationary time series. *Applied Stochastic Models in Business and Industry*, 38(3): 512-544.
3. N. Sun, C. Yang, and R. Zitikis (2020). A statistical methodology for assessing the maximal strength of tail dependence. *ASTIN Bulletin: The Journal of the IAA*, 50(3): 799-825.