TECHNISCHE UNIVERSITÄT MÜNCHEN
FAKULTÄT FÜR MATHEMATIK
LEHRSTUHL FÜR FINANZMATHEMATIK

Selected topics in credit risk: Realistic modeling of correlations and new pricing approaches for credit products

Amelie Angelika Hüttner

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Noam Berger Steiger
Prüfer der Dissertation:	1.	Prof. Dr. Matthias Scherer
	2.	Prof. Dr. Steven Vanduffel Vrije Universiteit Brussel, Belgien (schriftliche Beurteilung)
	3.	Prof. Dr. Ralf Werner Universität Augsburg

Die Dissertation wurde am 18.04.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 02.07.2019 angenommen.

Abstract

This thesis focuses on several related topics from two areas in the field of credit risk. On the one hand, it is concerned with the simulation and modeling of correlation matrices for financial applications, on the other hand, new pricing approaches for credit products are developed.

We develop an algorithm for the simulation of Perron–Frobenius correlation matrices, which additionally allows to control the distribution of the eigenvalues of the generated matrices. The employed construction principle enables us to prove that the proportion of Perron–Frobenius correlation matrices in the set of all correlation matrices is $1/2^{d-1}$ in dimension d. The generated matrices tend to exhibit (in large dimensions) all empirically observed stylized facts of financial correlation matrices, thus the algorithm is a valuable tool for risk management or for the assessment of portfolio selection strategies. In a use case, we examine the persistent empirical relation between graph-based portfolio selection techniques and the classical Markowitz minimum variance portfolio, showing that this relation is due to the special structure of financial correlation matrices, not due to a fundamental link between the two concepts.
Subsequently, we present a new approach for the joint modeling of financial assets adapted from geostatistics. The central idea is the joint modeling of the data as a Gaussian random field, and describing the dependence structure via a covariance resp. correlation function depending on the distance between observations. The key benefit of this ansatz is the possibility to easily include new data points (i.e. firms), which has appealing benefits for covariance resp. correlation matrix estimation and missing data imputation.

In the second part of this thesis, we first focus on the valuation of CDS options in a structural credit model based on a jump-diffusion process. Compared to existing approaches, this model has the advantage of providing realistic dynamics for the CDS spread. CDS options and other European optionalities like extension risk are priced in an efficient Monte Carlo scheme based on Brownian bridges.
Finally, sharp analytical lower bounds for the price of a convertible bond are derived in two defaultable Markov diffusion models. The central idea is to Europeanize the American conversion right, leading to analytical solutions and consequently to a large computational gain. The sharpness of the lower bound is demonstrated in two real-world examples.

Zusammenfassung

Die vorliegende Doktorarbeit beschäftigt sich mit verschiedenen Themen der Kreditrisikomodellierung, zum einen mit der Simulation und Modellierung von realistischen Korrelationsmatrizen für Finanzanwendungen, zum anderen mit neuen Bewertungsansätzen für zwei kreditrisikobehaftete Finanzprodukte, nämlich CDS-Optionen und Wandelanleihen.

Es wird ein Algorithmus für die Simulation von Perron–Frobenius-Korrelationsmatrizen entwickelt, welcher es zudem erlaubt, die Verteilung der Eigenwerte der Matrizen vorzugeben. Das verwendete Konstruktionsprinzip ermöglicht es, den Anteil an Perron–Frobenius-Korrelationsmatrizen in beliebiger Dimension explizit zu bestimmen. Die erzeugten Matrizen weisen (in großen Dimensionen) näherungsweise alle empirisch beobachteten Eigenschaften von Finanz-Korrelationsmatrizen auf, wodurch sich der Algorithmus ausgezeichnet für Aufgaben im Risikomanagement oder für die Beurteilung von Investmentstrategien eignet. In einem Anwendungsbeispiel untersuchen wir graphenbasierte Investmentstrategien und zeigen, dass der wiederholt empirisch beobachtete Zusammenhang zwischen diesen und dem klassischen Minimum-Varianz-Portfolio nach Markowitz nicht fundamental, sondern in der speziellen Struktur von Finanz-Korrelationsmatrizen begründet ist.
Anschließend beschäftigen wir uns mit einem neuen, ursprünglich aus der Geostatistik stammenden Ansatz für die gemeinsame Modellierung von Finanzdaten, welcher auf der gemeinsamen Modellierung der Daten als Gauß'sches Zufallsfeld basiert, wobei die Abhängigkeitsstruktur durch eine abstandsabhängige Kovarianz- bzw. Korrelationsfunktion beschrieben wird. Er erlaubt es, in einfacher Weise weitere Datenpunkte (d.h. Firmen) in gegebene Analysen einzubeziehen, was in verschiedenen Anwendungen, wie etwa beim Schätzen von Kovarianz- bzw. Korrelationsmatrizen und beim Einfügen fehlender Daten in einem Datensatz, von Vorteil ist.

Im zweiten Teil der Arbeit befassen wir uns zunächst mit der Bewertung von CDS-Optionen in einem auf einem Sprung-Diffusionsprozess basierenden strukturellen Kreditrisikomodell, welches gegenüber bisherigen Ansätzen den Vorteil hat, dass es realistische Pfade für CDS-Spreads erzeugt. Die Bewertung von CDS-Optionen sowie anderer europäscher Optionalitäten wie dem 'Extension Risk' erfolgt über einen effizienten, auf Brown'schen Brücken basierenden Monte-Carlo-Ansatz.
Schlussendlich werden analytische Formeln für eine untere Schranke des Preises einer Wandelanleihe in zwei Kreditrisikomodellen hergeleitet. Hierbei wird das typischerweise amerikanische Recht, die Anleihe in Aktien umzutauschen, zu einem europäischen Ausübungsrecht beschränkt, was die Berechnungszeiten für Preise erheblich verkürzt. Anhand zweier Beispiele wird demonstriert, dass die gegebenen Formeln unter realen Marktbedingungen eine sehr gute Annäherung an den exakten Preis liefern.

Acknowledgements

First and foremost, I thank my supervisor Matthias Scherer for the opportunity to pursue a PhD in mathematical finance and for the freedom to work on several interesting topics. Throughout my PhD he has always been very supportive in all matters concerning research, conference visits, and teaching.

I am equally grateful to my mentor Jan-Frederik Mai, who sparked my interest in new fascinating research directions and supported me throughout my PhD with his valuable advice and prompt and helpful feedback on a wide variety of topics.

Moreover, I'd like to thank my further coauthors Benedikt Gräler and Stefano Mineo for the fruitful collaboration.

I'd like to thank Rudi Zagst and Matthias Scherer for providing me with a teaching position. I'd also like to thank all my colleagues, past and present, at the chair, but also our neighboring chairs, for making work at the chair very enjoyable, and for lots of interesting discussions on math and many things more.

Equal thanks are due to Wolfgang Klopfer, Jochen Felsenheimer and Ulrich von Altenstadt at XAIA Investment for providing me with the opportunity to work part-time in the financial industry alongside doing my PhD, which has resulted in a most fruitful exchange between theory and practice which has greatly benefited my academic work.

Last, but definitely not least, I'd like to thank my family and friends who have supported me in many ways in these last years. My deepest gratitude goes to my parents and Ludwig, who are always there for me, come hell or high water, for their loving and encouraging support.

Contents

1 Introduction

Credit risk modeling is a very diverse area in mathematical finance. This field of study encompasses a wide variety of problems, ranging from the pricing of credit-risky products referencing on single entities or on baskets of obligors, to the modeling of joint defaults in baskets, as well as to models capturing a realistic evolution of credit spreads, to just name the most prominent ones. Given this variety of modeling tasks, credit risk is often split up into several aspects, cf. Schönbucher (2003), such as default risk, recovery risk (referring to the amount lost in case of default), default correlation risk (the risk of correlated defaults in a basket), or market risk (the risk of loss due to changing prices), for example.

Previously, especially following the credit crisis, the major focus has been on the modeling and assessment of default risk of single assets or baskets of assets, cf. e.g. Burtschell *et al.* (2009); Meissner (2008) and references therein. However, the correct assessment of market risk (or spread risk as referred to in Cont and Kan (2011)), i.e. the risk of a (joint) downturn of credit quality of the reference entities without any defaults, which has received comparably less attention in the context of credit derivatives, is just as important: In an illuminating example, Cont and Kan (2011) find that, while only 8 defaults occurred in the CDX index[1] in 2003-2011, a change in credit spread levels inducing the same loss as one default happens approximately twice per year. This shows that modeling the dependence in credit spread changes should receive a similar attention as modeling the dependence of defaults in a basket.

The present thesis revolves around several topics in the wide field of credit risk, and decomposes naturally into two parts, one focusing on problems related to spread risk, concretely the simulation and modeling of correlation matrices (Chapters 2 and 3), the other focusing on new pricing approaches for two products subject to default risk (Chapters 4 and 5).

Although the problems presented in the first part are transferable to other asset classes, the focus of all chapters is mainly on application to credit default swaps (CDS) or related instruments, except for Chapter 5, which is concerned with the valuation of convertible bonds. For the convenience of the reader, we give a brief introduction to CDS and the basic principles of their valuation in the following section.

[1] The CDX index contains the 125 most liquidly traded credit default swaps referencing on North-American investment grade firms.

1.1 Basics of CDS valuation

A CDS is essentially an insurance contract between two parties, the protection buyer and the protection seller, against a credit event of a reference entity. Depending on the specification of the contract, such a credit event may for example be the default of the reference entity, or a restructuring event. The mathematical modeling of CDS (and other assets subject to credit risk) is typically restricted to default being the only possible credit event. The protection buyer has to pay a standardized premium c up to maturity of the contract or default, whichever comes earlier, whereas the protection seller compensates the protection buyer for losses incurred in case of a default during the lifetime of the contract. Premium payments are usually made quarterly in arrears at contractually fixed dates, the premium being standardized to 100 bps or 500 bps. A schematic of the payment streams is displayed in Figure 1.1.

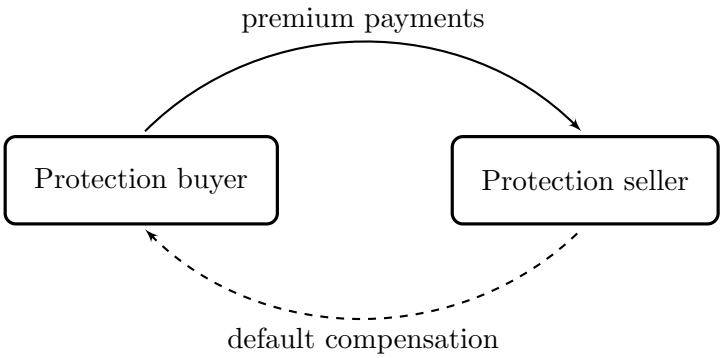

Figure 1.1: Illustration of CDS payment streams.

Denoting the expected discounted values of the payments made by the protection buyer and seller as the premium leg ($EDPL$) and the default leg ($EDDL$), respectively, today's value of a CDS with maturity T from the viewpoint of the protection buyer is then

$$CDS_{0,T}(c) = EDDL_{0,T} - c \cdot EDPL_{0,T},$$

with T denoting the maturity date of the contract. The value of the CDS at inception of the contract has to be paid upfront.
CDS are often quoted in terms of the so-called *running spread* $s_{0,T}$, which is the premium level c at which the contract has zero value at inception:

$$s_{0,T} = \frac{EDDL_{0,T}}{EDPL_{0,T}}.$$

Chapters 2 and 3 present applications focusing on CDS spread or upfront data, whereas Chapter 4 presents a new approach for valuing options to enter into a CDS with a prespecified spread at a future time point.

1.2 Presented results

The results presented in this thesis decompose into two parts: Part I, consisting of Chapters 2 and 3, is concerned with the realistic simulation and modeling of correlation matrices for financial applications. Part II, consisting of Chapters 4 and 5, is more practically oriented and focuses on two pricing problems in the field of credit risk.
The required mathematical concepts and considered financial products in the different chapters are quite diverse and are therefore introduced in the respective chapters in order to keep the presentation self-contained.

Chapter 2 builds upon joint work with Jan-Frederik Mai and Stefano Mineo, published in the articles Hüttner and Mai (2019) and Hüttner *et al.* (2018). We develop a simulation algorithm for correlation matrices with the Perron–Frobenius property, one of four stylized facts exhibited by financial correlation matrices. The algorithm is able to generate all such matrices, and from its construction principle the volume of the set of Perron–Frobenius correlation matrices can be deduced. Our algorithm further allows to specify the distribution of eigenvalues, hence the presence of another stylized fact of financial correlation matrices, namely a largest eigenvalue explaining more than 30% of the total variance, can be explicitly controlled for. Specifying a realistic distribution for the eigenvalues, also the two remaining stylized facts, a distribution of pairwise correlations that is significantly shifted to the positive and the scale-free property of the associated minimum spanning tree, tend to be present in large correlation matrices drawn from our algorithm. Consequently, the presented algorithm is a powerful tool in applications where a broad class of correlation matrices possibly associated with financial time series has to be simulated, e.g. when backtesting portfolio selection strategies. In a use case, the presented algorithm is employed, together with other simulation algorithms, to assess graph-based portfolio selection strategies: Several empirical studies have documented a strong relation between the classical Markowitz approach and centrality in a graph deduced from the correlation matrix. We show that there is no fundamental connection between the two portfolio selection approaches. Instead, using different simulation algorithms for correlation matrices with certain attributes, we show that this persistent observation originates in the special structure of financial correlation matrices.

Chapter 3 is based on a joint project with Matthias Scherer and Benedikt Gräler, cf. Hüttner *et al.* (2019), and is concerned with the joint modeling of dependent credit spreads using an approach adapted from geostatistics. Under the assumption that observations are jointly distributed as a Gaussian random field, which is completely characterized by its mean and covariance function, a modeling approach for realistic covariance resp. correlation matrices is at the core of the presented ansatz: Correlation matrices are constructed from distance-parameterized correlation functions, following Tobler's first law of geography, stating that observations made at closer locations are more related than observations made at locations further apart. As soon as the distance of a new location to the sample locations is known, it can be incorporated easily into

existing analyses without a possibly costly re-estimation of the model. This has appealing implications for several applications, two of which are showcased explicitly, namely estimation and parameterization of large correlation matrices and imputation of missing data.
Translating this approach traditionally applied in a two- or three-dimensional coordinate system to the higher-dimensional framework necessary for the modeling of financial data entails several challenges: Obviously, the key question is the design of a meaningful financial distance measure, for which we propose two general approaches. Further, we thoroughly discuss the question of model validity in higher-dimensional frameworks, which is related not only to the choice of covariance function, but also to the metric used in the financial distance, and to potentially necessary preliminary transformations of the coordinate space.

In the practically oriented part, Chapter 4 introduces a new valuation approach for single-name CDS options, developed in joint work with Matthias Scherer and published in Hüttner and Scherer (2016). CDS options allow their holder to buy or sell a CDS of a given reference entity at a prespecified spread at some future time point. Existing methods primarily focus on obtaining Black-type formulas for the price, similarly as for equity options, under the unrealistic assumption of lognormally distributed CDS spreads, or rely on models which produce somewhat unrealistic paths for the evolving CDS spreads. We propose the use of a structural credit risk model whose driving firm value process is a double-exponential jump diffusion for the modeling of CDS options. It supplies realistic CDS spread paths and admits the use of an efficient Monte Carlo approach based on Brownian bridges for CDS options and other European optionalities on (any analytic functional of) the firm value.

Finally, Chapter 5 presents analytical formulas for lower bounds for the price of a convertible bond in selected defaultable Markov diffusion models. A convertible bond is a regular bond which grants its holder the option to exchange the bond for a prespecified number of shares in a given time period. Additionally, a so-called soft call right, which de facto corresponds to enforcing conversion into shares if the stock price exceeds a certain level in a given time period, is often present. Due to the involved American optionalities, typical pricing methods are quite time-consuming PDE or tree schemes. Based on Europeanizing the American conversion option, analytical formulas for a lower bound are derived in two models, which turn out to be quite sharp in most realistic market circumstances, and constitute a considerable improvement in computation time. This makes the presented formulas highly attractive for convertible bond investors that screen a large number of convertible bonds for their investment decisions. This chapter builds upon and extends a joint project with Jan-Frederik Mai published in Hüttner and Mai (2018).

Part I

Realistic simulation and modeling of correlations for financial applications

2 Simulating realistic correlation matrices

2.1 Motivation

The simulation of realistic correlation matrices is important in several financial applications, for example for backtesting portfolio selection strategies or in risk management. However, when compared to purely random correlation matrices, those obtained from financial data sets tend to exhibit a very special structure. Various empirical studies have documented a battery of stylized facts of empirical correlation matrices from financial data sets:

(S1) **Large first eigenvalue:**
Considering the spectrum, random matrix theory predicts a certain range $[\lambda^-, \lambda^+]$ and density f_λ for the eigenvalues of a random correlation matrix constructed from data matrices with iid entries[1], which is usually violated by the eigenvalues of market correlation matrices. The largest empirical eigenvalue lies well above the theoretical upper bound λ^+, cf. Bouchaud and Potters (2011); Laloux *et al.* (1999); Plerou *et al.* (1999, 2002); Bun *et al.* (2017), and typically explains more than 30% of the total variance.

(S2) **Perron–Frobenius property:**
The Perron–Frobenius Theorem states that entrywise positive matrices $A \in \mathbb{R}_+^{n\times n}$ have a unique real largest (in absolute value) eigenvalue $\lambda_1 > 0$ of multiplicity 1, whose associated eigenvector $\mathbf{v}_1$ has all entries positive. Moreover, there are no other eigenvectors of A with this property (except positive multiples of $\mathbf{v}_1$). There exist matrices, although not entrywise positive, that also exhibit these properties. Research in linear algebra, cf., e.g., Johnson and Tarazaga (2004); Tarazaga *et al.* (2001), is concerned with identifying the set of all matrices displaying the properties stated in the Perron–Frobenius Theorem; however, so far only sufficient conditions have been found.
We consider a relaxed version of the Perron–Frobenius Theorem for non-negative matrices without additional structure. These matrices have a real largest eigenvalue $\lambda_1 \geq 0$ (not necessarily unique in absolute value and not necessarily simple) and a corresponding eigenvector $\mathbf{v}_1$ with only non-negative entries and at least

[1] These bounds and the density are dependent on the ratio of the number of simulated time series to their length.

one positive component (other non-negative eigenvectors may exist).[2]
Correlation matrices always have a real largest eigenvalue $\lambda_1 \geq 1$ (which is a.s. unique). Further, our results in Section 2.3.2 imply that the sets of correlation matrices having the Perron–Frobenius property and the strong Perron–Frobenius property only differ by a set of measure zero. Thus, when stating that a correlation matrix displays the (strong) Perron–Frobenius property, we refer to the property of having a dominant eigenvector with all entries non-negative (positive).
Boyle *et al.* (2014) observe in a data set of S&P1500 stocks that the major percentage of market correlation matrices exhibits a dominant eigenvector with only positive entries, and this percentage has been increasing up to 100% in their considered time period from 1994 to 2013.

(S3) **Distribution of pairwise correlations is shifted to the positive:**
The pairwise correlation entries of market correlation matrices typically display a smooth, unimodal distribution with a positive mean, with almost no mass on large negative correlations, e.g. Kazakov and Kalyagin (2016) find that the average of pairwise correlations is around 0.3 in several stock markets in the period 2003-2014, and Plerou *et al.* (2002) find that the distribution of pairwise correlations of US stocks is centered around a positive value in different time intervals in 1962-1996, illustrating the persistence of this stylized fact.

(S4) **Scale-free property of the corresponding minimum spanning tree:**
A $d \times d$ correlation matrix C can be viewed as the adjacency matrix of a weighted undirected complete graph $G_w(C)$ on d vertices. The pairwise correlation ρ_{ij} is translated to the length (weight) of the edge connecting nodes i and j in the graph[3] using a weight function w, which is strictly monotone and usually decreasing.[4] Thus the higher the correlation between i and j, the closer the respective nodes are in the graph. The information in this complete graph is then reduced to its *minimum spanning tree* ($\mathrm{MST}(C)$), cf. Figure 2.1: This is the connected subgraph on all vertices of the original graph without cycles having the minimal overall length (weight). The MST is unique if all edge weights of the original graph are different, i.e. if all entries of the correlation matrix are different, cf. Matoušek and Nešetřil (2007, Chapter 5.4, Ex. 4), which is typically the case for observed financial correlation matrices.
Vandewalle *et al.* (2001); Bonanno *et al.* (2003) investigate MSTs constructed on financial correlation matrices, and find that these MSTs exhibit the special structure of a so-called scale-free graph: This scale-free property means that the *degree distribution* of the MST, i.e. the distribution of the number of direct neighbors of

[2] For irreducible non-negative matrices the sharper statements of the original Perron–Frobenius Theorem hold.

[3] Diagonal entries imply self-loops in $G_w(C)$, which are not meaningful in most applications and hence ignored.

[4] A popular weight function is the so-called *correlation distance* $w(x) = \sqrt{2(1-x)}$ proposed by Mantegna (1999). Sometimes correlations are directly used as weights, i.e. $w(x) = x$, cf. Peralta and Zareei (2016).

the nodes in the graph, follows a power law[5]. Graphically speaking, this results in a higher probability of observing nodes with many neighbors (i.e. a high degree) than in 'random' graphs, cf. Vandewalle *et al.* (2001), and consequently, as the number of edges in a tree is fixed, in a higher number of leaves (vertices with only one neighbor) in the MST.[6]

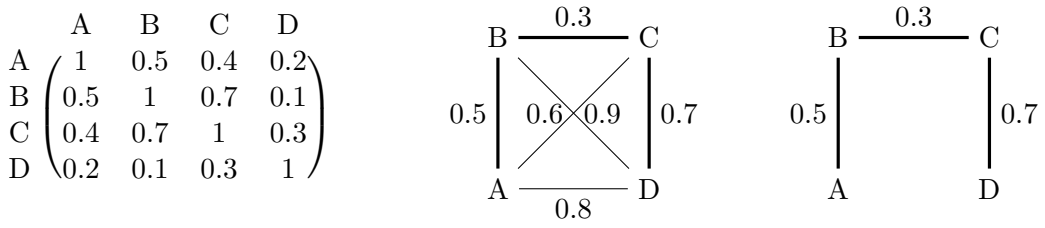

Figure 2.1: Procedure for obtaining the MST of a correlation matrix, illustrated here using the weight function $w(x) = 1 - x$.

The literature so far has focused primarily on stock data. In a data set of historical CDS data of 395 firms, we found that the correlation matrices of randomly drawn CDS portfolios with $d \in \{5, 10, 20, 50\}$ constituents also display these stylized facts, cf. Hüttner *et al.* (2018): The first eigenvalue explains about 40% of the variance: 47% in $d = 5$, declining in d to 37% in $d = 50$, for $n = 1,000,000$ randomly drawn portfolios of size d. More than 99.9% of correlation matrices of portfolios with $d = 20$ constituents ($n = 1,000,000$ draws) exhibited the Perron–Frobenius property. Further, the mean of pairwise correlations was always positive, and on average equal to 0.33, for $n = 1,000,000$ randomly drawn portfolios of different sizes from $d = 5$ to $d = 50$ constituents. Figure 2.2 contrasts the histogram of pairwise correlations of all 395 assets to that of a random correlation matrix of the same size drawn from the uniform distribution on the set of correlation matrices as introduced in Section 2.2.1. And finally, comparing the number of leaves encountered in $n = 1,000,000$ simulations of MSTs from random correlation matrices drawn from the uniform distribution with the number of leaves encountered in the same number of MSTs from correlation matrices of 20 randomly drawn assets from our data pool, we find that MSTs based on market data exhibit in general a higher number of leaves, cf. Figure 2.2.

To illustrate just how special financial correlation matrices are, we simulate $d \times d$ correlation matrices from the uniform distribution on the set of correlation matrices for various d, cf. Section 2.2.1, and find that the percentage of correlation matrices displaying stylized facts (S1) or (S2) or both quickly vanishes with increasing dimension, cf. Table 2.1.

[5]This means that the number of nodes having k neighbors is proportional to k^E, where E is the exponent of the power law.

[6]Barabási and Albert (1999) argue that this behaviour originates in a growing network with preferential attachment, i.e. new nodes are more likely to attach to highly connected nodes in the existing network, an interpretation that seems intuitive in a financial context. (Note that there exist other generating mechanisms that may result in a scale-free network.)

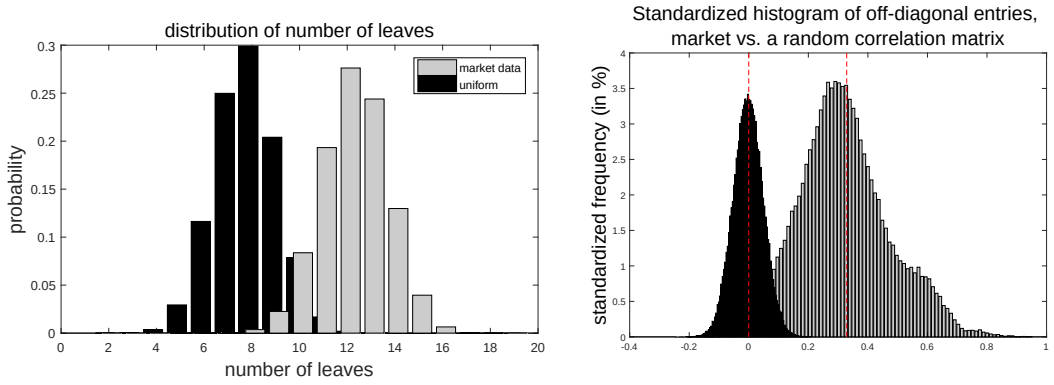

Figure 2.2: Left: Number of leaves encountered in uniformly simulated (black) vs. CDS data-based (gray) MSTs. Right: Standardized frequencies of the off-diagonal elements of a uniformly simulated (black) and our market correlation matrix (gray). The latter is significantly shifted to the right.

	$d=3$	$d=4$	$d=5$	$d=7$	$d=10$	$d=15$
% displaying 1.	100	99.99	99.91	93.26	26.69	0.01
% displaying 2.	24.98	12.47	6.22	1.56	0.20	0.01
% displaying 1. and 2.	24.98	12.47	6.21	1.46	0.05	0

Table 2.1: Simulation of $n = 1,000,000$ correlation matrices from the uniform distribution: Percentage of correlation matrices displaying stylized facts (S1) (as a proxy, we check if the first eigenvalue explains at least 30% of total variance) and (S2) declines fast with increasing dimension d.

2.1.1 Open research problem: Simulation of realistic correlation matrices for financial applications

Surprisingly, to the best of our knowledge, to date there exists no simulation algorithm that is able to reproduce all of the stylized facts (S1)-(S4), or even more than just one. In Section 2.2, we review known approaches for generating correlation matrices, and find that there is only one algorithm which is able to control for the presence of one of the stylized facts, namely (S1). For all other algorithms, too little is known about the distribution of the generated correlation matrices to control for the presence of any the stylized facts (S1)-(S4).

To overcome this issue, one often opts for simulating correlation matrices from a factor model calibrated to a specific market, which will of course exhibit very similar characteristics as the correlation matrix of the market the model was calibrated to. However, correlation matrices of other markets might display characteristics that are similar in their general nature, but differ in size, like displaying a large first eigenvalue compared to the rest of the spectrum, but explaining a different fraction of the total variance, or the distribution of pairwise correlations being shifted to the positive, but centered

around a different value. Consequently, such a procedure generates only a very narrow subclass of realistic correlation matrices.
Some use cases, however, require to simulate randomly from the set of all correlation matrices which might potentially be associated with some financial time series. This is especially the case in risk management, when one seeks to take into account model risk by assuming that the currently prevailing market conditions might change. For these purposes, a simulation algorithm that is able to produce a broad class of realistic correlation matrices is desirable.

Another use case that illustrates the importance of taking the special properties of financial correlation matrices into account is the assessment of graph-based portfolio selection techniques. Pioneered by Mantegna (1999), the application of graph-based approaches has become increasingly popular in the finance literature, and Onnela *et al.* (2003); Pozzi *et al.* (2013); Kaya (2015); Peralta and Zareei (2016) explore their usefulness for optimal investment purposes. The basic idea is to select assets according to their centrality in a graph[7] deduced from the covariance resp. correlation matrix, and the above-mentioned studies have documented a strong connection between an asset's centrality in the covariance-deduced graph and its weight in the corresponding minimum variance portfolio (MVP) of classical Markowitz theory for stock (return) data. This might lure investors into believing that there is a fundamental connection between these two concepts. However, we demonstrate in Section 2.4 that this suspicion cannot be confirmed, and the coherence of these findings must instead originate in the special properties of financial correlation matrices.

In a joint project with Jan-Frederik Mai, published in the article Hüttner and Mai (2019), we developed a simulation algorithm that generates correlation matrices with the Perron–Frobenius property (S2). It additionally allows for a specification of the eigenvalue distribution, hence allowing to control for the presence of (S1). We prove that our algorithm is able to generate all correlation matrices with the Perron–Frobenius property, and the construction principle applied in the algorithm enables us to prove that $1/2^{d-1}$ of all $d \times d$ correlation matrices exhibit this property. On the other hand, the construction principle makes it hard to derive properties of the distribution of the generated correlation matrices, or to adjust the algorithm for taking into account also (S3) and (S4). Nevertheless, depending on the chosen eigenvalue distribution, a major percentage of the correlation matrices generated from the presented algorithm exhibits also (S3), a realistic distribution of pairwise correlations, and, for large correlation matrices ($d > 1000$), also (S4), the power-law-like degree distribution of the MST, tends to be fulfilled.

With this simulation algorithm for realistic correlation matrices at hand, we are able to assess graph-based portfolio selection techniques and analyze which specific features of empirical correlation matrices may cause the persistent observed relation between graph centrality and MVP weights, hereby extending our earlier study based on joint

[7] This is typically the complete weighted undirected graph corresponding to the matrix or its MST.

work with Jan-Frederik Mai and Stefano Mineo, published in the article Hüttner *et al.* (2018).

Consequently, certain parts of this chapter will exhibit considerable conformity with the above references.

Throughout this chapter, we will use the following notation for (the sets of) covariance resp. correlation matrices:

$$\begin{aligned}
\mathcal{K}_d &:= \left\{\Sigma \in \mathbb{R}^{d\times d} \,:\, \Sigma \text{ symmetric and non-negative definite}\right\} \\
&\quad = \text{set of covariance matrices in dimension } d, \\
\mathcal{C}_d &:= \left\{C \in \mathcal{K}_d \cap [-1,1]^{d\times d} \,:\, C_{i,i} = 1 \,\forall\, i \in \{1,\dots,d\}\right\} \\
&\quad = \text{set of correlation matrices in dimension } d, \\
\mathcal{C}_d^{PF} &:= \text{set of correlation matrices in dimension } d \text{ exhibiting the Perron–Frobenius} \\
&\qquad \text{property}, \\
\mathcal{C}_{d,\lambda} &:= \text{set of correlation matrices in dimension } d \text{ with eigenvalues } \lambda.
\end{aligned}$$

It holds that $\mathcal{C}_d \subset \mathcal{K}_d$, i.e. every correlation matrix is a covariance matrix. The manifold $\mathcal{C}_d$ may be identified with a compact subset of $\mathbb{R}^{d(d-1)/2}$, e.g. via the pairwise correlations, while $\mathcal{K}_d$ may be identified with an unbounded subset of $\mathbb{R}^{d(d+1)/2}$. For every $\Sigma \in \mathcal{K}_d$ there is a unique $C \in \mathcal{C}_d$ such that

$$\Sigma = \operatorname{diag}(\sqrt{\Sigma_{ii}})\, C \operatorname{diag}(\sqrt{\Sigma_{ii}}),$$

$$\operatorname{diag}(\sqrt{\Sigma_{ii}}) = \begin{pmatrix} \sqrt{\Sigma_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{\Sigma_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\Sigma_{dd}} \end{pmatrix}.$$

The remainder of this chapter is organized as follows: Section 2.2 reviews known approaches for the generation of correlation matrices. Our algorithm for simulating Perron–Frobenius correlation matrices is presented in Section 2.3.2, where we further prove related statements and analyze the simulated matrices with respect to the presence of stylized facts (S3) and (S4). Section 2.4 studies graph-based portfolio selection techniques and, by applying the techniques to simulated correlation matrices exhibiting different subsets of the stylized facts (S1)-(S4), demonstrates that there is no fundamental connection between graph-centrality and MVP-weights. Instead, the persistent empirical findings in favor of such a relation must originate from the special properties of financial correlation matrices. Section 2.5 summarizes and concludes.

2.2 Previous approaches to simulating correlation matrices

Ideally, an algorithm for simulating realistic correlation matrices for financial applications would be able to reproduce all of the above-mentioned stylized facts. However,

the choice of simulation algorithms is limited, and narrows even further if one intends to generate correlation matrices with specific features. In the following, we review the currently available simulation procedures, and discuss whether they are able to reproduce the stylized facts (S1)-(S4) or might be modified accordingly.

2.2.1 Simulating from the uniform distribution on $\mathcal{C}_d$

Joe (2006), Lewandowski *et al.* (2009), and Ghosh and Henderson (2003) propose different approaches for simulating uniformly from the set $\mathcal{C}_d$ of $d \times d$ correlation matrices:

- Joe (2006) proposes to parameterize $C \in \mathcal{C}_d$ via $d(d-1)/2$ *partial correlations* $\rho_{i,j|L}$, with $\rho_{i,j|L}$ being the correlation between variables i and j, with the effects of variables contained in the set $L = \{i+1, \ldots, j-1\}$ removed. Generating the partial correlations $\rho_{i,i+k|i+1,\ldots,i+k-1}$ independently from $\text{Beta}((d+1-k)/2, (d+1-k)/2)$ distributions on $(-1,1)$, $k \in \{1, \ldots, d-2\}$ and transforming to pairwise correlations yields a correlation matrix C drawn from the uniform distribution on $\mathcal{C}_d$, $C \sim \mathcal{U}(\mathcal{C}_d)$. Lewandowski *et al.* (2009) extend this to different sets of $d(d-1)/2$ partial correlations.
- Ghosh and Henderson (2003) follow a different approach to obtain $C \sim \mathcal{U}(\mathcal{C}_d)$, which the call the *onion method*: The core principle is to extend a $(d-1) \times (d-1)$ correlation matrix to a $d \times d$ correlation matrix by adding a suitable row / column vector $(\mathbf{q}, 1) \in \mathbb{R}^d$. This procedure is linked to elliptical distributions, cf. also Lewandowski *et al.* (2009), as $\mathbf{q} = \tilde{C}^{\frac{1}{2}} \cdot \sqrt{b} \cdot U$, where $\tilde{C}$ is the $(d-1) \times (d-1)$ correlation matrix to be extended, b is a $\text{Beta}((d-1)/2, (d-k)/2)$-distributed random variable, and U is a random variable drawn from the uniform distribution on the unit sphere in dimension $d-1$.

Simulating uniformly from the set of correlation matrices in dimension d produces correlation matrices with no further inherent structure other than positive (semi-)definiteness. Nothing is known about the distribution of eigenvalues of such matrices, and Table 2.1 shows that with increasing dimension it is highly unlikely that matrices generated from this distribution exhibit the Perron–Frobenius property. Via the ansatz of Joe (2006), one obtains that pairwise correlations are identically $\text{Beta}(d/2, d/2)$-distributed on $(-1,1)$, thus $C \sim \mathcal{U}(\mathcal{C}_d)$ do not exhibit (S3). Finally, Figure 2.2 indicates that they also do not exhibit (S4).

2.2.2 Simulating correlation matrices with given eigenvalues: the `randcorr` algorithm

This algorithm, introduced in Bendel and Mickey (1978), subsequently improved by Davies and Higham (2000) and implemented as `gallery('randcorr',...)` in Matlab, is sometimes also referred to as Bendel–Mickey algorithm. The central idea is to start

from a diagonal matrix Λ of eigenvalues, and then subsequently apply Givens rotations that force a diagonal element to become one:

$$C = G_{d-1} \cdot \ldots \cdot G_1 \Lambda G_1' \cdot \ldots \cdot G_{d-1}',$$

$$G_\theta^{(i,j)} = \begin{pmatrix} I_{i-1} & 0 & \ldots & \ldots & 0 \\ 0 & \cos(\theta) & 0 & -\sin(\theta) & 0 \\ \vdots & 0 & I_{j-i-1} & 0 & \vdots \\ \vdots & \sin(\theta) & 0 & \cos(\theta) & 0 \\ 0 & \ldots & \ldots & 0 & I_{d-j} \end{pmatrix} \tag{2.1}$$

where I_n is the n-dimensional identity matrix, and $G_k := G_{\theta_k}^{(i_k,j_k)}$ is a Givens rotation with rotation angle θ_k in the (i_k, j_k)-plane that forces the i_k-th diagonal entry of the resulting matrix $G_k \cdots G_1 \Lambda G_1' \cdots G_k'$ to become 1. These operations step by step introduce the unit diagonal without altering the trace and the eigenvalues, i.e. transform Λ into a correlation matrix. As this might produce correlation matrices where some entries are zero according to Davies and Higham (2000), and this typically never happens in the market (or a.s. never happens when drawing from the uniform distribution), Λ is first multiplied with a random orthogonal matrix $Q \in \mathcal{O}(d)$ drawn from the Haar distribution on the orthogonal group $\mathcal{O}(d)$, cf. Stewart (1980) for a simulation algorithm, i.e.

$$C = G_{d-1} \cdot \ldots \cdot G_1 Q \Lambda Q' G_1' \cdot \ldots \cdot G_{d-1}' = V \Lambda V'.$$

With the eigenvalue decomposition of C in mind, the eigenvector matrix V is thus constructed as the product of a random orthogonal matrix and $d-1$ Givens rotations. Davies and Higham (2000) identify the following efficient and numerically stable approach to solve for $c_k := \cos(\theta_k)$, $s_k := \sin(\theta_k)$, denoting

$$W^{(k-1)} := G_{k-1} \cdots G_1 Q \Lambda Q' G_1' \cdots G_{k-1}',$$
$$V^{(k-1)} := G_{k-1} \cdots G_1 Q :$$

The rotation indices $i_k < j_k$ of G_k are chosen such that[8] either $w_{i_k,i_k}^{(k-1)} < 1 < w_{j_k,j_k}^{(k-1)}$ or $w_{j_k,j_k}^{(k-1)} < 1 < w_{i_k,i_k}^{(k-1)}$, and should force the i_k-th diagonal entry of the resulting matrix $W^{(k)}$ to be 1, i.e. the equation

$$c_k^2 w_{i_k,i_k}^{(k-1)} - 2 c_k s_k w_{i_k,j_k}^{(k-1)} + s_k^2 w_{j_k,j_k}^{(k-1)} = 1$$

must hold. This can be transformed into a quadratic equation in the tangent, from which they recover

$$\begin{aligned} &(w_{j_k,j_k}^{(k-1)} - 1) t_k^2 - 2 t_k w_{i_k,j_k}^{(k-1)} + w_{i_k,i_k}^{(k-1)} - 1 = 0, \\ &\tan(\theta_k) = t_k = \frac{w_{i_k,j_k}^{(k-1)} \pm \sqrt{(w_{i_k,j_k}^{(k-1)})^2 - (w_{i_k,i_k}^{(k-1)} - 1)(w_{j_k,j_k}^{(k-1)} - 1)}}{w_{j_k,j_k}^{(k-1)} - 1}, \\ &c_k = \pm (1 + t_k^2)^{-\frac{1}{2}}, \quad s_k = c_k t_k. \end{aligned} \tag{2.2}$$

[8] This is always possible as the trace equals d and does not change.

The conditions on the rotation indices ensure that t_k is well-defined. Further, since the value of the square root in the numerator is larger than $w_{i_k,j_k}^{(k-1)}$ in absolute value, one of the possible solutions for the tangent is positive, the other negative. We denote the two solutions by $t_k^+ > 0$ and $t_k^- < 0$ in the remainder of this chapter.

As the eigenvalues are specified prior to forming the correlation matrix, this algorithm deliberately allows to control for the presence of (S1). Apart from the distribution of eigenvalues and related quantities like the determinant as discussed in Holmes (1991), little is known about the distribution of the matrices generated by the `randcorr` algorithm, cf. Davies and Higham (2000). Consequently, it is very hard to derive statements about the distribution of pairwise correlations and the presence of stylized facts (S3) and (S4). Simulation studies showed that, even when specifying a realistic distribution for the eigenvalues, the generated correlation matrices exhibit only (S1), but not the other stylized facts, cf. Section 2.4.3 below.

However, from the `randcorr` algorithm, a simulation algorithm for correlation matrices with the Perron–Frobenius property (S2) can be developed:
Davies and Higham (2000) always opt for $c_k = +(1+t_k^2)^{-\frac{1}{2}}$ in (2.2) and do not acknowledge that $c_k = -(1+t_k^2)^{-\frac{1}{2}}$ is also a valid solution. Being able to choose the sign of the cosine here, however, gives us the necessary flexibility to adjust the algorithm for the generation of Perron–Frobenius correlation matrices. Further, they only state the tangent solutions if the i_k-th diagonal entry is forced to 1. If instead the j_k-th diagonal entry should be forced to 1, the tangent solutions become

$$t_k = \frac{-w_{i_k,j_k}^{(k-1)} \pm \sqrt{(w_{i_k,j_k}^{(k-1)})^2 - (w_{j_k,j_k}^{(k-1)} - 1)(w_{i_k,i_k}^{(k-1)} - 1)}}{w_{i_k,i_k}^{(k-1)} - 1}. \tag{2.3}$$

In the remainder of this chapter, we consider the `randcorr` algorithm to be augmented in this respect.

2.2.3 Random gram matrices

The basic idea of this ansatz discussed in Holmes (1991); Marsaglia and Olkin (1984) is to construct a correlation matrix $C = XX'$ from a data matrix X of linearly independent rows of norm 1. The distributions of eigenvalues and pairwise correlations are unknown, except for the case where the rows of X are distributed uniformly on the sphere, where Marsaglia and Olkin (1984) derive results on the distribution of pairwise correlations. In this case, the percentage of generated matrices with positive dominant eigenvector vanishes quickly with increasing dimension d, like for the uniform distribution. It is unclear, if, or under which conditions on the elements of X, this algorithm is able to generate Perron–Frobenius correlation matrices, and if this procedure can be modified accordingly.

A related algorithm for the generation of covariance matrices whose diagonal resp. off-diagonal elements follow distributions with prespecified moments is given in Hirschberger *et al.* (2007), which, however, cannot be modified readily to produce correlation matrices with pairwise correlation entries having specific moments due to the constraint that diagonal entries have to be 1. A further constraint one should impose in the latter method is that the generated correlation matrices should have full rank, like those observed in the market, which is a further restriction that hinders the translation of this approach to correlation matrices.

2.2.4 Perturbation about or adding noise to a given correlation matrix

Marsaglia and Olkin (1984) discuss how to generate correlation matrices perturbed about a given mean $C \in \mathcal{C}_d$. Hardin *et al.* (2013) provide an algorithm for adding noise to a given correlation matrix such that the resulting matrix is still positive definite.
In both approaches, the generated matrices are centered around a given correlation matrix. With a similar argument as for factor correlation matrices, correlation matrices generated like this constitute only a narrow subclass of realistic correlation matrices. No general statements can be made about the distribution of eigenvalues or pairwise correlations. Further, even when starting from a mean correlation matrix C that has the Perron–Frobenius property, due to the irregular shape of the set of Perron–Frobenius correlation matrices[9], we cannot expect that all matrices generated from these procedures will display a dominant eigenvector with only positive entries.

2.2.5 Factor model correlation matrices

As mentioned above, we take the view that the simulation of factor model correlation matrices as discussed e.g. in Fan *et al.* (2008); Cizeau *et al.* (2001) relies too strongly on the specific characteristics of the market being analyzed to be considered a completely random approach of sorts. The simulated matrices typically display (S1) and (S3), with the exact characteristics of course depending on the market the model was calibrated to. Correlation matrices generated from factor models may or may not display the Perron–Frobenius property: Of the two factor models considered in Section 2.4.3, one model yields correlation matrices with the Perron–Frobenius property, the other not. It is further worth noting that, with respect to stylized fact (S4), the scale-free property, factor models are likely not the way to go: Bonanno *et al.* (2003) observes that correlation matrices with a one-factor structure tend to exhibit tree structures that are a lot denser (more leaves, all nodes very close to the central one) than those obtained from market correlation matrices. Concerning multi-factor models, the analysis of simulated

[9] Boyle *et al.* (2014) show that it is not convex in $d > 3$. See also Johnson and Tarazaga (2004); Tarazaga *et al.* (2001) for results on the sets of (symmetric) square matrices exhibiting the Perron–Frobenius properties.

3-factor-model correlation matrices in Section 2.4.3 implies that similar tree structures are to be expected.

2.2.6 Generating correlation matrices exhibiting (S1)-(S4)

To summarize, only the `randcorr` algorithm explicitly allows to control the distribution of eigenvalues, hence the presence of (S1). The distribution of pairwise correlations in the presented algorithms is either unknown or does not display the properties specified in (S3). Concerning the generation of correlation matrices whose MSTs exhibit the scale-free property (S4), to the best of our knowledge there is no algorithm available, and due to the generating mechanism of the MST we expect the task of finding such correlation matrices to be highly complex.

For the generation of correlation matrices with the Perron–Frobenius property (S2), no algorithm is available so far, although basically all observed financial correlation matrices are of this type nowadays. To the best of our knowledge, Boyle *et al.* (2014); Boyle and N'Diaye (2018) are the only articles concerned with this subclass of correlation matrices: Apart from a broad analysis of market correlation matrices, they give an explicit characterization of 3×3 Perron–Frobenius correlation matrices[10], some related results in dimensions 4 and 5, and an intuition for our Theorem 2.3.9 on the proportion of Perron–Frobenius correlation matrices in arbitrary dimension.

Whereas Boyle *et al.* (2014) in their theoretical considerations start directly from an arbitrary correlation matrix, we find it more natural to approach the problem of generating Perron–Frobenius correlation matrices starting from the eigendecomposition of a correlation matrix:

$$C = V \Lambda V', \quad C \in \mathcal{C}_d,$$

where $\Lambda = \operatorname{diag}(\lambda)$ is the diagonal matrix of eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_d)$, $\lambda_1 \geq \ldots \geq \lambda_d$, and the columns of V contain the corresponding eigenvectors, i.e. $\mathbf{v}_1$, the first column of V, corresponds to the dominant eigenvector.

As we have noted in Section 2.2.2, the `randcorr` algorithm is an excellent starting point for devising a simulation procedure for Perron–Frobenius correlation matrices.

[10] They show that a necessary and sufficient condition in $d = 3$ for a correlation matrix to have a positive dominant eigenvector is that the sum of any two pairwise correlations is positive.

2.3 Simulating Perron–Frobenius correlation matrices

2.3.1 Auxiliary results on correlation matrices simulated from the `randcorr` algorithm

As stated above, not much is known about the distribution of matrices generated by the `randcorr` algorithm. However, we are able to prove the following two lemmas, about the support of the distribution of correlation matrices generated by the `randcorr` algorithm and about the occurrence of zero entries in their dominant eigenvector, which are required in our proof of the proportion of Perron–Frobenius correlation matrices in Theorem 2.3.9 below.

Lemma 2.3.1 (full support of `randcorr`)
For fixed eigenvalues λ, the support of the distribution of $d \times d$ correlation matrices generated by the `randcorr` algorithm is $\mathcal{C}_{d,\lambda}$.
The eigenvalues of any $C \in \mathcal{C}_d$ are all non-negative and sum up to d. Thus, drawing the eigenvalues λ from a distribution that has full support on the set of feasible eigenvalues for $C \in \mathcal{C}_d$, namely $d \cdot \mathcal{S}_d$, where $\mathcal{S}_d := \{(x_1, \ldots, x_d) \in [0,1]^d : \sum_{i=1}^d x_i = 1\}$ is the d-dimensional unit simplex, the support of the `randcorr` algorithm is $\mathcal{C}_d$.

Proof
This follows from the facts that

1) the Haar distribution, which is used to generate the random Q in the initializing step of the algorithm, has full support on the orthogonal group $\mathcal{O}(d)$.

2) $\mathcal{C}_{d,\lambda}$ is a subset of the set $\mathcal{K}_{d,\lambda}$ of covariance matrices with these eigenvalues λ.

Then, considering the function $f_\lambda : \mathcal{O}(d) \to \mathcal{K}_{d,\lambda}$, $Q \mapsto Q\Lambda Q'$, $\Lambda = \operatorname{diag}(\lambda)$, that maps an orthogonal matrix $Q \in \mathcal{O}(d)$ to a covariance matrix with eigenvalues λ, we find that due to 1) this has full support $\mathcal{K}_{d,\lambda}$, containing the set $\mathcal{C}_{d,\lambda}$, i.e. we can get every correlation matrix already in the first step of the algorithm when multiplying with a random orthogonal Q. Together with the full support of the eigenvalue distribution, the second statement follows immediately. □

Lemma 2.3.2 (a.s. no zero entries in $\mathbf{v_1}$)
For fixed eigenvalues λ, the dominant eigenvector of a correlation matrix C generated from the `randcorr` algorithm a.s. has no zero entries.

Proof
The `randcorr` algorithm corresponds to a sequence of functions

$$Q\Lambda Q' = W^{(0)} \mapsto G_1 Q\Lambda Q' G_1' = W^{(1)} \mapsto G_2 G_1 Q\Lambda Q' G_1' G_2' = W^{(2)} \mapsto \ldots \mapsto C = V\Lambda V',$$
$$Q = V^{(0)} \overset{f_1}{\mapsto} G_1 Q = V^{(1)} \overset{f_2}{\mapsto} G_2 G_1 Q = V^{(2)} \overset{f_3}{\mapsto} \ldots \overset{f_{d-1}}{\mapsto} G_{d-1} \cdots G_1 Q = V,$$

with $\Lambda = \operatorname{diag}(\lambda)$ fixed in advance.

1) From the eigendecomposition of a correlation matrix, we know:

$$V\Lambda V' = \begin{pmatrix} v_{1,1} & \dots & v_{1,n} \\ \vdots & \dots & \vdots \\ v_{n,1} & \dots & v_{n,n} \end{pmatrix} \begin{pmatrix} \lambda_1 & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & \lambda_n \end{pmatrix} \begin{pmatrix} v_{1,1} & \dots & v_{n,1} \\ \vdots & \vdots & \vdots \\ v_{1,n} & \dots & v_{n,n} \end{pmatrix}$$
$$= \begin{pmatrix} \sum_{i=1}^n \lambda_i v_{1,i}^2 & \sum_{i=1}^n \lambda_i v_{1,i} v_{2,i} & \dots \\ \sum_{i=1}^n \lambda_i v_{1,i} v_{2,i} & \sum_{i=1}^n \lambda_i v_{2,i}^2 & \dots \\ \vdots & \vdots & \vdots \\ \dots & \dots & \sum_{i=1}^n \lambda_i v_{n,i}^2 \end{pmatrix} = C.$$

To match the diagonal entries to 1, we need

$$\sum_{l=1}^d \lambda_l v_{i,l}^2 = 1, \quad \forall\, i = 1, \dots, d.$$

2) Wlog. assume that G_k sets the i_k-th diagonal entry of $W^{(k)}$ to 1. (Similar considerations apply if the j_k-th diagonal entry of $W^{(k)}$ is set to 1.) The i_k-th row of V is not modified by later Givens rotations G_l, $l > k$, once the i_k-th diagonal entry is rotated to 1, as i_k is not selected as a rotation index again, and the i_k-th row and column of G_l, $l > k$, equal the i_k-th unit vector. Thus, G_k enforces $\sum_{l=1}^d \lambda_l v_{i_k,l}^2 = 1$.

3) As stated above, any d-dimensional correlation matrix can be parameterized by (at least) $d(d-1)/2$ parameters, i.e. $\mathcal{C}_d$ is a $d(d-1)/2$-dimensional manifold, thus can be locally identified[11] with $\mathbb{R}^{d(d-1)/2}$. Fixing the eigenvalues fixes $d-1$ of these parameters, so the remaining $d(d-1)/2 - (d-1)$ can be attributed to V. Further, $\mathcal{O}(d)$ is a $d(d-1)/2$-dimensional manifold[12], cf. Absil *et al.* (2008, Section 3.3.2). Thus, for fixed eigenvalues it suffices to study how the functions f_k reduce the dimension of the set of feasible matrices $V^{(k)}$. To this end, define

$$\begin{aligned} \mathcal{V}_\lambda^{(k)} &:= \{V \in \mathcal{O}(d) : V = G_k \cdots G_1 Q \text{ generated by Algorithm 2.3.4 with fixed} \\ &\qquad \text{eigenvalues } \lambda\} \\ &\overset{2)}{=} \{V \in \mathcal{O}(d) : \sum_{l=1}^d \lambda_l v_{i,l}^2 = 1, i \in \{i_1, \dots, i_k\}\}, \\ \mathcal{V}_\lambda^{(0)} &:= \mathcal{O}(d), \end{aligned}$$

i.e. we have $f_k : \mathcal{V}_\lambda^{(k-1)} \to \mathcal{V}_\lambda^{(k)}$, $V^{(k-1)} \mapsto G_k(Q, \lambda) V^{(k-1)}$, highlighting the dependence of $G_k = G_k(Q, \lambda)$ on (Q, λ). (Remember G_k depends on the eigenvalues

[11] This is implicitly used in Joe (2006)'s derivation of the uniform distribution on $\mathcal{C}_d$, as sketched in Section 2.2.1.

[12] Graphically speaking, an orthogonal matrix $Q \in \mathcal{O}(d)$ can be parameterized in terms of $d(d-1)/2$ angles describing the directions of the d orthogonal column vectors.

λ and the random orthogonal matrix Q via t_k, cf. Equation (2.2).) From this definition, we see that $\mathcal{V}_\lambda^{(k)}$ is a $(d(d-1)/2-k)$-dimensional subset of $\mathcal{O}(d)$.

4) From the `randcorr` algorithm, we have

$$V = V^{(d-1)} = G_{d-1} \cdots G_1 Q \qquad \Leftrightarrow Q = V^{(0)} = G_1' \cdots G_{d-1}' V.$$

Now consider the set of feasible eigenvector matrices having a zero entry in the dominant eigenvector,

$$\begin{aligned}\mathcal{V}_{\lambda,0}^{(d-1)} &:= \{V \in \mathcal{V}_\lambda^{(d-1)} : v_{j,1} = 0 \text{ for at least one } j \in \{1,\dots,d\}\} \\ &= \Big\{V \in \mathcal{O}(d) : \sum_{l=1}^{d} \lambda_l v_{i,l}^2 = 1, i \in \{i_1,\dots,i_k\}, v_{j,1} = 0 \text{ for at least one} \\ &\qquad j \in \{1,\dots,d\}\Big\} \subset \mathcal{V}_\lambda^{(d-1)},\end{aligned}$$

which has dimension $\delta < d(d-1)/2 - (d-1)$. If $V = V^{(d-1)} \in \mathcal{V}_{\lambda,0}^{(d-1)}$, then $Q = V^{(0)}$ necessarily lies in the set

$$\begin{aligned}&\big\{G_{\phi_1}^{(m_1,n_1)} G_{\phi_2}^{(m_2,n_2)} \cdots G_{\phi_{d-1}}^{(m_{d-1},n_{d-1})} V : V \in \mathcal{V}_{\lambda,0}^{(d-1)}, \phi_i \in [-\pi,\pi], \\ &(m_i,n_i) \subset \{1,\dots,d\}, i = 1,\dots,d\big\} \subset \mathcal{O}(d),\end{aligned}$$

which is a true subset whose dimension is at most $\delta + (d-1) < d(d-1)/2$. This is a set of measure zero in $\mathcal{V}_\lambda^{(0)} = \mathcal{O}(d)$ with respect to the Haar measure. Thus, we a.s. never draw Q such that the algorithm produces a dominant eigenvector having a zero entry. □

Remark 2.3.3 (Generalization of Lemma 2.3.2)
With the same logic, for an arbitrary $\mathcal{V}_{\lambda,*}^{(d-1)} \subset \mathcal{V}_\lambda^{(d-1)}$ of dimension $\delta < d(d-1)/2-(d-1)$ it holds that for $V = V^{(d-1)} \in \mathcal{V}_{\lambda,*}^{(d-1)}$, Q was necessarily drawn from a set of measure zero in $\mathcal{O}(d)$. Via the eigendecomposition[13] of C, we get that the 'pre-image' of any lower-dimensional subset of $\mathcal{C}_{d,\lambda}$ under the `randcorr` algorithm has measure zero in $\mathcal{K}_{d,\lambda}$. Drawing λ from an absolutely continuous distribution with full support $\mathcal{S}_d$, this statement generalizes to lower-dimensional subsets of $\mathcal{C}_d$.

2.3.2 Simulation algorithm for Perron–Frobenius correlation matrices

In the following we present an algorithm that generates $C \in \mathcal{C}_d^{PF}$ for an arbitrary dimension d. It is adapted from the `randcorr` algorithm outlined in Section 2.2.2. The two key observations that ensure we can indeed generate Perron–Frobenius correlation matrices are the following (dropping the subscript k of the rotation indices for notational simplicity):

[13] The eigendecomposition of C is unique up to different signs of columns of V, thus for a fixed C there is only finitely many V such that $V\Lambda V' = C$. These V constitute a set of measure zero in $\mathcal{V}_\lambda^{(d-1)}$.

1. The i-th row of V is not modified by later Givens rotations once the i-th diagonal entry is rotated to 1, as i is not selected as a rotation index again in the `randcorr` algorithm, and the i-th row and column of later Givens rotations equal the i-th unit vector.

2. In every step, we can choose one of the four possible Givens rotations[14] such that the i-th (resp. j-th) entry of $\mathbf{v}_1$, the first column of V (i.e. the dominant eigenvector), is forced to have a given sign when the corresponding diagonal entry is set to 1. Thus, we can force all entries of the dominant eigenvector $\mathbf{v}_1$ to have the same sign, which is sufficient for the generation of PF correlation matrices as $-\mathbf{v}_1$, too, is an eigenvector of the largest eigenvalue.
   To this end, note that a multiplication with a Givens rotation $G_\theta^{(i,j)}$ from the left affects only the i-th and j-th row, and specifically in the following way:

$$\begin{aligned} G_k V^{(k-1)} &= \begin{pmatrix} I_{i_k-1} & 0 & \dots & \dots & 0 \\ 0 & c_k & 0 & -s_k & 0 \\ \vdots & 0 & I_{j_k-i_k-1} & 0 & \vdots \\ \vdots & s_k & 0 & c_k & 0 \\ 0 & \dots & \dots & 0 & I_{d-j_k} \end{pmatrix} \begin{pmatrix} \vdots & \dots & \vdots \\ v_{i_k,1}^{(k-1)} & \dots & v_{i_k,d}^{(k-1)} \\ \vdots & \dots & \vdots \\ v_{j_k,1}^{(k-1)} & \dots & v_{j_k,d}^{(k-1)} \\ \vdots & \dots & \vdots \end{pmatrix} \\ &= \begin{pmatrix} \vdots & \dots & \vdots \\ c_k v_{i_k,1}^{(k-1)} - s_k v_{j_k,1}^{(k-1)} & \dots & c_k v_{i_k,d}^{(k-1)} - s_k v_{j_k,d}^{(k-1)} \\ \vdots & \dots & \vdots \\ s_k v_{i_k,1}^{(k-1)} + c_k v_{j_k,1}^{(k-1)} & \dots & s_k v_{i_k,d}^{(k-1)} + c_k v_{j_k,d}^{(k-1)} \\ \vdots & \dots & \vdots \end{pmatrix}. \end{aligned} \tag{2.4}$$

Algorithm 2.3.4 (Simulation of Perron–Frobenius correlation matrices)
Input: dimension d
Output: $C \in \mathcal{C}_d^{PF}$

(1) Draw λ uniformly on the set of feasible eigenvalues $d \cdot \mathcal{S}_d$, cf. Fang *et al.* (1990, Theorem 5.2(2)):

$$\begin{aligned} &\mathbf{e} = (e_1, \dots, e_d),\ e_i \sim \mathrm{Exp}(1) \text{ iid},\ i \in \{1, \dots, d\}, \\ &\lambda = (\lambda_1, \dots, \lambda_d) = d \cdot \mathrm{sort}(\mathbf{e}/\mathbf{e}'\mathbf{1}). \end{aligned}$$

(2) Draw a random orthogonal matrix Q, cf. Stewart (1980):

$$\begin{aligned} &X \in \mathbb{R}^{d \times d},\ x_{i,j} \sim \mathrm{N}(0,1) \text{ iid},\ i, j \in \{1, \dots, d\}, \\ &[Q, R] = \mathrm{qr}(X). \end{aligned}$$

[14] Once the rotation indices i and j are chosen, and it is decided which of the two diagonal entries should be set to 1, we have two possible solutions for the tangent, and two choices for the sign of the cosine, cf. Equations (2.2) and (2.3).

(3) Initialize:

$$\Lambda = \mathrm{diag}(\lambda), \qquad W = Q\Lambda Q', \qquad V = Q,$$

and sequentially compute the Givens rotations:
While any$(\mathrm{diag}(W) \neq 1)$,

(a) Uniformly select rotation indices $i < j$ from the sets $\{l \in \{1, \dots, d\} : w_{l,l} < 1\}$ and $\{l \in \{1, \dots, d\} : w_{l,l} > 1\}$.
Uniformly select one of the two indices. The algorithm then forces the diagonal entry $[GWG']_{i,i}$, resp. $[GWG']_{j,j}$, to 1.

(b.1) In the first iteration $k = 1$:
Uniformly select one of the two possible solutions for the tangent in Equation (2.2), resp. (2.3), i.e. $t \in \{t_1^+, t_1^-\}$.
Uniformly select the sign of the cosine $c = \pm\sqrt{1 + t^2}^{-1}$. This fixes the sign of the entries of $\mathbf{v}_1$.

(b.2) In iterations $k = 2, \dots, d-2$:
Uniformly select one of the two possible solutions to the quadratic equation in the tangent (2.2), resp. (2.3), i.e. $t \in \{t_k^+, t_k^-\}$.
Select the sign of $c = \pm\sqrt{1 + t^2}^{-1}$ such that the i-th, resp. j-th, entry of $\mathbf{v}_1$ has the desired sign, cf. Equation (2.4):

$$[GV]_{i,1} = c \cdot (v_{i,1} - t v_{j,1}) \geq 0 (\leq 0) \qquad \text{if } [GWG']_{i,i} \overset{!}{=} 1$$
$$[GV]_{j,1} = c \cdot (t v_{i,1} + v_{j,1}) \geq 0 (\leq 0) \qquad \text{if } [GWG']_{j,j} \overset{!}{=} 1$$

(b.3) In the last iteration $k = d - 1$:
Choose $t = t_{d-1} \in \{t_{d-1}^+, t_{d-1}^-\}$ such that $\mathrm{sign}(v_{i,1} - t v_{j,1}) = \mathrm{sign}(t v_{i,1} + v_{j,1})$.
Select the sign of $c = \pm\sqrt{1 + t^2}^{-1}$ such that the i-th and j-th entry of $\mathbf{v}_1$ have the desired sign.

(c) Construct the Givens rotation G accordingly, cf. Equation (2.1), and save for the next iteration:

$$W = GWG', \qquad V = GV.$$

Remark 2.3.5 (Clarifications)

- In step (1), $e_i \sim \mathrm{Exp}(1)$ iid means the random variables e_i are independent and identically distributed according to the exponential law with mean 1. Here, one can instead draw λ from a desired eigenvalue distribution other than the uniform law on $d \cdot \mathcal{S}_d$.
- In step (2), $\mathrm{N}(0, 1)$ denotes the standard normal distribution, and $\mathrm{qr}(\cdot)$ refers to the QR-decomposition, where the diagonal entries of R are positive, cf. Stewart (1980).

- In step (3), 'uniformly select' refers to a draw from the discrete uniform distribution on the respective sets.
- When referring to the sign of a quantity, we only distinguish between $a \geq 0$ and $a \leq 0$, i.e. $a = 0$ is compatible with either class, corresponding to the definition of the (non-strong) Perron–Frobenius property. In any case, $t, c \neq 0$ by construction, and Lemma 2.3.2 ensures that a.s. $[GV]_{i,1} \neq 0$, resp. $[GV]_{j,1} \neq 0$.
- In the case that $[Q\Lambda Q']_{l,l} = 1$ for some $l \in \{1, \ldots, d\}$, which a.s. never happens[15], the number of iterations required in step (3.b) decreases by the cardinality of the set $\{l \in \{1, \ldots, d\} : [Q\Lambda Q']_{l,l} = 1\}$, if the corresponding entries $q_{l,1}$ all have the same sign. Otherwise, $Q\Lambda Q'$ has to be left- and right-multiplied with another random orthogonal matrix before proceeding with step (3), as the algorithm will not modify the signs of these $q_{l,1}$.

Theorem 2.3.6 (Validity of Algorithm 2.3.4)
Algorithm 2.3.4 generates Perron–Frobenius correlation matrices. The generated matrices a.s. exhibit a dominant eigenvector with positive entries.

To prove Theorem 2.3.6, we require the following statements:

- The algorithm a.s. produces no zero entries in $\mathbf{v}_1$. This ensures that the algorithm a.s. produces correlation matrices with the strong Perron–Frobenius property, and in step (3.b.1) the first rotation indeed fixes the sign of the entries of $\mathbf{v}_1$.
  This follows directly from Lemma 2.3.2.
- In step (3.b.3), at least (and a.s. exactly) one of the possible rotations G_{d-1} simultaneously fixes the sign of the remaining two entries of $\mathbf{v}_1$. (For the previous rotations there is nothing to show in this respect, since they are specifically chosen to fix the sign of one entry of $\mathbf{v}_1$, and the remaining entries will be affected by later rotations.)
  This is shown in the following two lemmas.

Lemma 2.3.7 (Relation of tangent solutions in step (3.b.3))
The solutions t^+_{d-1}, t^-_{d-1} of (2.2) in step (3.b.3) of Algorithm 2.3.4 satisfy the relation $t^+_{d-1} = -(t^-_{d-1})^{-1}$.

Proof
As all diagonal elements in $W^{(d-2)}$ except $i = i_{d-1}$ and $j = j_{d-1}$ have been set to 1, and $\mathrm{trace}(W^{(d-2)}) = d$, we have $w^{(d-2)}_{ii} + w^{(d-2)}_{jj} = 2$. So the tangent equation (2.2) becomes

$$t_{d-1} = \frac{w^{(d-2)}_{ij} \pm \sqrt{(w^{(d-2)}_{ij})^2 + (w^{(d-2)}_{jj} - 1)^2}}{w^{(d-2)}_{jj} - 1},$$

[15] This follows from a similar argumentation as in the proof of Lemma 2.3.2, as the set $\{Q \in \mathcal{O}(d) : [Q\Lambda Q']_{l,l} = 1 \text{ for some } l \in \{1, \ldots, d\}\}$ is a lower-dimensional subset of $\mathcal{O}(d)$.

and we get (dropping the superscript $(d-2)$ of the W-entries for notational simplicity)

$$\frac{w_{ij} - \sqrt{w_{ij}^2 + (w_{jj}-1)^2}}{w_{jj}-1} = -\frac{w_{jj}-1}{w_{ij} + \sqrt{(w_{ij})^2 + (w_{jj}-1)^2}}$$
$$\Leftrightarrow -(w_{jj}-1)^2 = -(w_{jj}-1)^2,$$

a true statement. Note that in step (3.b.3) of the algorithm forcing the i-th diagonal entry to 1 is equivalent to forcing the j-th diagonal entry to 1, as the trace is preserved and the remaining entry is set to 1 simultaneously. Thus wlog. we can use (2.2). □

Lemma 2.3.8 (Validity of step (3.b.3))
In step (3.b.3), at least (and a.s. exactly) one of the four possible Givens rotations G_{d-1} forces the remaining two entries of $\mathbf{v}_1$ to simultaneously be non-negative, resp. non-positive.

Proof
The final eigenvector matrix can be expressed as $V = G_{d-1}V^{(d-2)}$. Let $(i,j) = (i_{d-1}, j_{d-1})$ be the remaining diagonal entries of $W^{(d-2)}$ that are not equal to one, i.e. we need to force $v_{i,1}$ and $v_{j,1}$ simultaneously to have the sign specified in step (3.b.1) of the algorithm. From Equation (2.4), with $s_{d-1} = t_{d-1}c_{d-1}$, we know:

$$v_{i,1} = c_{d-1}\underbrace{(v_{i,1}^{(d-2)} - t_{d-1}v_{j,1}^{(d-2)})}_{=:A} \quad \text{and} \quad v_{j,1} = c_{d-1}\underbrace{(t_{d-1}v_{i,1}^{(d-2)} + v_{j,1}^{(d-2)})}_{=:B}.$$

To be able to force $v_{i,1}$ and $v_{j,1}$ simultaneously to have the same sign, we need to show that for at least one of the two possible solutions for t_{d-1}, the terms A and B have the same sign.
Denote $A^{\pm} = v_{i,1}^{(d-2)} - t_{d-1}^{\pm}v_{j,1}^{(d-2)}$ and $B^{\pm} = t_{d-1}^{\pm}v_{i,1}^{(d-2)} + v_{j,1}^{(d-2)}$. From Lemma 2.3.7 we know that the two possible solutions for t_{d-1} fulfill $t_{d-1}^{+} = -1/t_{d-1}^{-}$. This yields

$$A^- = v_{i,1}^{(d-2)} - t_{d-1}^- v_{j,1}^{(d-2)} = v_{i,1}^{(d-2)} + \frac{1}{t_{d-1}^+}v_{j,1}^{(d-2)} \Leftrightarrow t_{d-1}^+ A^- = B^+$$
$$B^- = t_{d-1}^- v_{i,1}^{(d-2)} + v_{j,1}^{(d-2)} = -\frac{1}{t_{d-1}^+}v_{i,1}^{(d-2)} + v_{j,1}^{(d-2)} \Leftrightarrow t_{d-1}^+ B^- = -A^+.$$

Thus we have for the signs of $A^{\pm}$ and $B^{\pm}$:

$$\left\{\begin{array}{l} A^+ > 0 \\ A^+ = 0 \\ A^+ < 0 \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} B^- < 0 \\ B^- = 0 \\ B^- > 0 \end{array}\right\}, \qquad \left\{\begin{array}{l} B^+ > 0 \\ B^+ = 0 \\ B^+ < 0 \end{array}\right\} \Leftrightarrow \left\{\begin{array}{l} A^- > 0 \\ A^- = 0 \\ A^- < 0 \end{array}\right\},$$

i.e. we always have (A^+, B^+) non-negative (non-positive), or (A^-, B^-) non-negative (non-positive). If either $A^+ = 0$ or $B^+ = 0$, both tangent solutions are feasible, as either $v_{i1} = 0$ or $v_{j1} = 0$ in this case, and the sign of c_{d-1} can be chosen to match the

sign of the remaining entry of $\mathbf{v}_1$ to the sign chosen in step (3.b.1) of the algorithm. If both $A^+ = 0$ and $B^+ = 0$, $v_{i1} = v_{j1} = 0$, i.e. all four Givens rotations are feasible. However, taking into account the statement of Lemma 2.3.2 that zero entries in $\mathbf{v}_1$ a.s. never occur, we a.s. have either (A^+, B^+) or (A^-, B^-) of the same sign (i.e. both positive, resp. negative). The sign of $c_{d-1} = \pm(1 + t_{d-1}^2)^{-\frac{1}{2}}$ can then be chosen accordingly to match the sign of the entries of $\mathbf{v}_1$ specified in step (3.b.1) of the algorithm, which means that a.s. exactly one of the four possible rotations G_{d-1} is able to force the remaining entries of the dominant eigenvector simultaneously to have the specified sign. □

The correlation matrices our algorithm produces are not uniform on the set of Perron–Frobenius correlation matrices, just as `randcorr` does not produce uniformly distributed correlation matrices, cf. Figures 2.3 and 2.4 for a simulation in $d = 3$. The simulated correlation matrices are represented by the 3-dimensional vector of pairwise correlations. Obviously, `randcorr` assigns most of the mass around the diagonals $\pm\rho_{12} = \pm\rho_{13} = \pm\rho_{23}$. Our algorithm inherits this behaviour. The simulations illustrate Boyle *et al.* (2014); Boyle and N'Diaye (2018)'s theoretical result that the sum of each pair of ρ_{ij}, $i \neq j$, $i, j \in \{1, 2, 3\}$, is positive for $C \in \mathcal{C}_3^{PF}$.

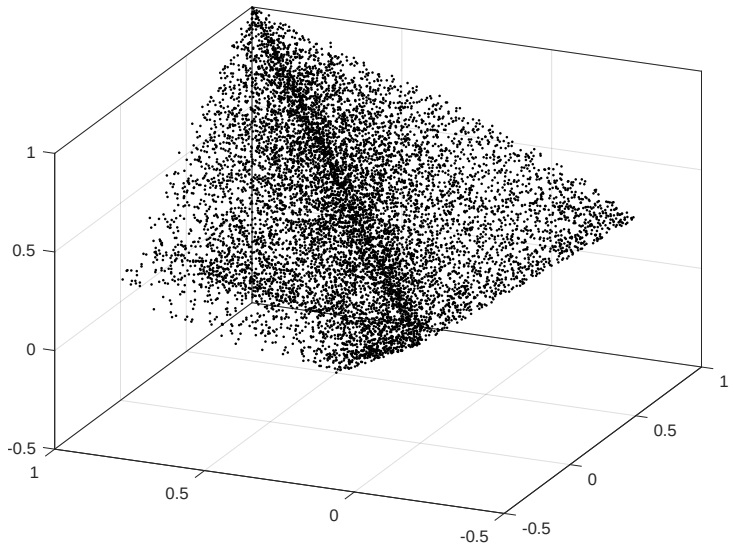

Figure 2.3: Distribution of pairwise correlations for correlation matrices simulated from our algorithm in $d = 3$ with eigenvalues uniformly distributed on d times the d-simplex.

As can be derived from Lemma 2.3.1, the distribution of correlation matrices generated by Algorithm 2.3.4 has full support $\mathcal{C}_d^{PF}$. Further, the following interesting fact about the volume of the set $\mathcal{C}_d^{PF}$ follows directly from the construction principle used in Algorithm 2.3.4. Table 2.1 and Boyle *et al.* (2014); Boyle and N'Diaye (2018) already give an indication for this.

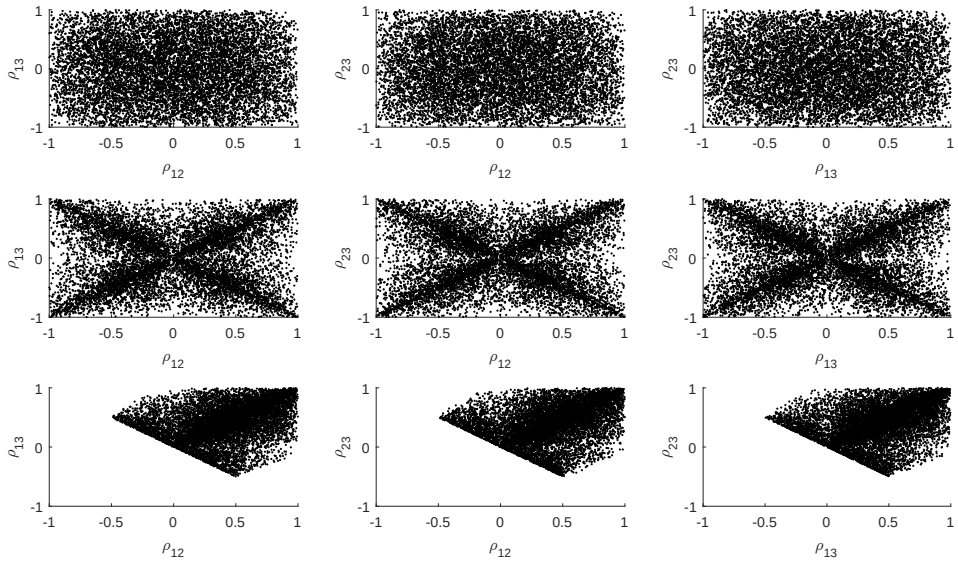

Figure 2.4: Distribution (projections to two-dimensional space) of pairwise correlations for correlation matrices in $d = 3$ simulated from the uniform distribution (top row), the `randcorr` algorithm with eigenvalues uniformly distributed on d times the d-simplex (middle row), and our algorithm with the same eigenvalue distribution (bottom row).

Theorem 2.3.9 (Proportion of Perron–Frobenius correlation matrices)
The proportion of Perron–Frobenius correlation matrices in the set of all correlation matrices in dimension d is $1/2^{d-1}$.

Proof
This follows from Lemmas 2.3.1 and 2.3.2, and a decomposition of the basic procedure into random and deterministic components, wlog. assuming that we always set the i_k-th diagonal entry to 1 in each iteration k:

1. The support of the `randcorr` algorithm is $\mathcal{C}_d$, cf. Lemma 2.3.1. Further, it a.s. requires $d - 1$ Givens rotations to force all diagonal entries to 1, as all diagonal entries of the initial matrix $W^{(0)} = Q\Lambda Q'$ a.s. differ from 1.
2. Zero entries in $\mathbf{v}_1$ a.s. never occur, cf. Lemma 2.3.2.

3. We can decompose the `randcorr` algorithm (and thus analogously also Algorithm 2.3.4) into the following two steps:

   a) Simulate $\omega := (\Lambda, Q, \sigma) \in \Omega$,

   $$\Omega = \{\Lambda = \mathrm{diag}(\lambda), \lambda \in \mathbb{R}^d_{+,0} : \lambda'\mathbf{1} = d\} \times \mathcal{O}(d) \times \Big\{\{(i_k, j_k)\}_{k\in\{1,\dots,d-1\}} : (i_k, j_k) \in \{1, \dots, d\}^2, i_k \neq j_k \forall k, \{i_1, \dots, i_{d-1}\} \subset \{1, \dots, d\}\Big\}.$$

   The eigenvalues λ are drawn from an absolutely continuous distribution with full support $d \cdot \mathcal{S}_d$, the orthogonal matrix Q is drawn from the Haar measure, and σ denotes a particular sequence of pairs of rotation indices $\{(i_k, j_k)\}_{k\in\{1,\dots,d-1\}}$. Remembering that Algorithm 2.3.4 draws the rotation indices sequentially, it is important to note that the sequence σ can indeed be chosen knowing only the initial matrix $W^{(0)}$, i.e. Q and λ: wlog. i_1 and j_1 are chosen uniformly from the sets $\{k : w^{(0)}_{kk} > 1\}$ and $\{k : w^{(0)}_{kk} < 1\}$, respectively. Since multiplication with a Givens rotation G_k affects only the i_k-th and j_k-th rows and columns of $W^{(k-1)}$ and does not alter the trace of a matrix, we know the diagonal elements of $W^{(1)} = G_1 W^{(0)} G_1'$ are $\{w^{(1)}_{kk}\}_{k\in\{1,\dots,d\}} = \{w^{(0)}_{kk}\}_{k\in\{1,\dots,d\}\setminus\{i_1,j_1\}} \cup \{1, w^{(0)}_{i_1,i_1} + w^{(0)}_{j_1,j_1} - 1\}$, and i_2 and j_2 can wlog. be chosen uniformly from the sets $\{k : w^{(1)}_{kk} > 1\}$ and $\{k : w^{(1)}_{kk} < 1\}$, respectively. The remaining pairs of rotation indices are chosen sequentially in the same manner.

   b) Given ω, for each Givens rotation G_k, $k \in \{1, \dots, d-1\}$, there are four potential choices $G^{(l_k)}_k$, from which the `randcorr` algorithm chooses uniformly, i.e. the algorithm chooses $\mathbf{l} := \mathbf{l}(\omega) \in \{1, 2, 3, 4\}^{d-1}$, $\mathbf{l} = (l_1, \dots, l_{d-1})$, where we identify $G^{(1)}_k$ with the Givens rotation obtained when taking the positive t^+_k of the two tangent solutions t_k and the positive sign for the cosine c_k, $G^{(2)}_k$ with the one obtained when taking $t_k = t^+_k$ and c_k negative, $G^{(3)}_k$ with the one obtained when taking $t_k = t^-_k$ and c_k positive, and $G^{(4)}_k$ with the one obtained when taking $t_k = t^-_k$ and c_k negative. We denote $A = A(\mathbf{l}) := G^{(l_{d-1})}_{d-1} \cdots G^{(l_1)}_1$ in the sequel.

4. By (1) and (3), we know:

   $$\mathcal{C}_d \overset{\text{a.s.}}{=} \left\{\cup_{\mathbf{l}(\omega)\in\{1,2,3,4\}^{d-1}} A(\mathbf{l}(\omega)) Q \Lambda Q' A(\mathbf{l}(\omega))' : \omega \in \Omega\right\},$$

   and by the construction principle employed in Algorithm 2.3.4, we have:

   $$\mathcal{C}^{PF}_d \overset{\text{a.s.}}{=} \left\{\cup_{\mathbf{l}(\omega)\in N(\omega)} A(\mathbf{l}(\omega)) Q \Lambda Q' A(\mathbf{l}(\omega))' : \omega \in \Omega\right\},$$

where

$$N(\omega) = \{\mathbf{l}(\omega) \in \{1,2,3,4\}^{d-1} :$$
$l_1 \in \{1,2,3,4\}$ arbitrary, determines the sign of the entries of $\mathbf{v}_1$,
$l_k, k = 2,\dots,d-2$ is one of those elements in $\{1,2,3,4\}$ that ensures that the updated i_k-th entry of the first eigenvector has the desired sign; Lemma 2.3.2 ensures there are a.s. two of these (compare step (3.b.2) of Algorithm 2.3.4), and
l_{d-1} is a.s. uniquely determined by Lemma 2.3.8$\}$.

Therefore, $|N(\omega)| = 4 \cdot 2^{d-3} \cdot 1 = 2^{d-1}$, and the proportion of Perron–Frobenius correlation matrices is $|N(\omega)|/4^{d-1} = 1/2^{d-1}$. □

Using Theorem 2.3.9 and the normalizing constant for the uniform distribution on $\mathcal{C}_d$ as given in Joe (2006), we can calculate the volume[16] of the set of d-dimensional Perron–Frobenius correlation matrices $\mathcal{C}_d^{PF}$, interpreted as a subset of $\mathbb{R}^{d(d-1)/2}$:

$$\mathrm{Vol}(\mathcal{C}_d^{PF}) = 2^{\frac{1}{6}(d-1)(2d^2-d-6)} \prod_{k=1}^{d-1} B((k+1)/2, (k+1)/2)^k,$$

where $B(\cdot,\cdot)$ is the beta function.
Table 2.2 lists this volume for dimensions $d \in \{3,4,5,7,10\}$.

dimension	$d=3$	$d=4$	$d=5$	$d=7$	$d=10$
$\mathrm{Vol}(\mathcal{C}_d^{PF})$	0.3886	1.4622	1.4083	0.4353	0.0013

Table 2.2: Volume of $\mathcal{C}_d^{PF}$ for $d \in \{3,4,5,7,10,15\}$.

2.3.3 Presence of other stylized facts for different eigenvalue distributions

It would be desirable to further develop the algorithm such that a generation of correlation matrices exhibiting stylized facts (S1)-(S3) (or even (S1)-(S4)) described in the introduction is possible, i.e. that additionally to the Perron–Frobenius property and a realistic eigenvalue distribution, the generated correlation matrices' off-diagonal entries have a realistic distribution, and their corresponding MSTs exhibit a power law degree distribution. Several empirical studies have found that realistic distributions for pairwise correlations exhibit a positive mean of about 0.3-0.5, cf. e.g. Kazakov and Kalyagin (2016), (and are unimodal, and negative correlations typically are not large in absolute value).

[16] This formula corrects a typo in the exponent made in the respective formula in our paper Hüttner and Mai (2019).

Remark 2.3.10 (No relation between (S2) and (S3))
As a correlation matrix C with all entries positive has a positive dominant eigenvector by the Perron–Frobenius Theorem, it might be tempting to think that the Perron–Frobenius property already encompasses stylized fact (S3) (distribution of pairwise correlations is significantly shifted to the positive) or vice versa. However, the following example in $d = 5$ illustrates that correlation matrices may have several negative correlations and still exhibit the Perron–Frobenius property, while the mean of pairwise correlations is even slightly negative:

$$C = \begin{pmatrix} 1.0000 & -0.2704 & -0.1370 & -0.1774 & 0.4315 \\ -0.2704 & 1.0000 & -0.2277 & -0.0651 & 0.3619 \\ -0.1370 & -0.2277 & 1.0000 & 0.7374 & 0.4277 \\ -0.1774 & -0.0651 & 0.7374 & 1.0000 & 0.4955 \\ 0.4315 & 0.3619 & 0.4277 & 0.4955 & 1.0000 \end{pmatrix},$$

$$\mathbf{v}_1(C) = \begin{pmatrix} 0.5118 \\ 0.6162 \\ 0.5980 \\ 0.0017 \\ 0.0260 \end{pmatrix}, \quad \lambda(C) = \begin{pmatrix} 2.1194 \\ 1.3314 \\ 1.2692 \\ 0.2470 \\ 0.0329 \end{pmatrix}, \quad \frac{1}{10}\sum_{i=1}^{4}\sum_{j=i+1}^{5} \rho_{i,j} = -0.0067.$$

On the other hand, the theoretical results for $d = 4$ in Boyle and N'Diaye (2018) state that the number of negative correlations in Perron–Frobenius correlation matrices is limited to at most 3, and restrictions on their relative positioning apply. A quick simulation study using Algorithm 2.3.4 (simulation of $n = 10000$ $d \times d$ correlation matrices for $d \in \{4, 5, 6\}$, $\lambda/d \sim \mathcal{U}(\mathcal{S}_d)$) shows that these results do not generalize to higher d: There exist Perron–Frobenius correlation matrices with more negative than non-negative pairwise correlation entries: For 8 of the 10000 simulated 6×6 Perron–Frobenius correlation matrices, 9 of the 15 pairwise correlations were negative.
Finally, the following counterexample shows that, vice versa, a positive mean of pairwise correlations (S3) does not imply the Perron–Frobenius property (S2):

$$C = \begin{pmatrix} 1.0000 & 0.6551 & 0.6028 & 0.1704 \\ 0.6551 & 1.0000 & 0.6112 & -0.4248 \\ 0.6028 & 0.6112 & 1.0000 & 0.1485 \\ 0.1704 & -0.4248 & 0.1485 & 1.0000 \end{pmatrix},$$

$$\mathbf{v}_1(C) = \begin{pmatrix} -0.5739 \\ -0.5943 \\ -0.5605 \\ 0.0571 \end{pmatrix}, \quad \lambda(C) = \begin{pmatrix} 0.0999 \\ 0.4000 \\ 1.2500 \\ 2.2501 \end{pmatrix}, \quad \frac{1}{6}\sum_{i=1}^{3}\sum_{j=i+1}^{4} \rho_{i,j} = 0.2939.$$

In general, as stated in Section 2.2.2, it is very hard to derive statements about the distribution of pairwise correlations simulated from the `randcorr` algorithm, and consequently this holds true also for our Algorithm 2.3.4. It seems very hard to manipulate the algorithm further to take the positive shift of pairwise correlations into account.

Still, in the following we aim to provide some enlightenment about the distributions of pairwise correlation entries for different choices of eigenvalue distributions. To this end, we consider the following distributions for the eigenvalues λ:

1) the uniform distribution of eigenvalues, i.e. λ/d simulated uniformly on the d-simplex $\mathcal{S}_d$,

2) (a special version of) the power-law distribution suggested by Bouchaud and Potters (2011); Bun *et al.* (2017) that was found to capture the eigenvalue distribution of empirical correlation matrices fairly well, where the entries of $\tilde{\lambda} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_d)$ are simulated independently from the density[17]

$$f(x) = \frac{2}{(x+1)^3}, \tag{2.5}$$

and $\lambda = d \cdot \tilde{\lambda}/\|\tilde{\lambda}\|$, and

3) a variant of the above power-law where the largest eigenvalue is fixed at 40% of total variance, i.e. $\lambda_1 = 0.4 \cdot d$ and $\tilde{\lambda}_i, i = 2, \ldots, d$, are simulated from density (2.5) and rescaled as described above such that $\lambda_2 + \ldots + \lambda_d = 0.6 \cdot d$.

Presence of (S3)

We simulate $n = 10,000$ Perron–Frobenius correlation matrices, with eigenvalues simulated according to these distributions (i.e. we alter step (1) of Algorithm 2.3.4), and study the distributions of the pairwise correlation entries.

1) When $\lambda/d \sim \mathcal{U}(\mathcal{S}_d)$, pairwise correlations tend to be unrealistically small for large correlation matrices. The range of pairwise correlations shrinks with increasing dimension d of the matrix. The mean of pairwise correlations is slightly shifted to the positive, but approaches zero for increasing d, cf. Table 2.3, which is too small compared to empirical observations of the average pairwise correlation in large financial data sets. Further, in empirical financial correlation matrices, one typically finds more positive than negative correlations, which are larger in absolute value, whereas the absolute size of the observed negative correlations is small. This is not the case for correlation matrices simulated from our algorithm with uniform eigenvalue distribution. See Figure 2.5 for a plot of empirical densities of pairwise correlations in $d = 100$.

2) Using Bouchaud and Potters (2011)'s power law distribution for the simulation of eigenvalues as described above, one gets very diverse distributions for the pairwise correlation entries, cf. Table 2.3 and Figure 2.5. Overall, the distributions of pairwise correlation entries of the simulated correlation matrices with this eigenvalue

[17] Bouchaud and Potters (2011); Bun *et al.* (2017)'s power law relies on the lower bound λ^- for the eigenvalues of a random correlation matrix in the sense of random matrix theory. We set $\lambda^- = 0$ here, which conforms to the random matrix theory limit when the data matrix is $N \times N$ and $N \to \infty$.

distribution tend to exhibit several realistic features[18]: Regardless of dimension, pairwise correlations on average lie in the range [-0.4,0.6], with the mean shifted to the positive. However, the minimum and mean pairwise correlation is on average smaller than what is typically observed in market correlation matrices.

3) Using Bouchaud and Potters (2011)'s power law plus a fixed largest eigenvalue explaining 40% of the total variance, we find that pairwise correlation entries of the simulated correlation matrices exhibit quite realistic distributions: Regardless of dimension, the average pairwise correlation is approximately 0.35, fluctuating in a range of [0.25,0.53], which is in line with the observations in empirical studies, cf. Kazakov and Kalyagin (2016); Plerou *et al.* (2002) and Section 2.4. The distribution functions are smooth and unimodal, with more weight on positive correlations than on negative ones, however, large negative correlations are on average more likely than typically observed in correlation matrices of financial data sets, see Figure 2.5.

1) Uniform	avg min	range min	avg mean	range mean	avg max	range max
$d = 25$	-0.3946	[-0.7597,-0.0379]	0.0886	[0.0196,0.3355]	0.5262	[0.2871,0.8403]
$d = 50$	-0.3614	[-0.6230,-0.1723]	0.0518	[0.0124,0.2072]	0.4426	[0.2791,0.6693]
$d = 75$	-0.3321	[-0.5280,-0.2034]	0.0372	[0.0140,0.1362]	0.3930	[0.2608,0.5961]
$d = 100$	-0.3094	[-0.5547,-0.2099]	0.0295	[0.0118,0.1037]	0.3588	[0.2438,0.5367]
2) Power-law	avg min	range min	avg mean	range mean	avg max	range max
$d = 25$	-0.4228	[-0.9233,0.9319]	0.2020	[0.0223,0.9644]	0.6657	[0.3369,0.9848]
$d = 50$	-0.4389	[-0.8270,0.8174]	0.1502	[0.0243,0.9070]	0.6011	[0.3260,0.9654]
$d = 75$	-0.4293	[-0.7942,0.8619]	0.1268	[0.0187,0.9297]	0.5609	[0.3162,0.9658]
$d = 100$	-0.4157	[-0.8099,0.9076]	0.1131	[0.0191,0.9587]	0.5315	[0.3364,0.9779]
3) Power-law	avg min	range min	avg mean	range mean	avg max	range max
$d = 25$	-0.4509	[-0.9701,0.0979]	0.3384	[0.2585,0.5272]	0.7782	[0.5893,0.9947]
$d = 50$	-0.4307	[-0.9679,0.0859]	0.3603	[0.2991,0.5023]	0.7502	[0.6038,0.9819]
$d = 75$	-0.4128	[-0.9464,0.0879]	0.3693	[0.3036,0.4705]	0.7311	[0.6084,0.9758]
$d = 100$	-0.3888	[-0.9805,0.0735]	0.3750	[0.3039,0.4883]	0.7144	[0.6058,0.9872]

Table 2.3: Average values of the minimum, mean, and maximum values of pairwise correlations in correlation matrices simulated from our algorithm with uniform and power-law eigenvalue distribution; $n = 10,000$ simulations in dimensions $d \in \{25, 50, 75, 100\}$.

Presence of (S4)

In the following, we check whether the MSTs of the simulated correlation matrices exhibit the scale-free property, using as diagnostics:

(i) the number of leaves,

(ii) the maximal degree, i.e. the maximal number of neighbors a node in the MST has, and

[18]Extreme cases where all pairwise correlations are larger than 0.9 are possible, but occur very rarely.

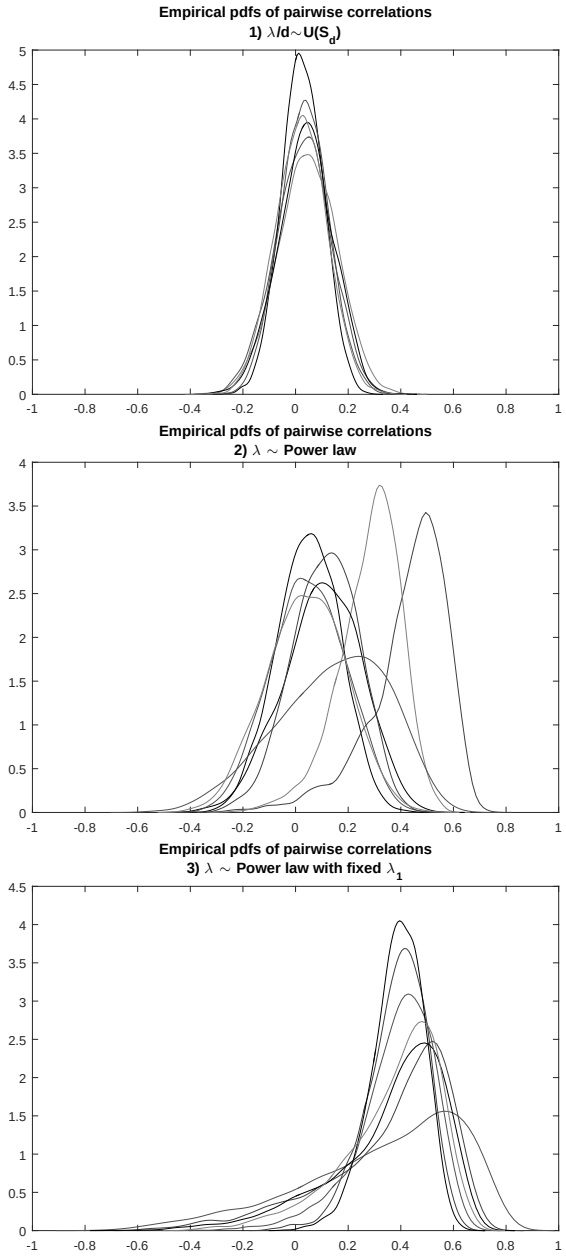

Figure 2.5: Empirical densities of pairwise correlations for several correlation matrices simulated from our Algorithm 2.3.4 in $d = 100$ with eigenvalues distributed 1) uniformly, i.e. $\lambda/d \sim \mathcal{U}(\mathcal{S}_d)$ (top), 2) according to density (2.5) (middle), and 3) with fixed largest eigenvalue of 40% of total variance and remaining eigenvalues distributed according to density (2.5) (bottom).

(iii) a log-log plot of the degree distribution, which would be approximately a straight line in the case of a power law degree distribution.

1) Concerning the tree structure of MSTs obtained from correlation matrices simulated from our algorithm, we find that when λ/d are distributed uniformly on the d-simplex for $d = 20$, the MSTs exhibits fewer leaves than MSTs derived from correlation matrices of d randomly drawn assets from our data set of 395 5Y-CDS log return time series of constituents of the four major credit indices (ITRX, ITRX HY, CDX, CDX XO) we use for the assessment of graph-based portfolio selection methods in Section 2.4 below, cf. Figure 2.6. Studying the degree distribution, i.e. the distribution of the number of neighbors, in $d = 100$, we find that it is not approximated well by a power law, as the scale-free property would postulate, but rather exhibits an exponential decay for large numbers of neighbors, which is typically encountered for 'random' graphs, cf. Vandewalle *et al.* (2001), with the maximal degree (i.e. the highest number of neighbors a node in the graph has) being at most 17 in $n = 10,000$ simulations. In $d = 1000$, the degree distribution is still not well approximated by a power law in $n = 1000$ simulations, but decays exponentially, cf. Figure 2.7.

2) When using the power law density (2.5) for the simulation of eigenvalues, we find a similar behaviour: Also with this eigenvalue distribution, MSTs derived from correlation matrices simulated from our algorithm exhibit fewer leaves than MSTs derived from financial correlation matrices in $d = 20$, cf. Figure 2.6. Again the degree distribution in $d = 100$ exhibits exponential decay, not the desired power law-like behaviour, with the maximal degree being at most 24 in $n = 10,000$ simulations. In $d = 1000$, we find that the degree distribution still decays exponentially, but less pronounced as in the case of uniformly distributed eigenvalues.

3) If, additionally to the power law eigenvalue distribution, we fix the largest eigenvalue at 40% of total variance, we still find fewer leaves than in market-derived MSTs in $d = 20$, cf. Figure 2.6, and exponential decay of the degree distribution in $d = 100$, with the maximal degree being at most 25 in $n = 10,000$ simulations. In large dimensions, however, the degree distribution approaches a power law-like behaviour, cf. Figure 2.7 for a log-log plot of simulated degree distributions in $d = 1000$.

So, whereas for eigenvalue distributions 1) and 2), the scale-free property cannot be detected, fixing the largest eigenvalue at 40% of total variance and the remaining eigenvalues distributed according to density (2.5), we find that the MSTs of large correlation matrices simulated from Algorithm 2.3.4 do indeed approach a power law-like degree distribution. When fitting a line to a log-log plot of the degree distribution[19], we find

[19] Although this is not the most reliable method for fitting a power law, we nevertheless stick with this approach for the sake of simplicity, as we only intend to provide an intuition about the behaviour of the degree distribution of the MSTs related to correlation matrices simulated from our Algorithm 2.3.4. See Clauset *et al.* (2009) for more information on the drawbacks of this approach for

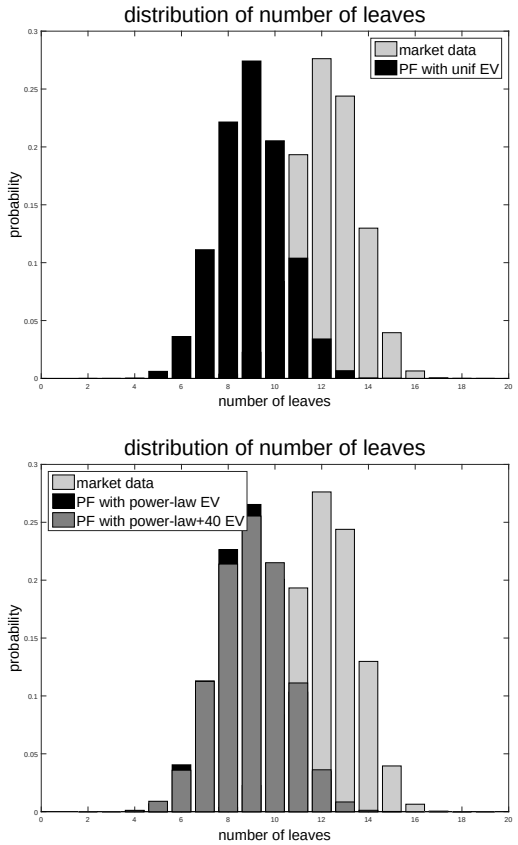

Figure 2.6: Histograms of the number of leaves of MSTs derived from correlation matrices simulated according to Algorithm 2.3.4 for different eigenvalue distributions, $d = 20$, in contrast to empirical correlation matrices. Top: Uniform distribution of eigenvalues. Bottom: Power law distribution of eigenvalues, once with fixed largest eigenvalue explaining 40% of total variance.

that the exponent of the fitted power law is approximately in $[-2.2, -2]$, which is (in absolute value) slightly below the values found in the empirical studies Vandewalle *et al.* (2001) (exponent -2.2±0.1) and Bonanno *et al.* (2003) (exponent -2.6).

fitting power-law distributions, and viable alternatives.

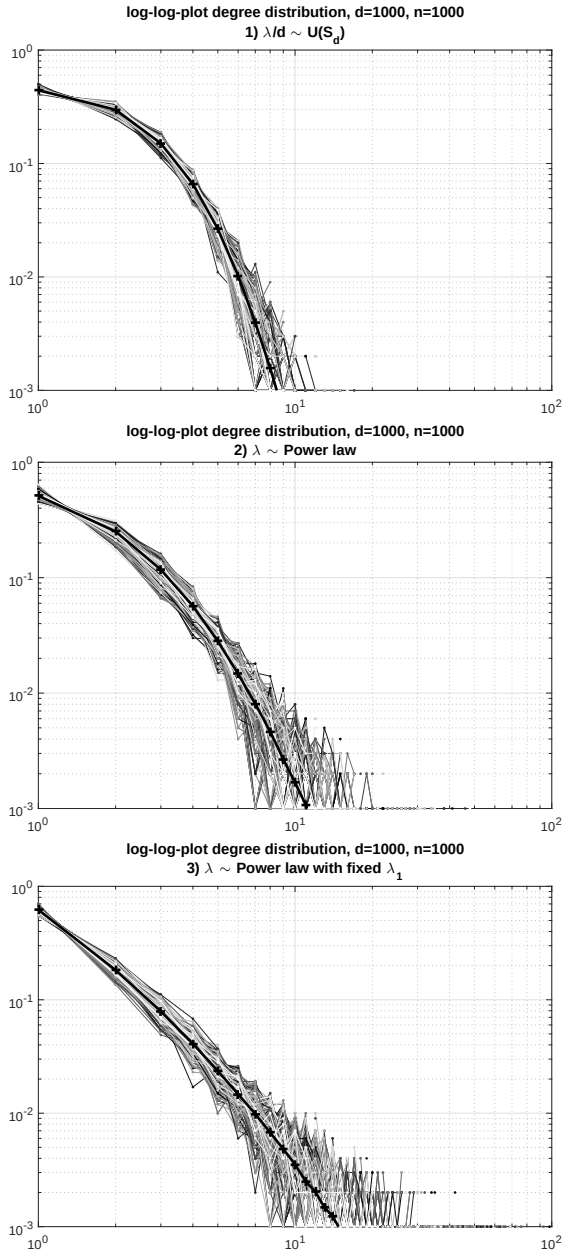

Figure 2.7: Degree distribution of MSTs derived from correlation matrices simulated according to Algorithm 2.3.4 for different eigenvalue distributions, $d = 1000$. Top: Uniform distribution of eigenvalues. Middle: Power law distribution of eigenvalues. Bottom: Power law distribution of eigenvalues with fixed largest eigenvalue explaining 40% of total variance. The thick black lines represent the respective means.

2.4 Application: Assessing graph-based portfolio selection techniques

Several empirical studies focusing on graph-based portfolio selection techniques have documented a strong relation between these methods and the classical Markowitz ansatz, cf. Onnela *et al.* (2003); Pozzi *et al.* (2013); Kaya (2015); Peralta and Zareei (2016). With different algorithms for simulating correlation matrices at hand, we are able to analyze whether this empirical relation stems from a fundamental connection between the two different approaches or is due to the special structure of market correlation matrices. We consider an investment universe of $d \in \mathbb{N}$ assets. Each asset $k = 1, \ldots, d$ is associated with its annualized log-return R_k, and investment decisions are based on the probability distribution of the vector $\mathbf{R} := (R_1, \ldots, R_d)$.

The problem of optimal investment, central in mathematical finance, was first formalized in the seminal work of Markowitz (1952, 1959). There, optimal investment is considered in terms of the first two moments of the distribution of $\mathbf{R}$, the covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ and, if desired, an expected return estimate $\mu \in \mathbb{R}^d$. A portfolio in the d assets is given by a vector $\mathbf{x} = (x_1, \ldots, x_d)' \in \mathbb{R}^d$ satisfying $\mathbf{1}'\mathbf{x} = x_1 + \ldots + x_d = 1$, with $\mathbf{1}$ denoting a d-dimensional column vector with all entries being equal to one. Component x_i gives the portfolio weight of asset i, with negative value corresponding to shortselling the asset. The side condition $\mathbf{1}'\mathbf{x} = 1$ demands that the portfolio is fully invested, shortselling being allowed. A Markowitz-optimal portfolio is one that minimizes portfolio variance $\mathbf{x}'\Sigma\mathbf{x}$ for a given expected target return $\mu'\mathbf{x} = c$, with c an input constant (resp. maximizes return for a given variance), the optimal solution of this quadratic optimization problem under linear side constraint being known in closed form. Among all these optimal portfolios, the so-called *minimum variance portfolio (MVP)*, denoted by $\bar{\mathbf{x}}$ in the sequel, is the one with smallest variance, and it depends solely on Σ (independent of μ, as the constraint $\mu'\mathbf{x} = c$ is omitted):

$$\bar{\mathbf{x}} = \bar{\mathbf{x}}(\Sigma) = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}. \tag{2.6}$$

This approach has been extended in different directions, for example to the optimization of alternative risk or return measures as in Rockafellar and Uryasev (2000), or to the inclusion of nonnegativity or cardinality constraints, or discrete-type constraints related to trading restrictions, which are highly relevant for practitioners as in, e.g., Jagannathan and Ma (2003); Li *et al.* (2006).

Recently, a more descriptive approach to portfolio selection has emerged: Pioneered by some remarkable works by Mantegna, e.g. Mantegna (1999), graph-based methods have found their way into finance literature, and recent studies, for example Onnela *et al.* (2003); Pozzi *et al.* (2013); Kaya (2015); Peralta and Zareei (2016), explore their usefulness for optimal investment purposes. In this context, portfolio selection is essentially

also based on the covariance matrix of returns[20] Σ, but relies on a more descriptive approach compared to the Markowitz paradigm: The matrix Σ is reduced to essential information in the form of a planar graph derived from it, typically the *minimum spanning tree (MST)* as introduced in the context of (S4). The resulting graph structure is used as an easy-to-grasp visualization of the essential aspects of interconnectedness between the assets, and investment decisions are based on the idea of choosing 'central' or 'non-central' assets from the graph, according to certain centrality measures; Intuitively, a variance-minimizing portfolio should consist of rather non-central vertices in this graph because, heuristically, these should form a well-diversified portfolio. This heuristic implies that there should be a significant relationship between centrality measurements in graphs and the MVP.
Indeed, Onnela *et al.* (2003) find empirically that the non-central assets in an MST computed from the historical stock return correlation matrix are prominently represented in the associated MVP. Similarly, Pozzi *et al.* (2013) detect that portfolio performance is improved if the constituent assets are selected among the non-central ones in an MST (or in a maximally filtered planar graph) derived from the correlation matrix. Using the same idea but a slightly differing technique, Kaya (2015) bases his analysis on a matrix containing pairwise mutual information of the assets for a more robust dependence measurement. He finds that more central assets yield higher returns, and concludes that portfolio selection should favor central names with low volatility, which is slightly opposite to the aforementioned references. Peralta and Zareei (2016) study the relation between Markowitz-optimal portfolios and graph-centrality not only empirically, but attempt at providing a heuristic algebraic connection between both concepts. Like Onnela *et al.* (2003); Pozzi *et al.* (2013), they find evidence for Markowitz-optimal portfolios favoring non-central assets. However, they also find that during certain time periods, in which the correlation between individual and systemic performance is high, more central assets gain more weight in Markowitz-optimal portfolios.

In this section we analyze whether there is a fundamental link between variance minimization according to the Markowitz ansatz and (different notions of) centrality in a graph based on the covariance resp. correlation matrix. On the one hand, we investigate potential algebraic links between the two approaches, as postulated in Peralta and Zareei (2016) for example, and find that an algebraic connection cannot be proven for more than 3 assets. On the other hand, we conduct our own study on historical CDS data, which, in line with the previously mentioned studies on stock return data, detects a strong connection between graph centrality and MVP weights. By means of a Monte Carlo study using the different simulation algorithms for correlation matrices introduced in the previous sections we are able to rule out a fundamental connection between graph centrality and MVP weights. Instead, we find that the persistent empirical findings in favor of such a connection must originate from the special structure of financial correlation matrices, as even portfolios constructed from correlation matrices displaying only

[20] Graph-based methods can more generally also be based on any matrix $\Sigma \in \mathbb{R}^{d\times d}$ containing pairwise dependence measurements. To ensure comparability with the classical Markowitz approach, we always take Σ to be the covariance matrix of asset returns.

a subset of the stylized facts (S1)-(S4) of financial correlation matrices did not exhibit a strong link between graph centrality and MVP weights.

In the following, we first introduce the considered centrality concepts in Section 2.4.1, before examining potential algebraic links in Section 2.4.2, and conducting the Monte Carlo studies in Section 2.4.3.

2.4.1 Centrality measures

We investigate the relation between 'central' resp. 'peripheral' assets in the graph associated with Σ and their weights in the corresponding MVP. An intuitive way of identifying non-central assets in a tree is to consider its **leaves**, i.e. vertices with only a single neighbor. We further consider the following, more sophisticated definitions of centrality:

- **Eigenvector centrality of a graph:** The adjacency matrix A of a finite connected graph is irreducible and has entries in $\{0,1\}$ with $a_{ij} = 0$ (resp. $a_{ij} = 1$) meaning that there is no (resp. an) edge between vertex i and vertex j. By the Perron–Frobenius Theorem, the largest eigenvalue of A is positive and the associated eigenvector $\mathbf{v}_1$ has positive components. Consequently, by normalizing $\mathbf{v}_1$ in such a way that $\mathbf{v}_1' \mathbf{1} = 1$, the dominant eigenvector $\mathbf{v}_1$ gives a probability distribution on the vertices. These probabilities can be interpreted as measurements of centrality in the graph, since $\mathbf{v}_1$ is the limit of $A^n \mathbf{1} / \mathbf{1}' A^n \mathbf{1}$, i.e. the normalized version of $A^n \mathbf{1}$, as $n \to \infty$. The i-th entry of $A^n \mathbf{1}$ gives precisely the number of all paths in the graph of length n starting at i (including stopovers, i.e. all paths of length $\leq n$ without stopovers). Consequently, the largest entry of $\mathbf{v}_1$ corresponds to the vertex from which most different paths are possible, i.e. which is most connected to other vertices.

  While this centrality notion is originally based on unweighted (and interesting only for incomplete) graphs, Peralta and Zareei (2016) heuristically extend it to the weighted graph $G_w(\Sigma)$ replacing (a_{ij}) by $(w(\Sigma_{ij}))$ for increasing w, see Section 2.4.2 for details and comments.

- **Mean occupation layer of a tree:** The central vertex of a tree T according to this criterion is defined as the vertex $r(T)$ minimizing the so-called *mean occupation layer*

$$\ell(T) := \frac{1}{d} \sum_{v=1}^{d} \mathcal{L}\big(r(T), v, T\big),$$

  where

$$\mathcal{L}(r, v, T) := \text{length of (unique) tree-path from } r \text{ to } v \text{ in terms of edges passed through,}$$

cf. Onnela *et al.* (2003). Intuitively, $\ell(T)$ gives the average length of a path in T from its root $r(T)$ to a vertex, and the root $r(T)$ is chosen such that this average length is minimal. Later on, we will apply this notion to a minimum spanning tree $\mathrm{MST}(C)$ derived from a correlation matrix C. In this case, we will abbreviate $r(C) = r(\mathrm{MST}(C))$ and $\ell(C) = \ell(\mathrm{MST}(C))$.
This notion is closely related to the concept of *closeness centrality* as explained in Newman (2008), where each vertex is assigned the average path length to all other vertices.

For other centrality concepts the interested reader is referred to Newman (2008); Kaya (2015).

It is important to note that the central vertex of a tree computed via the notion of eigenvector centrality may be different from the one computed via the notion of mean occupation layer, as Figure 2.8 shows.

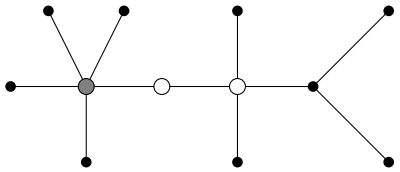

Figure 2.8: Central nodes implied by eigenvector centrality (gray) and mean occupation layer (white) may differ.

Further, Figure 2.8 shows that the central vertex according to the mean occupation layer criterion need not be unique. However, the number of MSTs computed from correlation matrices simulated from the uniform distribution $C \sim \mathcal{U}(\mathcal{C}_d)$ exhibiting this property vanishes for large d, and if the central vertex was not unique, the candidates for the central vertex were always two neighboring vertices.

2.4.2 Does an algebraic connection between centrality and MVP weights exist?

Eigenvector centrality

We first consider a possible relation between eigenvector centrality and MVP weights, which can be approached from the viewpoint of matrix algebra, cf. Peralta and Zareei (2016). By means of an eigenvalue decomposition of the correlation matrix $C = V\Lambda V'$, the MVP (2.6) associated with the covariance matrix $\Sigma = \mathrm{diag}(\sqrt{\Sigma_{ii}})\, C\, \mathrm{diag}(\sqrt{\Sigma_{ii}})$ can be rewritten as follows:

$$\bar{\mathbf{x}}(\Sigma) = \frac{\mathrm{diag}(1/\sqrt{\Sigma_{ii}})}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \left(\sum_{k=1}^{d} \frac{1}{\lambda_k} \mathbf{v}_k \mathbf{v}_k' \right) \mathrm{diag}(1/\sqrt{\Sigma_{ii}})\mathbf{1}, \tag{2.7}$$

where again $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d \geq 0$ denote the eigenvalues of C with associated orthonormal basis of eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_d$.

In the definition of eigenvector centrality, the entries of the dominant eigenvector associated with the adjacency matrix A of an unweighted connected graph are positive and allow to be interpreted as measurements of centrality. However, Peralta and Zareei (2016) consider the weighted graph $G_w(C)$ derived from a correlation matrix and relax the notion of eigenvector centrality in an intuitive, but algebraically questionable way. They consider the entries of the dominant eigenvector $\mathbf{v}_1$ of C (instead of A) as measurements of centrality in $G_w(C)$, although these entries need not be positive. Indeed, as we have shown in Theorem 2.3.9, the percentage of such matrices is $1/2^{d-1}$ in $\mathcal{C}_d$, hence declining fast with increasing d. This renders the interpretation of the entries of $\mathbf{v}_1(C)$ as measurements of centrality less intuitive. However, almost all empirical correlation matrices exhibit this stylized fact (S2), a dominant eigenvector with positive entries, cf. also Boyle *et al.* (2014). We are able to confirm this finding in our data set described in Section 2.4.3 consisting of 395 CDS time series: When considering the correlation matrices of randomly chosen portfolios of 20 assets, over 99.9% exhibited dominant eigenvectors with only positive entries.

Peralta and Zareei (2016) represent the minimum variance portfolio (2.7) as the sum of three parts:

$$\bar{\mathbf{x}}(\Sigma) = \frac{\operatorname{diag}(1/\sqrt{\Sigma_{ii}})}{\mathbf{1}'\Sigma^{-1}\mathbf{1}} \Big(I + N + R\Big),$$
$$I = \begin{bmatrix} \frac{1}{\sqrt{\Sigma_{11}}} \\ \vdots \\ \frac{1}{\sqrt{\Sigma_{dd}}} \end{bmatrix}, \quad N = \left(\frac{1}{\lambda_1} - 1\right)(\mathbf{v}_1' I)\,\mathbf{v}_1, \quad R = \sum_{k=2}^{d} \left(\frac{1}{\lambda_k} - 1\right)(\mathbf{v}_k' I)\,\mathbf{v}_k.$$

The term I is interpreted as individual performance part, because its i-th entry is decreasing in the volatility $\sqrt{\Sigma_{ii}}$ of asset i, while the term N is interpreted as containing information about the location of asset i in the network, and R is a remainder part. In their Corollary 1, from this representation Peralta and Zareei (2016) draw the conclusion that under the conditions

$$\lambda_1 > 1 \quad \text{and} \quad \mathbf{v}_1' I > 0, \tag{2.8}$$

non-central assets in $G_w(C)$ receive large weights in the minimum variance portfolio.

The given conditions (2.8) are introduced purely for technical reasons, namely to ensure that N has negative entries and the centrality measurements in $\mathbf{v}_1$ enter the MVP with negative sign. However, a closer look reveals that these conditions hold true for almost all correlation matrices:

- Since the eigenvalues of a correlation matrix are all real, non-negative, and sum up to its dimension, $\lambda_1 > 1$ holds almost surely. The only possible case of $\lambda_1 \leq 1$ is $\lambda_1 = \ldots = \lambda_d = 1$, which corresponds to having the identity as correlation matrix, and this case almost surely never happens.

- The condition $\mathbf{v}_1' I > 0$ is also fulfilled for almost all correlation matrices: If $\mathbf{v}_1' I < 0$, it suffices to take $-\mathbf{v}_1$ instead of $\mathbf{v}_1$. This, too, is an eigenvector corresponding to the largest eigenvalue, and orthogonal to all others. So the only case in which this condition is not fulfilled is $\mathbf{v}_1' I = 0$. For any given set of eigenvalues, with a similar argument as in the proof of Lemma 2.3.2, one finds that the set of eigenvector matrices for which this holds true is a lower-dimensional subset of the set of feasible eigenvector matrices of correlation matrices.[21] Thus, for an arbitrary correlation matrix, the condition $\mathbf{v}_1' I \neq 0$ is a.s. satisfied.

Further, Peralta and Zareei (2016) do not discuss the influence of the remainder term R in their decomposition of the MVP, which can be quite large and indeed offset the influence of the network centrality related part N, as we illustrate in Example 2.4.1 below. Indeed, according to Laloux *et al.* (1999), 'the composition of the least risky portfolio has a large weight on the eigenvectors with the smallest eigenvalues', as can be adumbrated also from the formulas for R and N, which contain the eigenvalues in the denominator.

Example 2.4.1 (An example in $d = 5$)
Consider the 5-dimensional correlation matrix

$$\Sigma = C = \begin{bmatrix} 1 & 0.2 & 0.4 & 0.1 & -0.3 \\ 0.2 & 1 & 0.4 & 0.1 & -0.1 \\ 0.4 & 0.4 & 1 & -0.7 & -0.2 \\ 0.1 & 0.1 & -0.7 & 1 & 0 \\ -0.3 & -0.1 & -0.2 & 0 & 1 \end{bmatrix}.$$

It can easily be checked numerically that C is positive definite and has leading eigenvalue $\lambda_1 = 1.9646$. The corresponding eigenvector is

$$\mathbf{v}_1 = (0.4035,\ 0.3517,\ 0.6737,\ -0.4106,\ -0.3016)',$$

and the condition $\mathbf{v}_1' I = \mathbf{v}_1' \mathbf{1} > 0$ is fulfilled. The MVP $\bar{\mathbf{x}} = \bar{\mathbf{x}}(\Sigma)$ is given and decomposed as

$$\bar{\mathbf{x}} = \begin{bmatrix} -0.1466 \\ -0.1828 \\ 0.6299 \\ 0.5548 \\ 0.1446 \end{bmatrix} = 0.0809 \left(\underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{I} + \underbrace{\begin{bmatrix} -0.1420 \\ -0.1238 \\ -0.2371 \\ 0.1445 \\ 0.1062 \end{bmatrix}}_{N} + \underbrace{\begin{bmatrix} -2.6701 \\ -3.1353 \\ 7.0234 \\ 5.7131 \\ 0.6816 \end{bmatrix}}_{R} \right).$$

[21] As can be seen from the proof of Lemma 2.3.2, for any fixed set of eigenvalues, the set of feasible eigenvector matrices $\mathcal{V}_\lambda$, i.e. the set of such orthogonal matrices V with $V\Lambda V'$ a correlation matrix, $\Lambda = \operatorname{diag}(\lambda)$, can be locally identified with (a subset of) $\mathbb{R}^{d(d-1)/2-(d-1)}$. Depending on the exact entries of I, the set of $V \in \mathcal{V}_\lambda$ satisfying also this additional constraint $\mathbf{v}_1' I = 0$ is a lower-dimensional subset of $\mathcal{V}_\lambda$ for almost all I. (Indeed, only for $I = \lambda$ it is not.)

It is observed that the largest weight in the MVP is assigned to asset 3, the most central asset according to the entries of the first eigenvector, as the remainder part R offsets the negative influence of the centrality measurements in part N.

Remark 2.4.2 (Further related work)
Many authors argue that the (normalized) dominant eigenvector, i.e. the eigenvector corresponding to the largest eigenvalue, of the correlation matrix C of stock returns provides a reasonable proxy for the so-called *market portfolio*; see the references in Boyle *et al.* (2018, 2014). The i-th component of the latter by definition equals the market share of asset i (among the d assets considered). The market portfolio plays an important role in Markowitz theory and the capital asset pricing model (CAPM). According to the mean-variance tautology in Roll's critique, cf. Roll (1977), the market portfolio lies on the efficient frontier (i.e. it is mean-variance efficient) if and only if the CAPM holds. This means that under the assumption of the CAPM framework, the market portfolio is mean-variance efficient in the sense of Markowitz. Apparently the market portfolio has non-negative components, while the dominant eigenvector of an arbitrary correlation matrix can have negative components, see Boyle *et al.* (2014) for examples and a thorough investigation of this issue. This shows that the dominant eigenvector in general is not equal to the market portfolio, and the aforementioned findings are merely approximations that work well empirically.

Leaves and mean occupation layer

The arguments presented in the previous paragraph already raise first doubts regarding a fundamental relation between centrality on a graph associated with the covariance matrix and the corresponding MVP weights. Whereas we have just dealt with a 'weighted' centrality measure on the complete graph $G_w(\Sigma)$, in the following we will focus on leaves and closeness centrality on the associated MST. Empirical findings are in favor of an existing relation between MST and MVP. Based on historical data, Onnela *et al.* (2003); Pozzi *et al.* (2013) find that non-central assets in MST(C) dominate Markowitz-optimal portfolios. For instance, it is claimed that 'the companies of the minimum risk Markowitz portfolio [MVP] are always located on the outer leaves of the [minimum spanning] tree', cf. (Onnela *et al.*, 2003, p. 1).

The following lemma shows that at least in the simplest case $d = 3$ there is a fundamental relation between MVP weights and the MST, if variances are ignored and only a correlation matrix is considered. The statement remains valid also for covariance matrices as long as all their diagonal entries are identical.

Lemma 2.4.3 (MVP and MST for $d = 3$)
Consider a 3×3 correlation matrix $C \in \mathcal{C}_3$.

(a) The MVP associated with the matrix $\Sigma = C$ is $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \bar{x}_3)'$, where

$$\bar{x}_i := \frac{(1 - C_{jk})(1 + C_{jk} - C_{ij} - C_{ik})}{D},$$
$$\text{with } (i, j, k) \text{ some permutation of } (1, 2, 3),$$
$$D := 4\,(1 - C_{13})\,(1 - C_{23}) - (1 + C_{12} - C_{23} - C_{13})^2.$$

(b) Let T be an MST associated with C, computed from $\mathrm{G}_w(C)$ with a decreasing weight function w. The unique[22] central vertex of T corresponds to the minimum weight in the MVP. More formally, letting $\{1, 2, 3\} = \{i, j, k\}$, if $C_{ij} = \min\{C_{12}, C_{13}, C_{23}\}$, then $\bar{x}_k = \min\{\bar{x}_1, \bar{x}_2, \bar{x}_3\}$.

Proof

(a) Straightforward computation shows the claim:

$$C^{-1} = \frac{1}{\det(C)} \begin{pmatrix} 1 - C_{23}^2 & C_{13}C_{23} - C_{12} & C_{12}C_{23} - C_{13} \\ C_{13}C_{23} - C_{12} & 1 - C_{13}^2 & C_{12}C_{13} - C_{23} \\ C_{12}C_{23} - C_{13} & C_{12}C_{13} - C_{23} & 1 - C_{12}^2 \end{pmatrix},$$
$$\det(C) \cdot \mathbf{1}'C^{-1}\mathbf{1} = D$$
$$\bar{\mathbf{x}} \overset{(2.6)}{=} \frac{C^{-1}\mathbf{1}}{\mathbf{1}'C^{-1}\mathbf{1}} = \frac{1}{D} \begin{pmatrix} (1 - C_{23})(1 + C_{23} - C_{13} - C_{12}) \\ (1 - C_{13})(1 + C_{13} - C_{23} - C_{12}) \\ (1 - C_{12})(1 + C_{12} - C_{23} - C_{13}) \end{pmatrix}$$

(b) By symmetry, it suffices to verify the statement for $k = 3$, i.e. we may assume wlog. that C_{12} is the smallest entry of C. We also assume wlog. that $C_{23} \geq C_{13}$ (the opposite case is treated symmetrically). We have to show (i) $\bar{x}_3 \leq \bar{x}_2$ and (ii) $\bar{x}_3 \leq \bar{x}_1$. Using part (a) and some basic algebra, the inequality (i) is seen to be equivalent to

$$C_{13}\big(C_{13} - (1 + C_{23})\big) \leq C_{12}\big(C_{12} - (1 + C_{23})\big). \tag{2.9}$$

The function $f_{23}(u) := u\,(u - (1 + C_{23}))$ is a parabola with global minimum at $u_{23} := (1 + C_{23})/2$. Since $C_{12} \leq C_{13} \leq u_{23}$ by assumption, it follows that $f_{23}(C_{12}) \geq f_{23}(C_{13})$, which is equivalent to (2.9), hence to (i). Using part (a) and some basic algebra, the inequality (ii) is seen to be equivalent to

$$C_{23}\big(C_{23} - (1 + C_{13})\big) \leq C_{12}\big(C_{12} - (1 + C_{13})\big). \tag{2.10}$$

The function $f_{13}(u) := u\,(u - (1 + C_{13}))$ is a parabola with global minimum at $u_{13} := (1 + C_{13})/2$. In order to verify (ii), it suffices to verify (2.10), which is equivalent to showing $f_{13}(C_{12}) \geq f_{13}(C_{23})$. If $C_{23} \leq u_{13}$, the assertion follows precisely as in the previous case (i). If not, then we have $C_{12} \leq u_{13} < C_{23}$. Since

[22] In the case $d = 3$ the two leaves i and j of an MST are obviously such that C_{ij} is the minimal entry of C. The MST is unique if this minimal entry is unique.

f_{13} is a parabola, the assertion holds true if and only if $C_{23} - u_{13} \leq u_{13} - C_{12}$, which is equivalent to $C_{23} - 1 \leq C_{13} - C_{12}$. This is a true statement, since the left-hand side is non-positive and the right-hand side non-negative by assumption. □

A statement similar to the one of Lemma 2.4.3(b), algebraically hard-coding a relation between centrality in MST(C) and weights in MVP(C), becomes more difficult to obtain in larger dimensions, as Example 2.4.4 shows in $d = 5$.

Example 2.4.4 (Non-centrality in MST ≠ large weight in MVP)
Consider again the 5-dimensional correlation matrix of Example 2.4.1, whose MVP is given by

$$\bar{\mathbf{x}} = (-0.1466,\ -0.1828,\ 0.6299,\ 0.5548,\ 0.1446)'.$$

In particular, the assets 3 and 4 have by far the largest weights in the MVP. However, it is readily checked that none of these two assets is a leaf in any MST associated with C. There is an MST with leaves 1 and 5, and an MST with leaves 2 and 5.

2.4.3 No fundamental link between centrality and MVP weights: evidence from Monte Carlo studies

In the previous section we have seen that a fundamental link between centrality and MVP weights cannot be established analytically for any of the considered centrality measurements. Instead, the presented examples raise serious doubts about a relation between these concepts.

In the following, we analyze the weights in MVPs as well as the different notions of centrality in graphs derived from the same correlation matrix for correlation matrices with specific features by means of a Monte Carlo study: We simulate $n = 1,000,000$ $d \times d$ correlation matrices for $d \in \{5, 10, 20, 50, 100\}$ from the following algorithms, and consider their corresponding MVPs and MSTs:

1. the uniform distribution on $\mathcal{C}_d$, cf. Section 2.2.1, where the simulated matrices do not display any of the stylized facts (S1)-(S4),
2. the `randcorr` algorithm, cf. Section 2.2.2, with the largest eigenvalue fixed at 40% of total variance and the remaining eigenvalues distributed according to the realistic density (2.5), which generates correlation matrices exhibiting (S1), but none of the other stylized facts,
3. our Algorithm 2.3.4, cf. Section 2.3.2, once with uniformly distributed eigenvalues, such that the generated matrices display only (S2), and once with the largest eigenvalue fixed at 40% of total variance and the remaining eigenvalues distributed

according to density (2.5), such that the generated matrices display stylized facts[23] (S1) and (S2), and on average also display (S3), and

4. factor model correlation matrices, cf. Section 2.2.5.

Further, we conduct an empirical study on historical CDS data, namely 5Y-CDS mid upfront time series of the constituents of the four major credit indices (ITRX EUR, ITRX XO, CDX IG, and CDX HY) observed daily from July 30, 2015 to May 2, 2017, complementing the studies in the literature which are all based on stock return data.

To assess whether peripheral assets according to the considered centrality concepts are overweighted in the corresponding MVPs, we consider the following random variables:

$$e_d(C) := \frac{\text{portfolio weight of the 20\% least } C\text{-eigenvector-central assets}}{0.2},$$
$$f_d(C) := \frac{\text{portfolio weight of } \mathrm{L}(C) \text{ in } \mathrm{MVP}(C)}{|\mathrm{L}(C)|/d},$$
$$h_d(C) := \frac{\sum_{v=1}^{d} \bar{x}_v(C)\, \mathcal{L}\big(r(C), v, C\big)}{\ell(C)},$$

where MSTs and MVPs are both constructed from the (simulated) correlation matrix C, $\mathrm{L}(C)$ denotes the set of all leaves of $\mathrm{MST}(C)$ and $|\mathrm{L}(C)|$ its cardinality, and $\sum_{v=1}^{d} \bar{x}_v(C)\, \mathcal{L}\big(r(C), v, C\big)$ refers to a MVP-weighted variant of the mean occupation layer $\ell(C)$ introduced in Section 2.4.1.

The random variable e_d is designed to assess a potential link between eigenvector centrality and MVP-weights: The numerator of e_d is the sum of the MVP-weights assigned to the 20% least central assets according to Peralta and Zareei (2016)'s version of eigenvector centrality. If the centrality measurements did not play a role in the construction of the MVP, we would expect that these assets get assigned a total weight of about 20% (since MVP weights sum up to 1), so the denominator is chosen in order to normalize e_d. A value of $e_d > 1$ thus indicates an overrepresentation of the 20% least central assets in the MVP.
The random variable f_d targets a potential link between leaves in the MST[24] and MVP-weights: The numerator of $f_d(C)$ gives the MVP-weight of the leaves in $\mathrm{MST}(C)$, while the denominator gives the share of leaves of $\mathrm{MST}(C)$ in all d assets. Intuitively, $f_d(C)$ is > 1 (< 1) if and only if the leaves are over- (under-) represented in the MVP.
Finally, $h_d(C)$ targets a potential relation between mean occupation layer and MVP-weights: It contrasts the MVP-weighted occupation layer in the numerator with the mean occupation layer in the denominator[25], with a value of $h_d(C) > 1$ (< 1) indicating

[23] (S4) is approximately fulfilled only for $d > 1000$, so not present in the simulated matrices for the considered values of d.

[24] For all considered simulation algorithms, and also in the empirical study, the entries of a correlation matrix are a.s. mutually distinct, so the resulting MST is unique.

[25] In case the central vertex is not unique, we randomly select one of the (two neighboring) candidate vertices.

that the MVP overweights (underweights) assets that are non-central according to the mean occupation layer criterion / closeness centrality.

In the empirical study, we also consider the empirical counterparts of these quantities:

$$\tilde{e}_d(\Sigma) := \frac{\text{portfolio weight of 20\% least } C\text{-eigenvector-central assets in MVP}(\Sigma)}{0.2}$$
$$\tilde{f}_d(\Sigma) := \frac{\text{portfolio weight of L}(C) \text{ in MVP}(\Sigma)}{|\mathrm{L}(C)|/d},$$
$$\tilde{h}_d(\Sigma) := \frac{\sum_{v=1}^{d} \bar{x}_v(\Sigma)\, \mathcal{L}\big(r(C), v, C\big)}{\ell(C)},$$

where Σ refers to the covariance matrix of the considered CDS investment log-return time series, C is the correlation matrix associated with Σ, and $\bar{\mathbf{x}}$ are the MVP-weights calculated from Σ.

Uniform distribution

Considering the potential link between eigenvector centrality and MVP-weights via the random variable e_d, we find indeed that there is a large probability $P(e_d > 1)$ for overrepresentation of the 20% least central assets for small d. This is in line with Peralta and Zareei (2016)'s result where they regress MVP weights on centrality measurements and find a significant negative relation. However, in all considered dimensions d there exists a nonempty set of correlation matrices that fulfill Peralta and Zareei (2016)'s technical conditions, and yet exhibit an underrepresentation of the 20% least central assets in the MVP. Moreover, the probability of underweighting these assets in the MVP increases with the dimension of the correlation matrix, cf. Table 2.6 and Figure 2.9, which visualizes the density of $e_d(C)$ as a histogram of its law based on $n = 1,000,000$ independent simulations.

As for a potential relation between leaves and MVP-weights, Figure 2.10 visualizes the density of $f_d(C)$ in terms of a histogram for the law of $f_d(C)$ based on $n = 1,000,000$ independent simulations. We observe that the mean of $f_d(C)$ is indeed greater than 1, indicating that there are more correlation matrices overweighting the leaves in MVP(C) than underweighting them. However, with increasing dimension d the mean $\mathbb{E}[f_d(C)]$ decreases and the probability of underweight $\mathbb{P}(f_d(C) < 1)$ increases. This suggests that there is not really a strong relation between MVP(C) and MST(C) for large d, unless one knows something about the structure of the correlation matrix which rules out certain subsets of $\mathcal{C}_d$. Indeed, there exist many correlation matrices that even underweight the leaves of MST(C) in MVP(C).

Note that, since the denominators of $e_d(C)$ and $f_d(C)$ are positive, there even exist portfolios where the overall weight on the 20% least central assets, or leaves, respectively, is negative, cf. Figures 2.9 and 2.10.

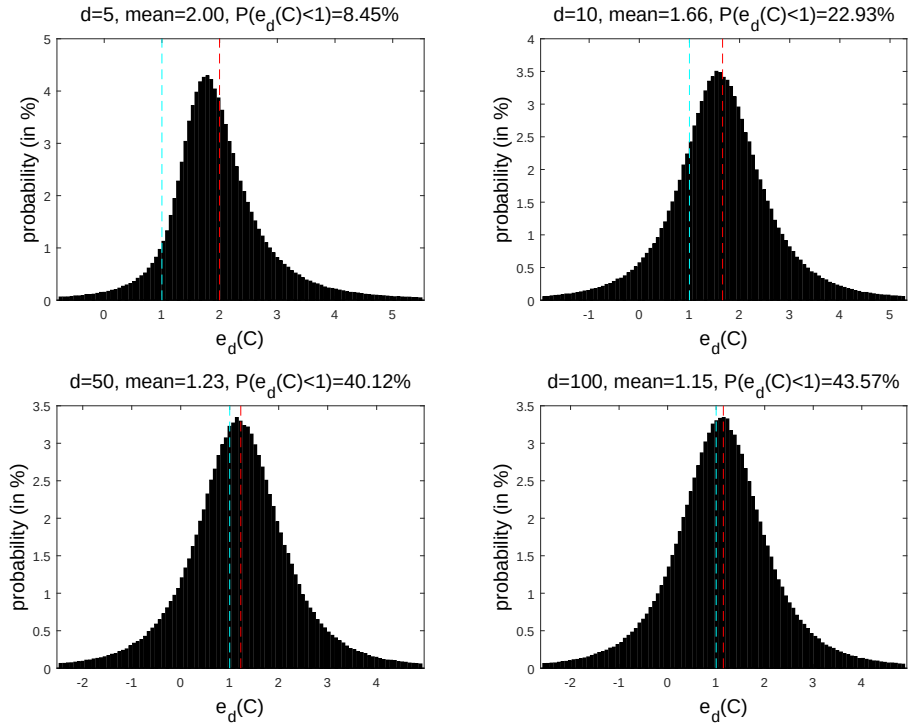

Figure 2.9: Histogram of the probability distribution of $e_d(C)$ with $C \sim \mathcal{U}(\mathcal{C}_d)$ based on $n = 1,000,000$ simulations, for $d \in \{5, 10, 50, 100\}$. The vertical, red line gives the mean, and the blue line represents the border 1 between over- and underrepresentation of the non-central assets.

We have seen that, for a huge number of correlation matrices, the associated MVP is dominated by non-leaf assets. A related, but clearly much weaker question is, whether there exists at least one leaf which is overweighted in the MVP. To this end, instead of $f_d(C)$, we repeat the analysis with the random variable

$$g_d(C) := \max_{B \subset L(C)} \left\{ \frac{\text{portfolio weight of } B \text{ in MVP}(C)}{|B|/d} \right\}, \quad C \sim \mathcal{U}(\mathcal{C}_d).$$

Figure 2.11 shows that the answer to this question is by far more affirmative, i.e. for almost every correlation matrix there is at least one leaf prominently represented in the MVP, and for $d \geq 50$ this statement becomes practically certain. This statement is universal, i.e. follows from the structure of $\mathcal{C}_d$ and has nothing to do with empirical data.

Focusing on the more sophisticated concept of mean occupation layer introduced in Section 2.4.1 to relate centrality in MST(C) and the associated MVP weights, Onnela

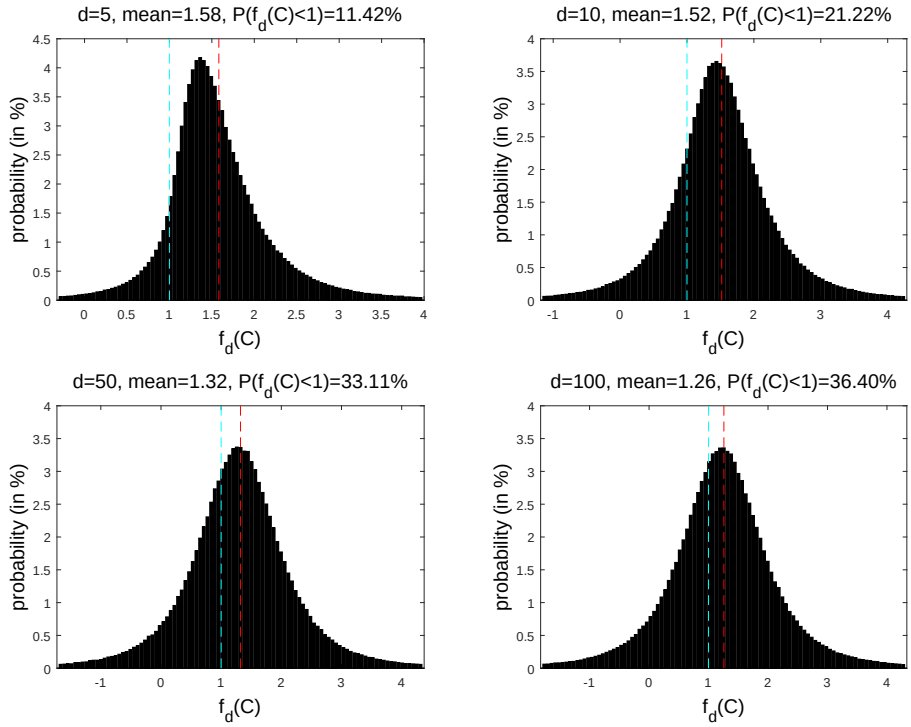

Figure 2.10: Histogram of the probability distribution of $f_d(C)$ with $C \sim \mathcal{U}(\mathcal{C}_d)$ based on $n = 1,000,000$ simulations, for $d \in \{5, 10, 50, 100\}$. The vertical, red line gives the mean, and the blue line represents the border 1 between over- and underrepresentation of the leaves.

et al. (2003) find that the *MVP-weighted portfolio layer*

$$\sum_{v=1}^{d} \bar{x}_v(\Sigma)\, \mathcal{L}\big(r(C), v, C\big)$$

is larger than the mean occupation layer $\ell(C)$. In other words, the MVP assigns more weight to non-central assets than an equally-weighted basket does (i.e. more weight on non-central assets than on central ones). However, studying the random variable $h_d(C)$ for correlation matrices drawn from the uniform distribution, $C \sim \mathcal{U}(\mathcal{C}_d)$, a fundamental relation can not be detected, as shown in Figure 2.12 and Table 2.6: While for a lower number of assets the probability of overweighting non-central assets in the MVP is substantial, this finding is not persistent for larger dimensions. For portfolios consisting of 100 assets, only in about 57% of the cases non-central assets are dominating the MVP.

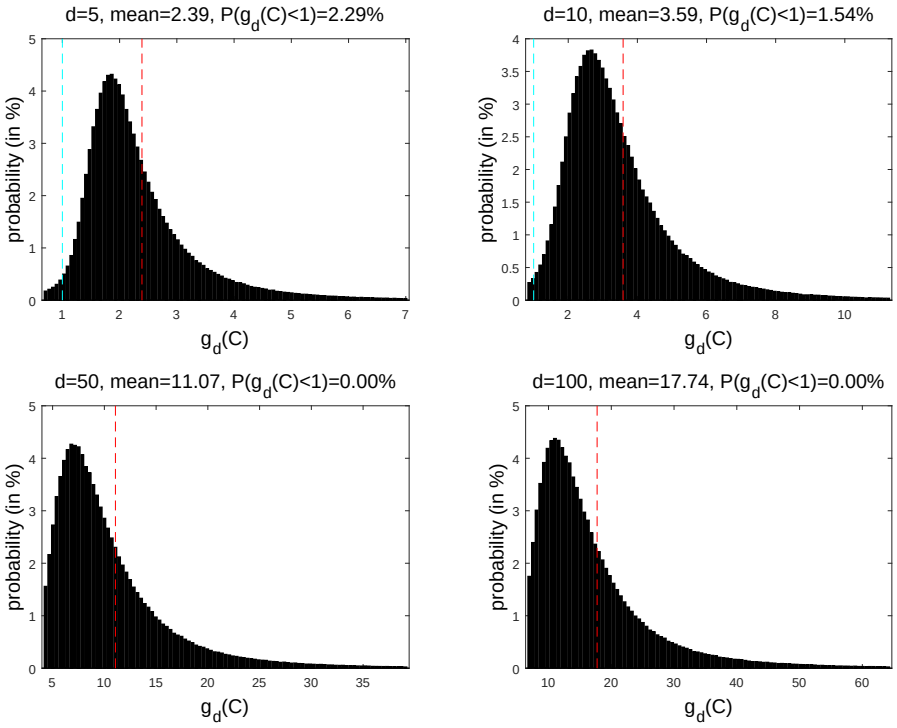

Figure 2.11: Histogram of the probability distribution of $g_d(C)$ with $C \sim \mathcal{U}(\mathcal{C}_d)$ based on $n = 1,000,000$ simulations, for $d \in \{5, 10, 50, 100\}$. The vertical, red line gives the mean, and the blue line represents the border 1 between over- and underrepresentation of the leaves.

Empirical results from CDS portfolios

In the light of the analysis in the previous paragraph, focusing on completely random correlation matrices without any special structure, we conclude that the strong relations between non-centrality in a graph and an MVP, both associated with the correlation matrix C, as observed by Onnela *et al.* (2003); Pozzi *et al.* (2013); Kaya (2015); Peralta and Zareei (2016) are purely data-dependent.
Looking at historical data of credit default swaps (CDS), we are able to confirm this suspicion: We find that large portfolios tend to exhibit a strong overweighting of non-central assets. Portfolios underweighting non-central assets are only found for small to moderate portfolio sizes. Our data set[26] consists of 5Y-CDS mid upfront time series of the constituents of the four major credit indices, namely ITRX EUR, ITRX XO, CDX IG, and CDX HY, observed daily from July 30, 2015 to May 2, 2017. For each asset we consider the trading strategy of selling 5Y CDS protection. Notice that CDS maturities

[26]Source: ICE Data Services.

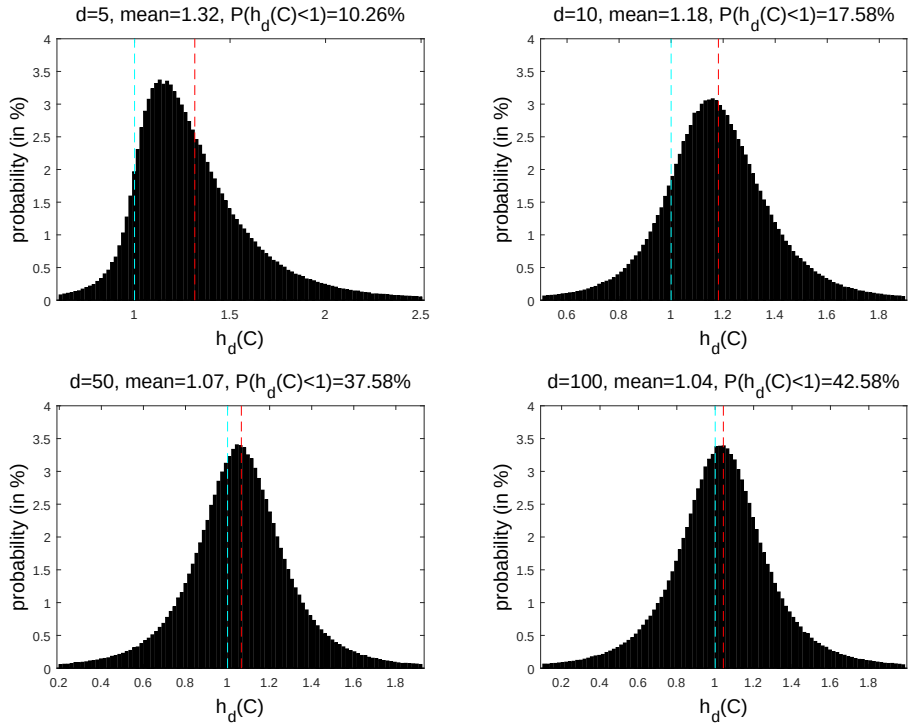

Figure 2.12: Histogram of the probability distribution of $h_d(C)$ with $C \sim \mathcal{U}(\mathcal{C}_d)$ based on $n = 1,000,000$ simulations, for $d \in \{5, 10, 50, 100\}$. The vertical, red line gives the mean, and the blue line represents the border 1 between over- and underrepresentation of the non-central assets.

are standardized to be always on 20 June or 20 December of a year. Furthermore, the observed market price (= upfront) of a 5Y CDS switches from the CDS with maturity in June (December) to the one with maturity in December of the same year (June of the next year) on 20 September (20 March). On these CDS roll dates 20 March and 20 September the trading strategy closes out the old CDS and rolls into the new one, in order to be in accordance with the observed market prices and to keep the duration of the CDS as constant as possible over time. If u_t denotes the upfront of a CDS on day t, we define the log-return at the next day $t+1$ by $\log\left(\frac{1-u_{t+1}}{1-u_t}\right)$. This is because the value $1-u_t$, sometimes called the *bond-equivalent value* of the CDS, can be considered the value of the investment at time t. Clearly, $-u_t$ is the value of the CDS, but the amount 1 needs to be held in cash because it is at stake in case of a potential credit event at t, followed by a CDS auction yielding zero recovery rate.[27] After deleting series with missing data, we are left with 395 assets.

[27] If this value was not held in cash, the investment must be considered levered, which we do not.

Table 2.4 shows that the quantities $\tilde{e}_d(\Sigma)$, $\tilde{e}_d(C)$, $\tilde{f}_d(\Sigma)$, $\tilde{f}_d(C)$, $\tilde{h}_d(\Sigma)$, and $\tilde{h}_d(C)$ are larger than 1 for all four indices, thus indicating an overweighting of leaves, respectively non-central assets according to both the mean occupation layer criterion and Peralta and Zareei (2016)'s eigenvector centrality. The correlation matrices of all indices fulfill both the plausible conditions (2.8) of Peralta and Zareei (2016) and the Perron–Frobenius property.

	$\tilde{e}_d(\Sigma)$	$\tilde{e}_d(C)$	$\tilde{f}_d(\Sigma)$	$\tilde{f}_d(C)$	$\tilde{h}_d(\Sigma)$	$\tilde{h}_d(C)$	constituents
ITRX EUR	3.11	10.31	2.56	4.82	1.35	1.50	123
ITRX XO	1.57	6.22	1.58	2.84	1.07	1.54	64
CDX IG	3.09	6.78	1.70	3.27	1.23	1.64	123
CDX HY	1.89	4.31	1.86	2.28	1.20	1.60	85

Table 2.4: In all major credit indices we detect a systematic overweighting of leaves resp. non-central assets in MVP(Σ) and MVP(C). The number of constituents of these indices (after deleting series with missing data) is given in the rightmost column.

To get a more profound impression, we further calculate $\tilde{e}_d(\Sigma)$, $\tilde{e}_d(C)$, $\tilde{f}_d(\Sigma)$, $\tilde{f}_d(C)$, $\tilde{h}_d(\Sigma)$, and $\tilde{h}_d(C)$ for $n = 1,000,000$ random drawings of $d = 20$ assets out of our pool comprising 395 firms. There are $\binom{395}{d} \approx 2.1547 \cdot 10^{33}$ possibilities, so enough that the probability of choosing the same set twice is negligible. Unlike the aforementioned references and Table 2.4, we cannot confirm a systematic overweighting of non-central assets for arbitrary baskets of CDS when the MVP is calculated from the covariance matrix Σ, cf. Figure 2.13. A significant number of the randomly chosen portfolios exhibits an underweighting of leaves, respectively peripheral assets. An example is given in Table 2.5.
However, for increasing dimension, the probability of finding an overweighting of non-central assets in MVP(Σ) increases, cf. Figure 2.14 and Table 2.6. This aligns with the results of Peralta and Zareei (2016); Pozzi *et al.* (2013); Onnela *et al.* (2003), who study data sets of 200, 300, and 477 stocks, respectively, and with our previous observations for the four major credit indices.
When the MVP is calculated from the correlation matrix C, almost all portfolios exhibit an overweighting of non-central assets, regardless of considered centrality measure, cf. Figure 2.13 and Table 2.6.

These findings indicate that the persistent empirical observation of a strong relation between centrality in a graph and the weights in an MVP derived from the same correlation matrix originates in the special structure of financial correlation matrices as described by stylized facts (S1)-(S4).
Considering the influence of these stylized facts on our quantities e_d, f_d, and h_d, it is hard to anticipate which of these features triggers the completely different behavior of e_d in market and random correlation matrices. For f_d, the scale-free graph structure of empirical correlation matrices implies that the denominator is larger than in the simulated case. However, as we typically observe $f_d > 1$ for empirical correlation matrices, there

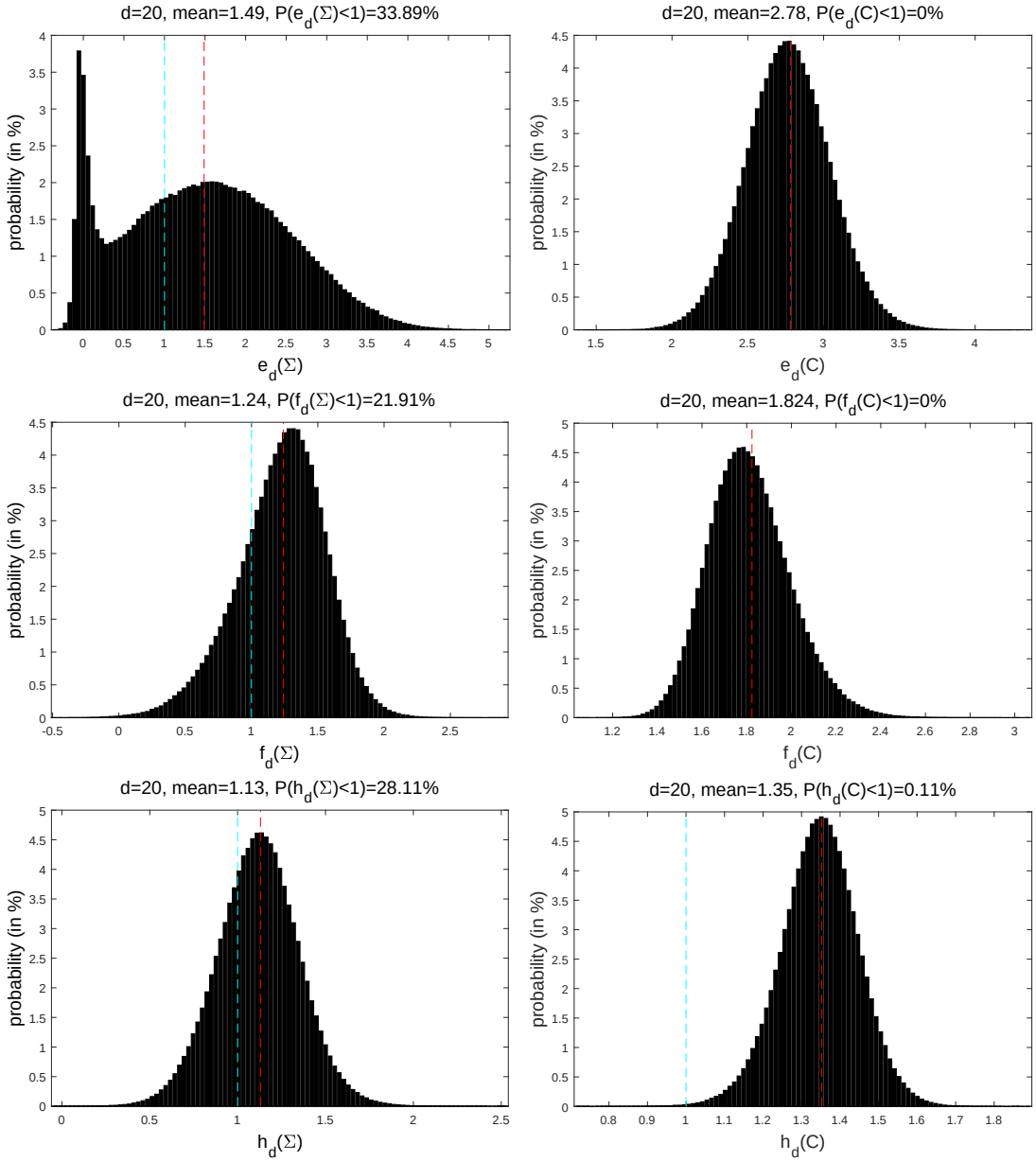

Figure 2.13: Histograms of the probability distributions of $\tilde{e}_{20}(\Sigma)$, $\tilde{e}_{20}(C)$ (top), $\tilde{f}_{20}(\Sigma)$, $\tilde{f}_{20}(C)$ (middle) and $\tilde{h}_{20}(\Sigma)$, $\tilde{h}_{20}(C)$ (bottom) with Σ, C being the covariance resp. correlation matrix of 20 randomly chosen CDS upfront time series out of the 395 considered entities. The vertical, red lines give the respective mean, and the blue lines represent again the border 1 between over- and underrepresentation of non-central assets.

firm	MVP(Σ)-weight	MST-layer	$\mathbf{v}_1$
Centrica	0.0796	1	0.2754
Halliburton	0.0151	3	0.1994
DISH DBS	-0.0336	2	0.2361
Koninklijke Ahold Delhaize	0.2249	1	0.2568
Whirlpool	0.0453	3	0.2113
Meritor	-0.0191	2	0.2069
Quest Diagnostics	0.3829	3	0.2065
BMW	-0.0090	2	0.2308
AirFrance-KLM	-0.0129	2	0.2044
Gas Natural	-0.0971	0	0.2795
Danone	0.2307	1	0.2263
Orange	0.2481	1	0.2623
Ziggo Bond Finance	-0.0389	1	0.2548
Best Buy	-0.0266	4	0.2074
Lincoln National	0.0078	1	0.1809
Deutsche Lufthansa	-0.0034	2	0.2314
Newmont Mining	0.0170	4	0.0387
Stena	-0.0162	2	0.2620
Astaldi	-0.0026	3	0.2118
Nordstrom	0.0079	5	0.1739

Table 2.5: The above portfolio exhibits a systematic underweighting of leaves, respectively peripheral assets, in the covariance-deduced MVP in the considered time period: $\tilde{e}_d(\Sigma) = 0.2392$, $\tilde{f}_d(\Sigma) = 0.8847$, $\tilde{h}_d(\Sigma) = 0.8777$.

must be an even stronger influence on the MVP weights that counters the influence of the higher portion of leaves. To analyze possible effects of the stylized facts on h_d, we first observe that the more leaves a tree structure on d nodes has, the smaller we expect its mean occupation layer to be. This expectation is extrapolated from observations of low-dimensional tree structures. Clearly a 'chain-like' graph has the highest mean occupation layer (with values $d(d+1)/(2d+1)$ for an odd number d of vertices, resp. $d/2$ for an even number of vertices), and a 'star-like' graph has the lowest possible mean occupation layer with value $(d-1)/d$. Therefore, the denominator is smaller in the empirical case, leading to a higher value of h_d. As in the quantity f_d, the numerator is affected both by MVP weights and graph structure. As we observe a higher value of h_d in the empirical case, higher MVP weights on the outskirts of the network compensate for the overall shortening of paths from the central node to the other nodes.

To analyze the effect of the observed features (S1)-(S4) on the quantities e_d, f_d, and h_d in more detail, we rerun the Monte Carlo studies in the following paragraphs, using simulation algorithms that are able to produce correlation matrices that display only a subset of these stylized facts, as opposed to completely random correlation matrices (which display none of them) and market correlation matrices (which exhibit them all).

Correlation matrices with a realistic eigenvalue structure

With the help of the `randcorr` algorithm of Davies and Higham (2000), cf. Section 2.2.2, we are able to analyze the influence of the presence of (S1), a large first eigenvalue. For dimensions $d \in \{5, 10, 20, 50, 100\}$, we generate $n = 1,000,000$ realistic simulations of eigenvalues $(\lambda_1, \ldots, \lambda_d)$, cf. Section 2.3.3: We fix the first eigenvalue $\lambda_1 = 0.4 \cdot d$, according to the typical size of the first eigenvalue in random portfolios drawn from our data set, and simulate $\lambda_2, \ldots, \lambda_d$ according to the special form (2.5) of the power-law type density given in Bouchaud and Potters (2011); Bun *et al.* (2017), which is found to capture the distribution of the bulk of eigenvalues of market correlation matrices fairly well, and rescale in order to cover the remaining 60% of total variance. An artificial spectrum simulated in this way is very similar to the observed spectrum of an arbitrary correlation matrix from our data set. In the next step, we generate for each $(\lambda_1, \ldots, \lambda_d)$ a random correlation matrix C having this particular spectrum according to the `randcorr` algorithm, and calculate $e_d(C)$, $f_d(C)$, and $h_d(C)$. As mentioned in Section 2.2, this procedure is able to reproduce stylized fact (S1), but not the others: Similar to $\mathcal{U}(\mathcal{C}_d)$ the percentage of simulated correlation matrices with the Perron–Frobenius property is small and decreases fast with increasing dimension d. Pairwise correlation entries have a bimodal distribution, symmetric about 0, with mean close to 0 for our simulated correlation matrices. The histogram of leaves for MSTs of correlation matrices with these realistically simulated eigenvalues looks very similar to that obtained from the uniform distribution, so on average graphs derived from correlation matrices simulated from this algorithm exhibit a lot fewer leaves than those derived from market correlation matrices, which hints at stylized fact (S4) also not being present. The results show that just the fact of displaying a realistically large first eigenvalue with a realistic distribution of the spectrum is not enough to explain the empirically observed relation between graph centrality and MVP weights. As for uniformly random simulated correlation matrices, the percentage of correlation matrices simulated according to the `randcorr` algorithm that exhibit a significant overweighting of central assets grows with dimension d, contrary to market correlation matrices where this percentage declines with d, cf. Figure 2.14 and Table 2.6.

Perron–Frobenius correlation matrices

Using our Algorithm 2.3.4 with two different eigenvalue distributions, namely the uniform distribution for eigenvalues, where $\lambda/d \sim U_{\mathcal{S}_d}$, with λ being the vector of eigenvalues and $\mathcal{S}_n$ being the d-simplex, and the power law (2.5) with first eigenvalue fixed at $0.4 \cdot d$, we are able to analyze the influence of stylized fact (S2). As demonstrated in Section 2.3.3, when simulating Perron–Frobenius correlation matrices with uniformly distributed eigenvalues, the generated matrices do not exhibit any other stylized fact than (S2): As our algorithm is a modification of the `randcorr` algorithm, the generated matrices similarly exhibit a small first eigenvalue (with respect to dimension d). Pairwise correlation entries follow a bimodal distribution with antimode at zero, which,

however, is no longer symmetric, but the mean is still close to zero and declining in d. Also stylized fact (S4) is absent, as the low number of leaves in the corresponding MSTs as compared to MSTs of market correlation matrices, and the exponential decay of the degree distribution show, cf. Section 2.3.3.

Regarding the behavior of our quantities $e_d(C)$, $f_d(C)$, and $h_d(C)$, we find that the presence of (S2) significantly lowers the probability of underweighting non-central assets for any of the considered centrality measures: Concerning eigenvector centrality, the simulation study in $e_d(C)$ shows that the probability of underweighting peripheral assets is low in general, with a slight decrease in d recognizable. Concerning leaves and mean occupation layer, both $f_d(C)$ and $h_d(C)$ exhibit a low probability of underweight. However, unlike as for empirical correlation matrices, this probability grows in d, cf. Table 2.6 and Figure 2.14.

Combining our Algorithm 2.3.4 with the eigenvalue distribution (2.5) and a fixed first eigenvalue explaining 40% of total variance yields correlation matrices with more realistic properties, as described in Section 2.3.3: On average, also stylized fact (S3) is present in the simulated matrices, and also the scale-free property (S4) tends to be present for large matrices with $d > 1000$. Here, our quantities $e_d(C)$, $f_d(C)$, and $h_d(C)$ behave as follows: Concerning $e_d(C)$, the probability of underweighting non-central assets is again low, and decreasing in dimension as for empirical correlation matrices. For f_d, probability of underweight is low and slightly increasing in dimension. However, it is slightly higher as for Perron–Frobenius correlation matrices with uniformly distributed eigenvalues. Finally, for h_d, the probability of underweight is lowered even further in comparison to Perron–Frobenius correlation matrices with uniformly distributed eigenvalues, and slightly increasing in dimension, cf. Table 2.6 and Figure 2.14.

This supports our findings of the previous sections that Algorithm 2.3.4 is a promising step towards simulating realistic correlation matrices for financial applications.

Factor model correlation matrices

Despite our concerns that this approach will not produce completely random realizations of correlation matrices with given properties, to gain some insight in whether such correlation matrices may yield a relation between centrality measurements and MVP weights, we rerun our simulation study with factor model correlation matrices:
Following a methodology similar to Fan *et al.* (2008, Section 4), we simulate correlation matrices corresponding to a one-factor model. The distributional characteristics of the parameters are obtained from a fit of a one-factor model to our CDS data set:

$$X_i(t) = b_i M(t) + \epsilon_i(t),$$

where M denotes the market factor (as a proxy we choose an equally weighted portfolio of all assets), $X_i, i = 1, \ldots, 395$ is the i-th time series in our data set, b_i its factor loading, and ϵ_i the associated time series of errors. We find that the factor loadings

b_i and the standard deviations σ_i of the error terms ϵ_i are both approximately gamma distributed, with parameters $\alpha_b = 0.4819$, $\beta_b = 1.6533$, and $\alpha_\sigma = 0.5400$, $\beta_\sigma = 0.0052$, respectively. In the simulation, the market factor is taken to be normally distributed[28], with mean and standard deviation matching the observed values, $\mu_M = 4.7 \cdot 10^{-5}$, and $\sigma_M = 0.0018$, factor loadings and error standard deviations are simulated from the above gamma distributions, and the error time series are simulated independently from normal distributions with zero mean and the respective simulated standard deviations. In $n = 1,000,000$ simulations, we obtain $d \in 5, 10, 20, 50, 100$ time series from a factor model with the above characteristics, and calculate the corresponding (sample) correlation matrices.

The largest eigenvalue of the simulated matrices explains on average about 40% of total variance for $d \in \{10, 20, 50, 100\}$, and about 45% for $d = 5$. Contrary to our expectations, correlation matrices simulated from this one-factor model do not regularly exhibit the stylized fact (S2): The proportion of simulated correlation matrices with Perron–Frobenius property steadily declines, from 62.50% in dimension $d = 5$ to 0.02% in dimension $d = 100$. The mean of pairwise correlations in our simulations is 0.23 on average, so stylized fact (S3) is typically present. The MSTs associated with the one-factor correlation matrices exhibit on average more leaves than those obtained from uniformly random correlation matrices, but fewer leaves than those associated with empirically observed correlation matrices.

Concerning eigenvector centrality, the MVPs related to the simulated factor correlation matrices almost certainly overweight the 20% least central assets, regardless of the number of assets considered, cf. e_d 1-factor in Table 2.6. Also leaves seem to be consistently overweighted: f_d is smaller than 1 for only a low percentage of the simulated correlation matrices, with only a slight growth in dimension. In terms of mean occupation layer, the probability of underweighting peripheral assets, $\mathbb{P}(h_d(C) < 1)$, grows with dimension, from 4.11% in dimension $d = 5$ to 27.71% in dimension $d = 100$. Thus, concerning h_d, correlation matrices simulated from this one-factor model unexpectedly behave similar to correlation matrices simulated uniformly or from the `randcorr` algorithm.

To shed more light on the behavior of the quantities $e_d(C)$, $f_d(C)$, and $h_d(C)$ for factor models, we repeat our analysis with correlation matrices simulated according to the characteristics described in Fan *et al.* (2008, Section 4): There, a Fama-French 3-factor model was fit to daily data of 30 stock portfolios obtained from French's website[29]. Factor loadings were found to approximately follow a trivariate normal distribution, and error standard deviations were found to approximately follow a gamma distribution.

In correlation matrices simulated from this model, we find that stylized facts (S1)-(S3) are present: Regardless of dimension, over 99% of the simulated correlation matrices exhibit the Perron–Frobenius property, the first eigenvector on average explains more

[28] This assumption would be questionable if one intends to describe our data set accurately. However, since we intend to construct time series from an artificial factor model, fitting the exact distribution of the factor returns is not crucial, and we stick with normality for the sake of simplicity.

[29] `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

than 60% of total variance, and the distribution of the pairwise correlation entries is shifted to the positive with a mean of 0.61. We further find that the MSTs associated with 3-factor correlation matrices typically exhibit more leaves than those associated with empirically observed correlation matrices, thus hinting at a denser graph structure than typically exhibited by scale-free trees. This is in line with Bonanno *et al.* (2003)'s finding that one-factor correlation matrices tend to exhibit denser MST structures than market correlation matrices[30].

Concerning the quantities $e_d(C)$, $f_d(C)$, and $h_d(C)$, we find that regardless of dimension d or centrality measure, peripheral assets are almost certainly overweighted in the MVPs associated with the 3-factor correlation matrices, cf. Table 2.6 and Figure 2.14.

2.4.4 Issues of graph-based asset allocation

Having demonstrated that graph-based portfolio selection mechanisms lack a fundamental connection to the traditional Markowitz approach, we briefly want to address potential problems that may arise in the context of graph-based portfolio selection.
As we remarked in Section 2.4.1, the choice of central assets in a MST for a portfolio depends heavily on the choice of centrality measure, and central nodes may differ for different centrality measures, cf. Figure 2.8.
A further aspect, which has not appeared in our study, as we focused on graphs derived from correlation matrices, is that the choice of dependence measure may heavily influence the graph structure, and different graphs may result from different dependence matrices, as illustrated in the next subsection.
Another aspect, which did not appear in our study, as the considered centrality measures did not depend on edge weights, is that the chosen weight function may influence centrality. An example for an edge-weight-dependent centrality measure is a modification of the mean occupation layer criterion of Section 2.4.1, where the length of the tree path is not measured in terms of passed edges, but in terms of the length of these edges.
Finally, it is important to remark that, by just taking into account correlations (or pairwise dependence measures), one loses the information captured by the marginal distribution of the assets or by higher-order dependence structures. An illustration of this is provided below. This criticism, however, applies also to the classic Markowitz approach relying only on the first two moments of the joint distribution of returns.

As a side remark, it is further worth noting that certain graphs derived from the correlation matrix correspond to clustering techniques, e.g. the MST corresponds to single linkage clustering. Issues of clustering-based portfolio selection have been documented e.g. in Lemieux *et al.* (2014).

[30] Bonanno *et al.* (2003) simulate correlation matrices from a one-factor model previously fitted to a large stock data set with the S&P500 index as market factor.

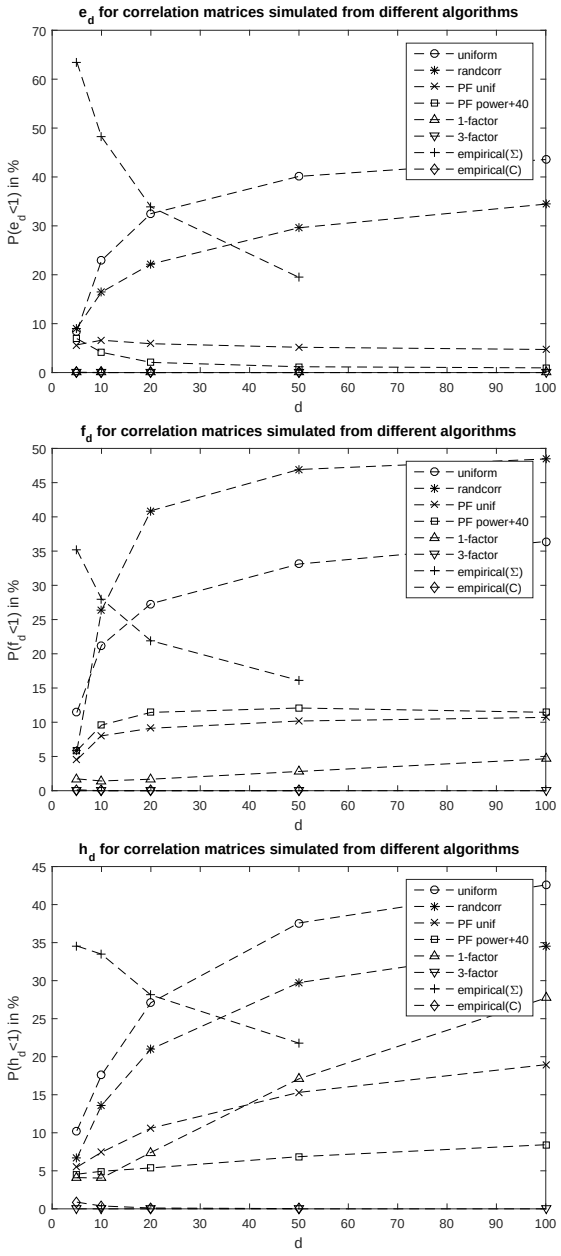

Figure 2.14: Probability of underweight for our quantities e_d (top), f_d (middle), and h_d (bottom), for different types of correlation matrices: uniform, `randcorr`, one-and 3-factor, and empirical.

$\mathbb{P}(\cdot_d < 1)$	$d = 5$	$d = 10$	$d = 20$	$d = 50$	$d = 100$
e_d uniform	8.45%	22.93%	32,50%	40.12%	43.57%
e_d randcorr	9.08%	16.45%	22.18%	29.61%	34.45%
e_d PFC+Unif	5.66%	6.61%	5.90%	5.17%	4.72%
e_d PFC+Power+40	6.94%	4.14%	2.08%	1.19%	0.96%
e_d 1-factor	0.07%	0.01%	0%	0%	0%
e_d 3-factor	0%	0%	0%	0%	0%
$e_d(\Sigma)$ empirical	63.51%	48.21%	33.89%	19.47%	-
$e_d(C)$ empirical	0%	0%	0%	0%	-
f_d uniform	11.42%	21.22%	27.31%	33.11%	36.40%
f_d randcorr	5.79%	26.35%	40.89%	46.92%	48.46%
f_d PFC+Unif	4.50%	8.01%	9.12%	10.18%	10.71%
f_d PFC+Power+40	5.83%	9.58%	11.44%	12.08%	11.45%
f_d 1-factor	1.73%	1.42%	1.68%	2.81%	4.66%
f_d 3-factor	0%	0%	0%	0%	0%
$f_d(\Sigma)$ empirical	35.18%	28.03%	21.91%	16.10%	-
$f_d(C)$ empirical	0.13%	0%	0%	0%	-
h_d uniform	10.26%	17.58%	27.10%	37.58%	42.58%
h_d randcorr	6.64%	13.60%	20.99%	29.74%	34.55%
h_d PFC+Unif	5.48%	7.44%	10.60%	15.31%	18.95%
h_d PFC+Power+40	4.55%	4.94%	5.38%	6.86%	8.43%
h_d 1-factor	4.11%	4.04%	7.41%	17.09%	27.71%
h_d 3-factor	0.03%	0.04%	0.01%	0%	0%
$h_d(\Sigma)$ empirical	34.56%	33.48%	28.11%	21.78%	-
$h_d(C)$ empirical	0.88%	0.38%	0.11%	0.03%	-

Table 2.6: Probability of underweighting non-central assets in terms of $e_d/f_d/h_d$, $\mathbb{P}(e_d/f_d/h_d(C) < 1)$, for different correlation matrices. Whereas the uniform and `randcorr` algorithms produce correlation matrices whose probability of underweighting non-central assets grows with dimension, factor models on the other hand almost certainly overweigh non-central assets.

Graph structure depends on chosen dependence measure

For graph-based portfolio selection methods, any risk measure can be used for the construction of Σ. The resulting dependence matrix will be symmetric, and, unlike in the Markowitz setting, positive definiteness is not required[31] in the selection algorithms. However, one has to keep in mind that different dependence measures may yield different MSTs, as can be seen from the following toy example: Consider $\mathbf{R} = (R_1, R_2, R_3)$, where R_i is lognormally distributed with parameters $\mu_i = 0$ and $\sigma_i > 0$ for $i = 1, 2, 3$, with $\sigma_1 = 0.5$, $\sigma_2 = 0.5$, and $\sigma_3 = 3$. The dependence structure of $\mathbf{R}$ is characterized

[31] Although not required, positive definiteness is a nice-to-have, as in this case the often used correlation distance provides a pseudometric on the set of considered assets.

by a Gaussian copula parameterized by the matrix

$$\begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix},$$

with $\rho_{12} = 0.1$, $\rho_{13} = 0.4$, and $\rho_{23} = 0.8$. Spearman's ρ and Kendall's τ are given as

$$\rho_S(R_i, R_j) = \frac{6}{\pi} \arcsin\left(\frac{\rho_{ij}}{2}\right), \qquad \tau(R_i, R_j) = \frac{2}{\pi} \arcsin(\rho_{ij}).$$

Both are strictly increasing transformations of the ρ_{ij}, so from $\rho_{12} < \rho_{13} < \rho_{23}$ it follows $\rho_S(R_1, R_2) < \rho_S(R_1, R_3) < \rho_S(R_2, R_3)$ and $\tau(R_1, R_2) < \tau(R_1, R_3) < \tau(R_2, R_3)$. On the other hand,

$$\begin{aligned} \mathrm{Cor}(R_1, R_3) &= \frac{(e^{\rho_{13}\sigma_1\sigma_3} - 1)}{\sqrt{(e^{\sigma_1^2} - 1)(e^{\sigma_3^2} - 1)}} \approx 0.017 \\ < \mathrm{Cor}(R_2, R_3) &= \frac{(e^{\rho_{23}\sigma_2\sigma_3} - 1)}{\sqrt{(e^{\sigma_2^2} - 1)(e^{\sigma_3^2} - 1)}} \approx 0.048 \\ < \mathrm{Cor}(R_1, R_2) &= \frac{(e^{\rho_{12}\sigma_1\sigma_2} - 1)}{\sqrt{(e^{\sigma_1^2} - 1)(e^{\sigma_2^2} - 1)}} \approx 0.089, \end{aligned}$$

so constructing the MST from (Pearson) correlations results in a different MST than construction from Spearman's ρ or Kendall's τ, as the central nodes differ, cf. Figure 2.15.

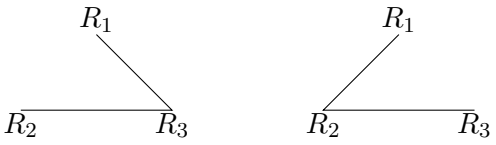

Figure 2.15: MSTs constructed from the different dependence measures using correlation distance as weight function. Left: ρ_S, τ. Right: Cor.

In an example with just three nodes, this may seem a minor issue at first glance. In practice, however, different tree structures for different dependence measures are often encountered, and the differences can be dramatic, as illustrated in Figure 2.16: The two MSTs constructed on return data of the SMI index constituents[32] using correlation resp. Spearman's ρ matrices in combination with a decreasing weight function are fundamentally different. Striking differences are e.g. the position of ZURN, which is rather central in the Spearman's ρ MST, but is a leaf in the correlation MST, or the branch

[32]Data from May 2015 to May 2017; Source: Bloomberg.

descending from BAER (UBSG, CSGN, ADEN), which is located in the center of the correlation MST, but rather peripheral in the Spearman's ρ MST. Table 2.7 presents the eigenvalues of the correlation resp. Spearman's ρ matrices, which differ only marginally, thus indicating that the two matrices are quite similar. The different tree structure is exclusively inferred by the marginal distributions of the return time series, which enter the calculation of the correlation coefficient, but not the calculation of Spearman's ρ. Table 2.7 further shows the annualized volatilities of the time series, indicating that the marginal distributions are diverse.

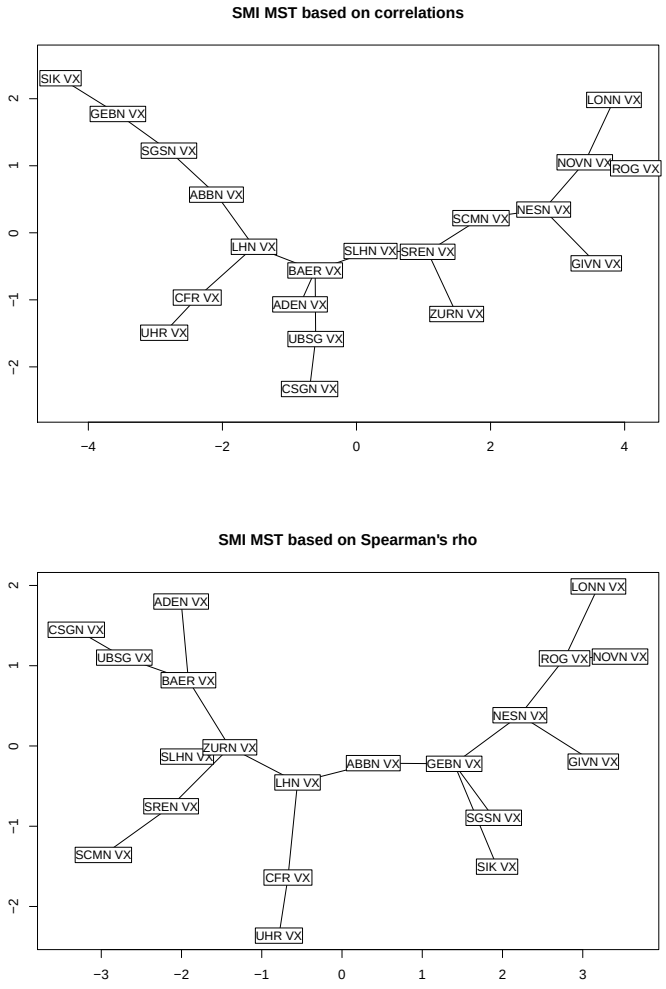

Figure 2.16: MSTs constructed from return data of the SMI Index constituents. Striking differences are for example the respective positions of ZURN and BAER in the networks.

EV corr.	EV rho	firm	vol.
0.1583	0.1461	ABB ('ABBN VX')	0.2046
0.1759	0.1703	Adecco Group ('ADEN VX')	0.2696
0.2139	0.1900	Julius Baer Gruppe ('BAER VX')	0.2777
0.238	0.2292	Cie. Fin. Richemont ('CFR VX')	0.2737
0.2582	0.2662	Credit Suisse Group ('CSGN VX')	0.3719
0.2756	0.2871	Geberit ('GEBN VX')	0.1874
0.2848	0.2976	Givaudan ('GIVN VX')	0.1874
0.3105	0.3281	Lafargeholcim ('LHN VX')	0.3203
0.3654	0.3568	Lonza Group ('LONN VX')	0.2332
0.3763	0.3824	Nestle ('NESN VX')	0.1597
0.3814	0.4001	Novartis ('NOVN VX')	0.1985
0.4133	0.4218	Roche Holding ('ROG VX')	0.1987
0.4343	0.4368	Swisscom ('SCMN VX')	0.1741
0.5092	0.4600	SGS ('SGSN VX')	0.1727
0.5668	0.5451	Swiss Life Holding ('SLHN VX')	0.2073
0.6174	0.5836	Swiss Re ('SREN VX')	0.1908
0.8395	0.8020	Sika ('SYNN VX')	0.2825
1.0862	1.0741	UBS ('UBSG VX')	0.3101
1.4723	1.3879	Swatch Group ('UHR VX')	0.2769
11.0227	11.2347	Zurich ('ZURN VX')	0.2381

Table 2.7: Left: The eigenvalues of the correlation (EV corr.) and Spearman's ρ (EV rho) matrices indicate that the two matrices are quite similar. Right: Annualized volatilities of the SMI return time series.

Influence of variances and higher-order dependence structures

Generally speaking, the 'performance' of a portfolio return $\mathbf{x}'\mathbf{R}$ should be a measurement depending on the full distribution of $\mathbf{R}$. Only taking into account a partial aspect of the latter distribution bears the risk of overlooking better performing portfolios. The first aspect to note is that the considered graph-based portfolio selection approaches are based solely on the correlation matrix C, cf. Onnela *et al.* (2003); Pozzi *et al.* (2013), whereas the MVP is derived from the covariance matrix Σ, i.e. the graph-based methods do not take into account information about the variances of the margins. The latter information has a massive effect on diversification when measured in terms of portfolio variance. In particular, if some components of $\mathbf{R}$ have a variance that is significantly larger than that of others, they are underweighted in the MVP irrespectively of the correlation matrix, which only has a secondary effect, see Example 2.4.5.

Example 2.4.5 (Influence of variances)
Consider the following covariance matrix Σ, with associated correlation matrix C:

$$\Sigma = \begin{pmatrix} 100 & -0.1 & 0 \\ -0.1 & 100 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & -0.001 & 0 \\ -0.001 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is easily seen that $\mathrm{MVP}(\Sigma)$ assigns the highest weight to asset 3, simply because it has by far the smallest variance. However, it is easily checked, cf. Lemma 2.4.3(b), that asset 3 is not a leaf in any MST computed from the associated correlation matrix C. In this artificial example, the effect of the correlation structure is clearly dominated by the effect of the one-dimensional margins (variances). The MST-based selection process simply overlooks the fact that assets 1 and 2 have a high variance.

Furthermore, all graph-based methods (but also Markowitz's approach) essentially rely on dependence information between bivariate pairs only. Consequently, they may be prone to overlook important characteristics of the distribution of $\mathbf{R}$ resulting from higher-order dependence structures beyond those observed through bivariate pair measurements (such as included in Σ). Typically, these effects are of secondary importance in practice, but there are cases in which they do matter, as the following example emphasizes. Consider the following two stochastic models for $\mathbf{R}$, denoted $\mathbf{R}^{(1)}$ and $\mathbf{R}^{(2)}$, which both have exactly the same covariance matrix.

(1) Each $R_i^{(1)}$ is normally distributed with mean $\mu = 0.08$ and standard deviation $\sigma = 0.3$, and the survival copula of $\mathbf{R}^{(1)}$ is given by

$$C(u_1, \ldots, u_d) = u_{[1]} \prod_{k=2}^{d} u_{[k]}^{2^{1-k}},$$

where $u_{[1]} \leq \ldots \leq u_{[d]}$ denotes the ordered list of $u_1, \ldots, u_d$, i.e.

$$\mathbb{P}(R_1^{(1)} > x_1, \ldots, R_d^{(1)} > x_d) = C\left(1 - \Phi\left(\frac{x_1 - \mu}{\sigma}\right), \ldots, 1 - \Phi\left(\frac{x_d - \mu}{\sigma}\right)\right),$$

where Φ denotes the distribution function of a standard normally distributed random variable.

(2) Each $R_i^{(2)}$ is normally distributed with mean $\mu = 0.08$ and standard deviation $\sigma = 0.3$, and the survival copula of $\mathbf{R}^{(2)}$ is given by

$$C(u_1, \ldots, u_d) = u_{[1]} \prod_{k=2}^{d} u_{[k]}^{\frac{1}{2}}.$$

Both copulas are within the family of Lévy-frailty copulas; see Mai and Scherer (2009) for background on the latter, and the standard textbook Nelsen (2006) for background

on copulas in general. Both models are such that $(R_i^{(1)}, R_j^{(1)})$ has the same distribution as $(R_i^{(2)}, R_j^{(2)})$, thus both models share the same covariance matrix Σ. All off-diagonal elements of Σ are equal, as are all its diagonal entries. Consequently, the MVP $\bar{\mathbf{x}}$ is an equally weighted portfolio in both cases. Regarding the portfolio derived from an MST, there is complete freedom. One finds an MST with $k \in \{2, \ldots, d-1\}$ arbitrary leaves: The considered portfolios exhibit constant correlation matrices, i.e. all pairwise correlations are equal. The corresponding complete graph has the same weight on all edges, thus any of its spanning trees is minimal.
While this example shows that the MST-based portfolio selection clearly needs further criteria, how different are the distributions of $\bar{\mathbf{x}}'\mathbf{R}^{(1)}$ and $\bar{\mathbf{x}}'\mathbf{R}^{(2)}$? Figure 2.17 illustrates that the variances and means of both portfolio returns are identical, but the shapes of their distributions differ dramatically. In particular, the second model is negatively skewed and has a significantly larger downside risk than the first.

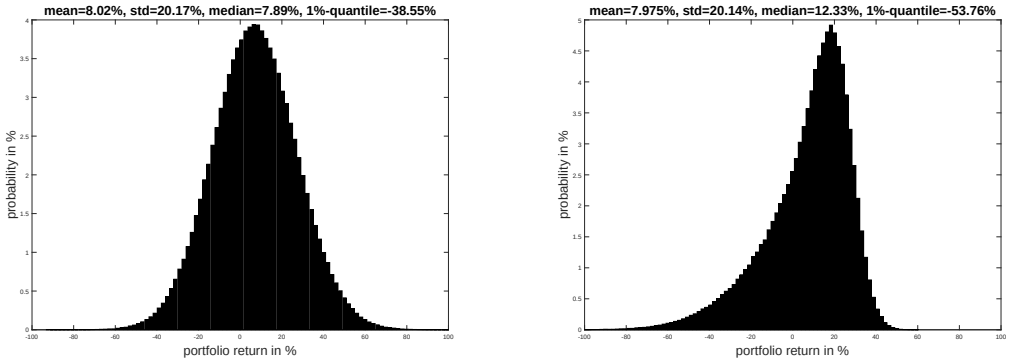

Figure 2.17: Visualization of the probability distribution of the MVP for $d = 20$. Left: model $\mathbf{R}^{(1)}$. Right: model $\mathbf{R}^{(2)}$.

2.5 Conclusion

In this chapter, we considered the realistic simulation of correlation matrices. First, we identified stylized facts (S1)-(S4) of financial correlation matrices, hereby unifying the results of several empirical studies on equity data all focused on single properties and confirming these results in our own study on CDS data. Surveying the literature on simulation algorithms for correlation matrices, we found that the choice of available simulation procedures able to reproduce these stylized facts is very limited: There is only one algorithm available, the `randcorr` algorithm, which is able to reproduce only stylized fact (S1), correlation matrices with a realistic eigenvalue structure. There are no simulation algorithms available which are able to generate completely random correlation matrices exhibiting any of the other stylized facts, or combinations thereof.

From the basis of the `randcorr` algorithm, we develop our Algorithm 2.3.4, which is

able to generate correlation matrices exhibiting stylized fact (S2), the Perron–Frobenius property. It is further able to take into account a realistic eigenvalue structure, i.e. generate correlation matrices which additionally exhibit (S1). We show that our algorithm is able to generate all such matrices (though not uniformly), and, using the construction principle employed in the algorithm, we prove that the proportion of Perron–Frobenius correlation matrices in the set of all correlation matrices is $1/2^{d-1}$ for $d \times d$ correlation matrices.

In a simulation study, we examined the matrices generated from our algorithm for the presence of stylized facts (S3) and (S4), a realistic distribution of pairwise correlations and the scale-free property of the associated MST. We found that a major percentage of the simulated correlation matrices with eigenvalues distributed according to Bouchaud and Potters (2011); Bun *et al.* (2017)'s power law will exhibit also a realistic distribution of pairwise correlations in addition to realistically distributed eigenvalues and the Perron–Frobenius property. Further, when additionally fixing the largest eigenvalue at a realistic value of 40% of the total variance, large (approx. $d > 1000$) correlation matrices simulated from our algorithm tend to exhibit also a power law-like degree distribution in their corresponding minimum spanning trees, as found in several empirical investigations of financial data sets. Thus the presented algorithm is a promising step in the direction of simulating realistic correlation matrices for financial applications.

Finally, we considered a particular use case where the simulation of correlation matrices with these realistic properties is crucial, namely the assessment of graph-based portfolio selection methods. We presented evidence that the persistent empirical observations of a strong relation between Markowitz-optimality and centrality in a graph derived from the related correlation matrix is purely data-dependent. The intuitive assumption that peripheral assets in a graph derived from the covariance/correlation matrix should form a well-diversified portfolio could not be confirmed for completely random correlation matrices: There exist portfolios assigning more weight to central assets for a small to moderate number of assets, and for increasing number of assets, the probability that the MVP associated with the considered correlation matrix underweights peripheral assets grows, seemingly approaching 50% in the limit. Simulation studies with correlation matrices displaying different subsets of the stylized facts (S1)-(S4) indicate that the presence of all stylized facts is required in order to replicate the behavior of the probability of underweight in empirical portfolios.

3 Geostatistical modeling for financial data

3.1 Motivation

Whereas the previous chapter focused on the simulation of financial correlation matrices, this chapter is dedicated to a new modeling ansatz for correlation matrices. It results from a novel approach to the joint modeling of financial assets adapted from geostatistics. Again, our focus lies primarily on CDS data, as the joint modeling of dependent credit spreads is crucial in many applications, ranging from quantitative risk management to portfolio optimization, and also to the pricing of portfolio credit derivatives. Numerous challenges are encountered in modeling and calibration, stemming e.g. from the large number of constituents in several credit portfolios of interest, which may additionally change over time[1], or from data availability and quality, which, unlike for equity data that is readily available for a large number of firms, is often still an issue for credit derivatives like CDS. Hence, desirable is a tractable and parsimonious, yet realistic model for the dependence structure that is able to cope with changing constituents and missing data.

So far, the main focus in the modeling of dependent credit spreads has been on the joint modeling of default times, see Burtschell *et al.* (2009); Meissner (2008) (and references therein) for an overview. However, Cont and Kan (2011) find that the main driver of losses from credit portfolios is not default risk, but rather the risk of a (joint) downturn of the credit spreads. They give a placative example in noting that, under common assumptions on the recovery rate, the loss incurred from the default of a single constituent in an equally weighted index of 125 CDS, like the CDX Investment Grade (IG) Index, roughly equals the loss incurred from a change in CDS spreads corresponding to the 99th percentile of daily changes in the period 2005-2009. To set this into perspective, they note that in the period of 2003-2011, only 8 constituents of all on- and off-the-run CDX IG series have defaulted, whereas a spread change corresponding to said 99th percentile occurred more than twice per year.

Different modeling approaches focusing on spread risk have been proposed in the literature: Cont and Kan (2011), who first noted the importance of modeling spread risk in credit derivatives, focus on accurate modeling of the stylized facts of CDS spreads resp. spread returns via time series models for the single-name spread (return) series and a multivariate t-distribution for the residuals. Oh and Patton (2018) measure systemic risk via CDS spreads utilizing a factor copula model, and Brechmann *et al.* (2013) and

[1] The constituents of the major European and US CDS indices are adjusted semiannually, for example.

Geidosch and Fischer (2016) use vine copula models for stress-testing of CDS and risk modeling of credit portfolios, respectively.
A common issue in these approaches is the need to re-estimate the models each time a new firm enters (or leaves) the portfolio, which may be computationally costly or even impossible in cases where little to no information on the new firm is available, and may change the modeling of the previously included firms. Cont and Kan (2011) resort to quasi maximum likelihood instead of maximum likelihood estimation to circumvent this problem. Further, a complete record of data is typically required for the estimation of the dependence structure, i.e. in case of a missing observation in one series, the whole data set is truncated.

In this chapter, we study an approach adapted from geostatistics that is able to overcome these issues. The essential idea is to model the dependent credit spreads jointly as realizations of a (Gaussian) random field, which is completely determined in terms of its mean and covariance functions. The latter is taken to be a low-parametric function of the distance between the observations, i.e. the underlying firms, where the exact notion of 'distance' in this context remains to be defined. A key benefit of this ansatz is the possibility to include new data points, i.e. to consider new companies in financial applications. Consequently, geostatistical modeling has appealing benefits in the contexts of covariance resp. correlation matrix estimation and missing data imputation that will be discussed in the application section below. The presented results are based on joint work with Benedikt Gräler and Matthias Scherer. An abbreviated discussion focusing primarily on financial applications of geostatistics is available in Hüttner *et al.* (2019).

It has already been discussed in the literature how financial modeling may benefit from spatial approaches. Several studies applying spatial models to stock data, e.g. Asgharian *et al.* (2013); Arnold *et al.* (2013); Fernandez (2011); Fernández-Avilés *et al.* (2012); Kou *et al.* (2016), reveal that spatial dependencies between the underlying entities can provide deeper insights about the dependence structure of stock returns. So far, the main focus of applications lies on contagion and systemic risk. The aspect of spatial dependence inherent in credit spreads has been considered in Blasques *et al.* (2016) and Keiler and Eder (2013). Both studies, and most of the above, employ a different approach than geostatistics, namely spatial autoregressive (SAR) models, which eventually follow a regression approach: Neighboring dependent variables are included in the regression using a spatial weights matrix typically constructed from a distance measure and then row standardized.
Concerning geostatistics, there are only few references applying this framework to financial data: Fernández-Avilés *et al.* (2012) investigate co-movements in stock markets, whereas Arbia and Di Marcantonio (2015) employ such techniques to forecast interest rates.[2]

We find that for CDS data, geostatistics is a promising alternative for the estimation of large covariance/correlation matrices, as it supplies a natural estimator which is guaranteed to be positive definite and is able to provide estimators for CDS correlation

[2]Gaussian random fields have also been applied in interest rate modeling, see Kennedy (1994, 1997).

coefficients of firms with a nonexistent record of CDS spreads. Further, we show that geostatistics supplies a useful tool for missing data imputation which compares favorably to existing approaches.

In the remainder of this chapter, we first introduce the geostatistical modeling ansatz in Section 3.2 and discuss the necessary adjustments when applying it in the high-dimensional framework that financial data sets entail in Section 3.3. In Section 3.4, we demonstrate the benefits of this method in covariance resp. correlation matrix estimation and the imputation of missing data on a data set of 98 time series of CDS par spreads. Section 3.5 discusses benefits and drawbacks of the approach, and offers an outlook on possible future research.

3.2 Introduction to geostatistics

In the following, we introduce the geostatistical modeling approach, which was originally designed to forecast data on two- or three-dimensional surfaces from few sample observations, and then discuss the necessary adjustments for its application in higher dimensions.
At this point, it is important to note that the notion of 'dimension' has a different meaning than in the previous chapter: Whereas we have previously focused on the 'dimension' d of a correlation matrix, i.e. its size or, put differently, the number of assets considered, here 'dimension' $\mathfrak{d}$ refers to the dimension of the underlying coordinate space on which the distance measure is defined. This parameter $\mathfrak{d}$ is of paramount importance for the validity of correlation functions as discussed in Section 3.3, whereas the number of assets d, i.e. the size of the covariance resp. correlation matrix to be estimated plays only a subordinate role.

Throughout this chapter, d denotes, consistently with the previous chapter, the number of considered (sample) locations resp. underlying firms, $\mathfrak{d}$ denotes the dimension of the considered coordinate space, and N is the number of observations of the respective time series.

3.2.1 Classical application in geosciences

For an introduction to geostatistical modeling, it is educational to consider first a typical problem in geosciences, like the prediction of the concentration of a variable of interest at a prespecified location, given measurements of this variable at a set of other locations, for locations in a two- or three-dimensional coordinate space. The described techniques can be executed using the `gstat` package for R, cf. Pebesma (2019) for an introduction. For a more detailed overview and some practical examples of their application in geosciences, see, e.g., Cressie (1993).

The main assumption is that the observations are taken from a random field

$$Z = \{Z(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^{\mathfrak{d}}\}, \quad \mathfrak{d} \leq 3,$$

and dependence is commonly expressed in terms of the so-called *(semi-)variogram* γ, or occasionally the *covariance function* k, defined as

$$\begin{aligned}\gamma(\mathbf{s}_1, \mathbf{s}_2) &:= \frac{1}{2}\mathbb{V}\big[Z(\mathbf{s}_1) - Z(\mathbf{s}_2)\big],\\ k(\mathbf{s}_1, \mathbf{s}_2) &:= \mathrm{Cov}\left(Z(\mathbf{s}_1), Z(\mathbf{s}_2)\right).\end{aligned}$$

The field Z is further assumed to be second-order stationary, which requires that the field has a constant mean and that the covariance, and consequently also the variogram, is a function of the difference between the locations:

$$\begin{aligned}\mathbb{E}\big[Z(\mathbf{s})\big] &= \mu\\ \mathrm{Cov}\left(Z(\mathbf{s}+\mathbf{h}), Z(\mathbf{s})\right) &= k(\mathbf{h}),\\ \frac{1}{2}\mathbb{V}\big[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})\big] &= \gamma(\mathbf{s}+\mathbf{h}, \mathbf{s}) = \gamma(\mathbf{h}), \quad \forall\, \mathbf{s}, \mathbf{h} \in \mathbb{R}^{\mathfrak{d}}.\end{aligned}$$

Second-order stationarity further yields the following relation between γ and k:

$$k(\mathbf{h}) = k(\mathbf{0}) - \gamma(\mathbf{h}). \tag{3.1}$$

If the semivariogram γ depends only on the absolute value $h := \|\mathbf{h}\|$ of the distance vector, i.e. points on spheres around some center have the same semivariogram value, the model is called *isotropic*. Isotropy is often an unrealistic assumption in real world situations, but in many cases a linear transformation of the underlying space restores isotropy in the model:

$$\gamma(\mathbf{h}) = \tilde{\gamma}(\|A\mathbf{h}\|), \quad \mathbf{h} \in \mathbb{R}^{\mathfrak{d}}, A \in \mathbb{R}^{\mathfrak{d}\times\mathfrak{d}}, \tag{3.2}$$

where A is said linear transformation and $\tilde{\gamma}$ is some isotropic semivariogram model. This is referred to as *geometric anisotropy*.

In order to obtain positive semidefinite model covariance matrices for observations made at any set of locations, covariance functions are required to be positive (semi-)definite. Accordingly, semivariogram functions are required to be conditionally negative definite, i.e.

$$\begin{aligned}&\sum_{i=1}^{d}\sum_{j=1}^{d} a_i \cdot k(\mathbf{s}_i - \mathbf{s}_j) \cdot a_j \geq 0, \quad \forall d \in \mathbb{N}, \mathbf{s}_i, \ldots, \mathbf{s}_d \in \mathbb{R}^{\mathfrak{d}}, a_1, \ldots, a_d \in \mathbb{R},\\ &\sum_{i=1}^{d}\sum_{j=1}^{d} a_i \cdot 2\gamma(\mathbf{s}_i - \mathbf{s}_j) \cdot a_j \leq 0, \quad \forall d \in \mathbb{N}, \mathbf{s}_i, \ldots, \mathbf{s}_d \in \mathbb{R}^{\mathfrak{d}}, a_1, \ldots, a_d \in \mathbb{R}, \sum_{i=1}^{d} a_i = 0.\end{aligned}$$

The semivariogram is estimated from observations via its sample version

$$\begin{aligned}\hat{\gamma}(\mathbf{h}) &= \frac{1}{2|N(\mathbf{h})|} \sum_{\mathbf{s}_i,\mathbf{s}_j \in N(\mathbf{h})} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2, \\ \hat{\gamma}(h) &= \frac{1}{2|N(h)|} \sum_{\mathbf{s}_i,\mathbf{s}_j \in N(h)} (Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2,\end{aligned} \tag{3.3}$$

cf. Cressie (1990, p.69), where $N(\mathbf{h})$ resp. $N(h)$ is the set of all sample pairs $\mathbf{s}_i$, $\mathbf{s}_j$ separated by $\mathbf{h}$, resp. by the distance $h = \|\mathbf{h}\|$ in the isotropic case and $|N(\cdot)|$ its cardinality. In practice, sample pairs are binned to achieve a more stable estimator. Once estimated from data and corrected for anisotropy if necessary, a valid model, typically a parametric one[3], has to be fitted to the empirical variogram in order to ensure conditional negative definiteness.

Most parametric isotropic semivariogram models only depend on three parameters τ, σ, and ρ, which define the so-called *nugget* τ^2, *(partial) sill* σ^2, and the *range* $1/\rho$, see Cressie (1993); Arbia and Di Marcantonio (2015). The nugget, named after the mining terminology referring to a naturally formed lump of gold, captures the variation on scales smaller than the minimum distance between sample locations, as this will not be registered in the estimation of the empirical semivariogram. More formally, it is defined as the limit of $\gamma(h)$ for h approaching zero. The sill is the limit of $\gamma(h)$ as h grows large, i.e. the variance of the field, and the range is the value of h where the sill is first reached. Some models reach a sill only asymptotically, in which case the range represents the distance where a certain percentage of the sill is reached. Two examples, the *exponential* and the *Gaussian* semivariogram, are given as follows:

$$\begin{aligned}\gamma_{\text{Exp}}(h) &= \begin{cases} 0 & h = 0, \\ \tau^2 + \sigma^2(1 - \exp(-\rho\, h)) & h > 0, \end{cases} \\ \gamma_{\text{Gau}}(h) &= \begin{cases} 0 & h = 0, \\ \tau^2 + \sigma^2(1 - \exp(-\rho^2\, h^2)) & h > 0. \end{cases}\end{aligned} \tag{3.4}$$

Like most variogram models, these are increasing in distance, i.e. the corresponding covariances are decreasing in distance. Combined with the fact that observed covariances between the sample locations are typically all positive, this reflects Tobler's 'first law of geography', cf. Tobler (1970), which essentially states that observations made at close locations are related more strongly than observations made at distant locations.

The ultimate goal in geoscience applications is to predict the variable of interest at unobserved locations $\mathbf{s}_{\text{new}}$. This is done via *kriging*, named after the mining engineer D. Krige, cf. Cressie (1990), a procedure determining the best unbiased linear predictor $p(\mathbf{s}_{\text{new}})$ minimizing the mean-squared prediction error $\mathbb{E}[(Z(\mathbf{s}_{\text{new}}) - p(\mathbf{s}_{\text{new}}))^2]$. The minimized mean-squared prediction error is referred to as the *kriging variance*.

[3] Non-parametric approaches to variogram fitting exist, cf. Cherry (1996) and references therein, but will not be considered here.

In case the mean of the field Z is known, this is referred to as *simple kriging*, and the simple kriging predictor and kriging variance are obtained as follows:

$$\min_{l_i, i=0,1,\ldots,d} \mathbb{E}\left[\left(Z(\mathbf{s}_{\text{new}}) - p_{\text{simple}}(\mathbf{s}_{\text{new}})\right)^2\right]$$
$$\Leftrightarrow \min_{l_i, i=0,1,\ldots,d} \mathbb{E}\left[\left(Z(\mathbf{s}_{\text{new}}) - \sum_{i=1}^{d} l_i Z(\mathbf{s}_i) - l_0\right)^2\right],$$

$$\Leftrightarrow \; p^*_{\text{simple}}(\mathbf{s}_{\text{new}}) = \mathbf{c}'\Sigma^{-1}(\mathbf{Z} - \mu\mathbf{1}) + \mu,$$
$$\sigma_{\text{simple}}(\mathbf{s}_{\text{new}}) = \mathbb{E}\left[\left(Z(\mathbf{s}_{\text{new}}) - p^*_{\text{simple}}(\mathbf{s}_{\text{new}})\right)^2\right] = k(\mathbf{s}_{\text{new}}, \mathbf{s}_{\text{new}}) - \mathbf{c}'\Sigma^{-1}\mathbf{c},$$

where $\mathbf{c} = (k(\mathbf{s}_{\text{new}}, \mathbf{s}_1), \ldots, k(\mathbf{s}_{\text{new}}, \mathbf{s}_d))'$ is the vector of model covariances between the new location and the sample locations, and $\Sigma = (k(\mathbf{s}_i, \mathbf{s}_j))_{i,j\in\{1,\ldots,d\}}$ is the model covariance matrix at the sample locations.
If the mean of the field is not known, it has to be estimated alongside the best unbiased linear predictor, which is referred to as *ordinary kriging*. For more involved approaches, e.g. more robust approaches when Z is not Gaussian, see Cressie (1993).
A popular example illustrating a typical application of geostatistics is the prediction of concentrations of zinc (and three other pollutants) in a flood plain along the river Meuse, cf. Pebesma (2019), the corresponding data set being supplied in the R package `sp`. Figure 3.1 illustrates the simple kriging predictions for zinc (log-)concentrations throughout the area of interest, obtained from data measured at comparatively few sample locations.

Remark 3.2.1 (Goodness of fit check)
Checking whether the chosen variogram model is a good fit to the data is typically done via *cross-validation*, cf. Cressie (1993, p.101ff). Essentially, one leaves out one data location and re-estimates the variable of interest at this location using the fitted variogram model via kriging for each of the d locations. Then, the following quantities

$$\frac{1}{d}\sum_{i=1}^{d} \frac{Z(\mathbf{s}_i) - p_{\text{simple}}(\mathbf{s}_i)}{\sigma_{\text{simple}}(\mathbf{s}_i)},$$
$$\sqrt{\frac{1}{d}\sum_{i=1}^{d}\left(\frac{Z(\mathbf{s}_i) - p_{\text{simple}}(\mathbf{s}_i)}{\sigma_{\text{simple}}(\mathbf{s}_i)}\right)^2},$$

can be interpreted as the average and mean-square standardized prediction errors, and should take on values of approximately 0 and 1, respectively.

3.2.2 Additional assumptions

Whereas classical geostatistics usually does not make any assumptions on the nature of the considered field, except that inference based on mean and covariance function is

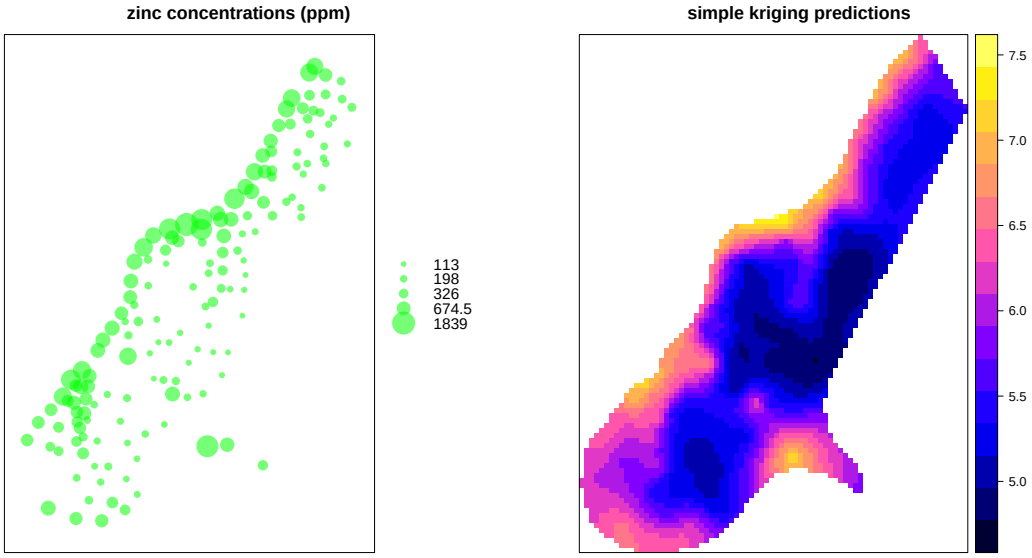

Figure 3.1: Left: Concentrations at the sample locations. Right: Interpolated concentrations obtained from simple kriging throughout the whole area of interest.

'accurate enough', we assume that Z is a *Gaussian* random field, which is completely determined by its mean and covariance function. This has the following nice implications:

- The joint distribution of the observations made at any set of locations is jointly Gaussian, with the covariance matrix essentially being parameterized by the pairwise distances between these locations, i.e. the dependence structure corresponds to a Gaussian copula[4] parameterized by distances.
- For the prediction at a new location, from the fact that the observations made at the sample locations are jointly Gaussian, one obtains that the predictive distribution (in a Bayesian view) at the new location is again Gaussian, cf. Rasmussen and Williams (2006). The simple kriging predictor and the kriging variance are the mean and variance of this predictive distribution.

We further assume that our field Z is isotropic and that its mean is known, i.e. prediction at unobserved locations can be done using simple kriging.

To summarize, geostatistical modeling proceeds as follows:

1. Estimate the sample variogram $\hat{\gamma}$.
2. Fit a valid variogram model γ.

[4] A copula is a d-dimensional distribution function with standard uniformly distributed margins. A Gaussian copula is the copula corresponding to a d-dimensional normal distribution.

3. Calculate model covariances based on distances and the fitted variogram model.
4. Calculate kriging predictors.

For financial applications, we have to add the non-trivial preliminary modeling task:

0. Define/select a suitable (financial) distance measure.

3.3 Adaption to financial data

In order to apply geostatistical methods to financial data sets, several of the modeling steps introduced above have to be adjusted. However, the crucial modeling step in translating geostatistical procedures to financial data is the definition of a suitable coordinate system, respectively distance measure. Informally speaking, we have to provide an answer to the question: How 'far' is company A from company B? In geosciences, one works with data measured at locations on a plane or sphere, and the association among the measurements is usually decreasing in distance. The Euclidean distance between the measurement locations furnishes a natural measure of closeness. For financial data, no such natural distance is available and it is a pivotal modeling step to design a meaningful one. We review several distances proposed in the literature in Section 3.3.1, and present two generally applicable approaches for defining a financial distance measure. Specific financial distance measures constructed along these lines that are tailored to our data set of CDS spreads are presented in Section 3.4.1 below.

After an appropriate distance measure has been defined, a permissible variogram resp. covariance function needs to be fitted to the data. The estimation of the sample variogram has to be adjusted slightly, as unlike in geosciences, many observations are available at each 'location', i.e. firm, cf. Section 3.3.2.
Further, the validity of covariance functions is closely tied to the chosen coordinate system resp. distance measure, and several of the parametric models considered in geoscience applications are no longer valid in the higher-dimensional coordinate systems $\mathfrak{d} > 3$ often necessary to adequately model financial distances. This issue is discussed in greater detail in Section 3.3.3.
Finally, although we have restricted ourselves to isotropic Z, we give a few remarks on how geometric anisotropy may be estimated in higher dimensions using techniques known from machine learning in Section 3.3.4.

3.3.1 Designing appropriate financial distance measures

Several financial distance measures have already been proposed in previous studies. Most of these studies employ SAR models, which incorporate spatial information differently than geostatistics. Nevertheless they are useful for choosing appropriate covariates for designing financial distance measures.

Fernández-Avilés *et al.* (2012) construct a financial distance between two countries based on foreign direct investments (FDI).[5] Using FDI data for the construction of financial distances seems quite promising on country level, but this approach cannot be adapted to corporates as the relevant data is not publicly available. Asgharian *et al.* (2013) investigate the performance of several distance measures, namely exchange rate volatility, interest rate volatility, bilateral trade, and geographical distances, with bilateral trade having the highest explanatory power for spatial dependence. For data sets where the assets are all denominated in the same currency, like our CDS data set introduced in Section 3.4.2, it does not make sense to study exchange rate and interest rate based distances. Similarly to FDI, data on bilateral trade between firms is not publicly available. Keiler and Eder (2013), one of the few studies concerning spatial dependence with a focus on credit spread data, use the (row-standardized) equity correlation matrix as spatial weights matrix in an SAR model. More generally, Fernandez (2011) proposes financial distance measures based on correlations between several covariates, namely the ratio of market cap to firm size, the market-to-book ratio, the ratio of total debt to total assets, debt maturity, and dividend yield.

This ansatz can be applied for the construction of financial distances from time series of covariates: The distance between entities i and j is taken to be their *correlation distance*, cf. Mantegna and Stanley (2000, Ch.13), which is often used as a weight function in the derivation of graphs from a correlation matrix as mentioned in Chapter 2:

$$d(i,j) = \sqrt{2(1-\hat{\rho}_{i,j})}, \tag{3.5}$$

where $\hat{\rho}_{i,j}$ is the sample version of Pearson's correlation coefficient. From the derivation in Mantegna and Stanley (2000), it can bee seen that the correlation distance corresponds to the Euclidean distance[6] between the standardized historical covariate vectors[7] $\tilde{\mathbf{R}}_i$:

$$\begin{aligned} d(i,j)^2 &= \left\| \tilde{\mathbf{R}}_i - \tilde{\mathbf{R}}_j \right\|^2 = \sum_{k=1}^{N} (\tilde{\mathbf{R}}_{i,k} - \tilde{\mathbf{R}}_{j,k})^2 = \sum_{k=1}^{N} \tilde{\mathbf{R}}_{i,k}^2 - 2\sum_{k=1}^{N} \tilde{\mathbf{R}}_{i,k}\tilde{\mathbf{R}}_{j,k} + \sum_{k=1}^{N} \tilde{\mathbf{R}}_{j,k}^2 \\ &= 2(1-\hat{\rho}_{i,j}). \end{aligned}$$

Consequently, the dimension $\mathfrak{d}$ of the coordinate system is very high, as it equals the length of the covariate time series.

Correlation-based distances can only be constructed when covariate time series with a sufficient number of observations are available for each underlying entity, which often excludes covariates that are only reported a few times per year, like balance sheet data. An alternative approach for the construction of financial distance measures from

[5] A major weakness in their analysis is that the chosen 'distance' is not a metric, which may result in model covariance matrices which are not positive definite.

[6] In fact, this is only a pseudometric, as $d(i,j) = 0$ for any two entities i and j with $\hat{\rho}_{i,j} = 1$.

[7] Fernandez (2011) use Spearman's correlation coefficient instead of Pearson's, so their distance measure corresponds to the Euclidean distances between the rank-transformed covariate vectors.

covariates is to directly take several covariates as coordinates and calculate the Euclidean distances between these coordinate vectors. The coordinate systems resulting from this ansatz are lower-dimensional compared to the correlation distance approach, the dimension depending on the number of considered covariates. Like in regression approaches, this might suffer from collinearity in the coordinates. By choosing covariates with moderate correlation among underlying entities, this issue can be minimized. Another issue, whose impact can be quite severe, is the differing scales of the chosen coordinates: If there is one covariate which is a lot larger than the others, the Euclidean distance between the coordinate vectors will be predominantly influenced by this largest figure, thus distorting the spatial influence. At the very least, the coordinates have to be normalized in scale to avoid this issue. Alternatively, one could use the *Mahalanobis distance* between locations $\mathbf{s}_1$ and $\mathbf{s}_2$,

$$d_{\mathrm{M}}(\mathbf{s}_1, \mathbf{s}_2) := \sqrt{(\mathbf{s}_1 - \mathbf{s}_2)' \Sigma^{-1} (\mathbf{s}_1 - \mathbf{s}_2)} = d_{\mathrm{Eucl.}}(\Sigma^{-\frac{1}{2}} \mathbf{s}_1, \Sigma^{-\frac{1}{2}} \mathbf{s}_2), \tag{3.6}$$

where Σ is the covariance matrix of the coordinates and $d_{\mathrm{Eucl.}}$ refers to the Euclidean distance. This corresponds to a 'whitening' of the data, cf. Kulis (2012).

The issue of different scales of the coordinates is quite similar to the issue of (geometric) anisotropy discussed in Section 3.3.4 below.

3.3.2 Sample variogram estimation

In geosciences, one often has only one single observation of the field, or at best a short time series. Financial data, in contrast, is available daily (or at even shorter time intervals), thus many observations of the same field are available. These observations are marred by temporal dependence in the single series, which has to be removed if one wants to take only pure spatial dependence into account.
Different approaches are possible to remove temporal dependence: Our method of choice is working with log-returns, which are often considered to be approximately free of spatial dependence, or with the residuals of different time series models fitted to the log-return series. This is explained in greater detail in Section 3.4.2, where our data set is introduced. Alternatively, one could work with series of weekly averages of log-returns.
Another, more complicated approach is to resort to space-time modeling of the field, which will not be discussed in greater detail here. There, the variogram resp. covariance function is a function of both spatial and temporal distance between the observations, and a crucial question is whether/how to model a potential relation between the spatial and temporal increments.

Having removed temporal dependence, we obtain independent observations $Z(\mathbf{s}_i, t)$ of the field at location i and time t, and we estimate the sample variogram similar as in

Equation (3.3):

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|N} \sum_{\mathbf{s}_i, \mathbf{s}_j \in N(h)} \sum_{t=1}^{N} \big(Z(\mathbf{s}_i, t) - Z(\mathbf{s}_j, t)\big)^2, \tag{3.7}$$

where, as in Equation (3.3), $N(h)$ refers to the set of firm pairs with (absolute value of) distance (vector) $h = \|\mathbf{h}\|$, and N is the number of (daily) observations for each firm.

On the technical side, as the `gstat`'s function for estimating the sample variogram is tailored to coordinate system dimensions $\mathfrak{d} \leq 3$ and expects only a single measurement per location, we have implemented a variant thereof that obtains a distance matrix as input, thus being able to deal with arbitrary coordinate system dimension, and takes the repeated measurements into account for determining the sample variogram. In this context, pairwise complete observations are sufficient for sample variogram estimation. Thus, in case of missing data, we need not exclude days with missing observations completely.

3.3.3 Fitting of a valid variogram model

When fitting a valid variogram model to the sample variogram, we have two major issues to consider: First, compatibility of the variogram model with the metric used, and second, compatibility of the variogram model with the dimension of the considered coordinate space.

Validity in dependence on the metric

Christakos and Papanicolaou (2000) discuss the permissibility of isotropic variogram resp. covariance functions in dependence on the chosen metric: Their key finding is that, under certain conditions, an isotropic covariance function is only valid in combination with the Euclidean norm.

Theorem 3.3.1 (Christakos and Papanicolaou (2000))
Let the covariance function k depend only on the distance between observations, i.e. $k(\mathbf{h}) = \tilde{k}(\|\mathbf{h}\|)$, for some norm $\|\cdot\|$, where $\tilde{k}$ has an even extension on $\mathbb{R}$ which is twice differentiable in an open neighborhood of 0. Then, $\|\cdot\|$ necessarily is the Euclidean norm on $\mathbb{R}^{\mathfrak{d}}$, where $\mathfrak{d}$ is the dimension of the coordinate space.

Note that the Gaussian variogram introduced in (3.4) satisfies the conditions of Theorem 3.3.1. If one decides to use another metric than the Euclidean, one has to check whether the fitted variogram resp. covariance model is indeed compatible with the chosen distance. We circumvent this issue by focusing on Euclidean distances. The Mahalanobis distances introduced above are Euclidean distances on a linearly transformed coordinate space, so these do not pose an issue.

Validity in dependence on $\mathfrak{d}$

The validity of variogram models in dependence on the dimension $\mathfrak{d}$ of the underlying coordinate space is considered, e.g., in Christakos (1984); Cressie (1993). In the case of second-order stationary fields, like the Gaussian fields considered in this chapter, this issue is equivalent to validating a fitted correlation function, and one may draw upon the rich theory available on stationary (isotropic) correlation functions, cf. Schoenberg (1938); Yaglom (1987); Abrahamsen (1997); Steerneman and van Perlo-ten Kleij (2005), to name only a few references. In the following, we review some helpful results for determining the validity of correlation functions in dependence on $\mathfrak{d}$.

Denote

$$\begin{aligned}\mathfrak{B}_\mathfrak{d} &= \{\text{class of correlation functions on } \mathbb{R}^\mathfrak{d}\},\\ \mathfrak{C}_\mathfrak{d} &= \{\text{class of } \textit{stationary} \text{ correlation functions on } \mathbb{R}^\mathfrak{d}\},\\ \mathfrak{D}_\mathfrak{d} &= \{\text{class of } \textit{isotropic} \text{ stationary correlation functions on } \mathbb{R}^\mathfrak{d}\}.\end{aligned}$$

The classes of such correlation functions that are valid in any dimension are denoted by the subscript ∞. It holds that

$$\mathfrak{D}_\infty \subseteq \ldots \subseteq \mathfrak{D}_2 \subseteq \mathfrak{D}_1,$$

i.e. valid isotropic correlation functions in dimension $\mathfrak{d}_2$ are also valid in dimension $\mathfrak{d}_1$ for $\mathfrak{d}_1 \leq \mathfrak{d}_2$, but not vice versa. The statement directly translates to variogram models, cf. Cressie (1993, Ch. 2.5.2). As a consequence, several of the parametric variogram models typically used in geoscience applications, where $\mathfrak{d} \leq 3$, are not valid in higher dimensions.

Indeed, whereas for $\mathfrak{d} \leq 3$ a whole battery of parametric variogram model families is available, the case $\mathfrak{d} > 3$ is not so extensively studied. If one has chosen a large number of coordinates, as might result from the correlation distance (3.5), it might be more convenient to work with correlation functions from (resp. variogram models corresponding to) the well-studied class of correlation functions valid in any dimension $\mathfrak{D}_\infty$. Using R's `gstat` package, this means one is limited to the exponential and Gaussian semivariograms introduced in (3.4). Gaussian semivariograms are shied from in geosciences, as these imply very smooth surfaces of the corresponding field, which is often considered unrealistic in these applications. However, a-priori there is no reason why this should be an unrealistic behaviour for financial data. Indeed, in our use cases in Section 3.4, Gaussian semivariograms fit the data quite well.

In some cases, however, one may want to fit a custom variogram function to the sample variogram obtained. In this case, Bochner's Theorem (and its equivalent for correlation functions, the Wiener–Khinchin Theorem) provides a useful link between positive (semi-) definite functions and measures, cf. Abrahamsen (1997):

Theorem 3.3.2 (Bochner's Theorem)
A function $r : \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}$ is positive (semi-)definite if and only if it admits the representation

$$r(\mathbf{h}) = \int_{\mathbb{R}^{\mathfrak{d}}} e^{i\mathbf{h}'\mathbf{x}} \mathrm{d}^{\mathfrak{d}} F(\mathbf{x}), \tag{3.8}$$

with F a non-negative bounded measure.

In terms of correlation functions, this statement is given as follows:

Corollary 3.3.3 (Wiener–Khinchin Theorem)
A function $\rho : \mathbb{R}^{\mathfrak{d}} \to \mathbb{R}$ is a correlation function in $C_{\mathfrak{d}}$ if and only if

$$\rho(\mathbf{h}) = \int_{\mathbb{R}^{\mathfrak{d}}} e^{i\mathbf{h}'\mathbf{x}} \mathrm{d}^{\mathfrak{d}} F_{\mathbf{X}}(\mathbf{x}) = \mathbb{E}[e^{i\mathbf{h}'\mathbf{X}}], \tag{3.9}$$

where $F_{\mathbf{X}}$ is a $\mathfrak{d}$-dimensional distribution function and $\mathbf{X} \sim F_{\mathbf{X}}$, i.e. any stationary correlation function is the characteristic function of some $\mathfrak{d}$-dimensional random variable $\mathbf{X}$.

For stationary isotropic correlation functions $\rho \in \mathfrak{D}_{\mathfrak{d}}$, a transformation to Polar coordinates and integration over all angles yields

$$\rho(\mathbf{h}) = r(\|\mathbf{h}\|) = \int_0^\infty 2^{\frac{\mathfrak{d}-2}{2}} \Gamma\left(\frac{\mathfrak{d}}{2}\right) \frac{J_{\frac{\mathfrak{d}-2}{2}}(x\|\mathbf{h}\|)}{(x\|\mathbf{h}\|)^{\frac{\mathfrak{d}-2}{2}}} \mathrm{d}G(x) =: \int_0^\infty \Lambda_{\mathfrak{d}}(x\|\mathbf{h}\|) \mathrm{d}G(x),$$

where $G(x) = \int_{\|\mathbf{x}\|<x} \mathrm{d}F_{\mathbf{X}}(\mathbf{x})$ and J is the Bessel function of the first kind, cf. Yaglom (1987) for a detailed proof.
Another representation of $\rho \in \mathfrak{D}_{\mathfrak{d}}$ due to Schoenberg (1938) is

$$\rho(\|\mathbf{h}\|) = \rho(h) = \int_0^\infty \frac{\Gamma(\frac{\mathfrak{d}}{2})}{\sqrt{\pi}\Gamma(\frac{\mathfrak{d}-1}{2})} \int_{-1}^1 e^{ihyv} (1-v^2)^{\frac{\mathfrak{d}-3}{2}} \mathrm{d}v \, \mathrm{d}F_{\|X\|}(y).$$

For continuous $F_{\mathbf{X}}$, the corresponding density f is called the spectral density of ρ and can be obtained as follows for stationary and stationary isotropic correlation functions $\rho \in \mathfrak{C}_{\mathfrak{d}}$ resp. $\rho \in \mathfrak{D}_{\mathfrak{d}}$:

$$f(\mathbf{x}) = (2\pi)^{-\mathfrak{d}} \int_{\mathbb{R}^{\mathfrak{d}}} e^{-i\mathbf{h}'\mathbf{x}} \rho(\mathbf{h}) \mathrm{d}^{\mathfrak{d}}\mathbf{h},$$
$$f(x) = (2\pi)^{-\frac{\mathfrak{d}}{2}} \int_0^\infty \frac{J_{\frac{\mathfrak{d}-2}{2}}(x\|\mathbf{h}\|)}{(x\|\mathbf{h}\|)^{\frac{\mathfrak{d}-2}{2}}} \|\mathbf{h}\|^{\mathfrak{d}-1} \rho(\|\mathbf{h}\|) \mathrm{d}\|\mathbf{h}\|.$$

So to verify if a function is indeed a valid stationary (and isotropic) correlation function in $\mathbb{R}^{\mathfrak{d}}$, a simple check is to verify that the corresponding spectral density is indeed the density of some probability measure.

3.3.4 Geometric anisotropy in higher dimensions

In this section we digress from the main thread of this chapter, where we have assumed isotropy of the considered field Z, and present, for the sake of completeness, a few ideas on how to identify the transformation A that corrects geometric anisotropy and restores isotropy in higher-dimensional coordinate systems $\mathfrak{d} > 3$.

Typically, in $\mathfrak{d} \in \{2, 3\}$, data-driven approaches are employed to estimate the (geometric) anisotropy matrix A, like the estimation of directional variograms, where distance vectors $\mathbf{h}$ pointing in different directions are used for sample variogram estimation along these directions in Equation (3.3). In this context, the sill is assumed to be constant over all directions, only the range varies.[8] A plot of the ranges in their corresponding directions then reveals the elliptic contour of the iso-variogram lines, cf. Figure 3.2 for an illustration, and the affine transformation A that restores isotropy[9] can be deduced:

$$A = \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\lambda \sin(\phi) & \lambda \cos(\phi) \end{pmatrix},$$

where ϕ is the angle between the major axis of the ellipse and the x-axis and λ is the ratio of the major to the minor axis of the ellipse.
The estimation of the directional variograms, however, requires a certain number of data points for each directional variogram to be feasible, a requirement which is getting increasingly hard to fulfill in higher dimensions $\mathfrak{d}$.

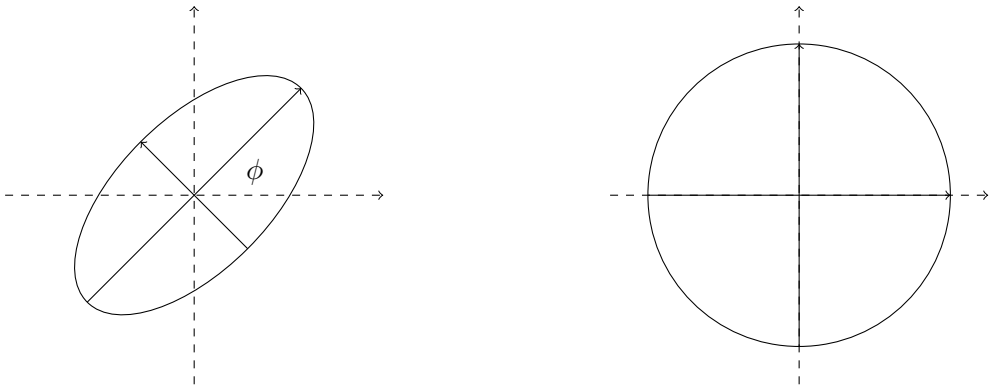

Figure 3.2: Transformation of the original anisotropic plot of the ranges in the considered directions to the isotropic case.

A potential way to overcome this issue is to employ *metric learning*, a different data-driven approach borrowed from machine learning applications, cf. Kulis (2012), for the

[8] The case of different sills in different directions, referred to as zonal anisotropy, is much harder to handle and will not be discussed here.

[9] Here, A aligns the axes of the ellipse with the coordinate axes. In case the original orientation of the coordinate space should be kept, A has to be multiplied from the left with a rotation matrix about $-\phi$.

estimation of geometric anisotropy in higher dimensions. Indeed, linear metric learning exactly corresponds to a formal estimation of the geometric anisotropy matrix A: Given a set of data points in a $\mathfrak{d}$-dimensional coordinate space and information on a more appropriate distance than the Euclidean, typically in the form of class labels or, simply put, information whether two data points are considered similar or not, linear metric learning approaches infer from the data a linear transformation A of the space such that the Euclidean distance in the transformed space is more appropriate for the task at hand. In our setup, this means it should more appropriately reflect Tobler's law. Formally, one learns a *Mahalanobis-type metric* d_A

$$d_A(x, y) = d_{\text{Eucl.}}(Ax, Ay) = \sqrt{x'A'Ay},$$

by minimizing some functional of A, subject to similarity/dissimilarity constraints (distance between similar/dissimilar points should be smaller/larger than some constant) or relative distance constraints ($d_A(i, j) < d_A(i, k)$). Several methods exploit that $A'A$ is positive definite. One sees that variogram estimation under this metric exactly corresponds to the case of geometric anisotropy as specified in (3.2). Further note the similarity to the classical Mahalanobis distance (3.6).
One of the most popular methods is the one proposed in Weinberger and Saul (2009), denoted LMNN (large margin nearest neighbors), which was originally designed to improve k-nearest neighbor classification. The essential idea, visualized in Figure 3.3, is that, after applying the transformation A, any data point should belong to the same class as its nearest neighbors.

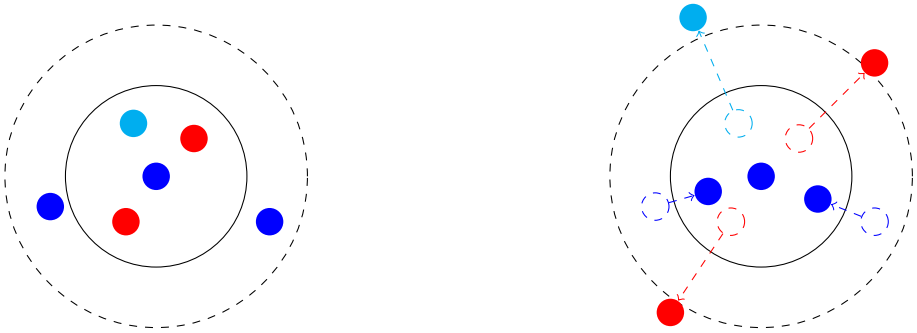

Figure 3.3: Illustration of locations in original and learned metric in LMNN.

Of course, a crucial question in financial applications is then how to define whether two firms are similar or not. A straightforward choice in our CDS data set of Section 3.4 below would be to define classes according to business sectors. In Figure 3.6 one sees that typically firms operating in the same sector are more correlated than firms operating in different sectors, which is also found in Perreault *et al.* (2019).

3.4 Applications

The possible applications of geostatistical techniques to financial data are manifold. Exemplarily, we focus on the problem of covariance resp. correlation matrix parameterization in Section 3.4.3 and on the interpolation of missing data in Section 3.4.4. Another possible application is the prediction of multivariate credit spreads.

In contrast to previous studies, which have predominantly focused on equity data, we carry out our analyses on a data set of liquidly traded corporate CDS, to see whether the joint modeling of dependent credit spreads can profit from the geostatistical approach introduced above. We propose a selection of financial distance measures for CDS data in Section 3.4.1, before introducing our data set and the required preprocessing steps in more detail in Section 3.4.2.

3.4.1 Financial distance measures

To design financial distance measures for CDS spread data, we first select meaningful covariates. A manifest choice is to use equity data, which is well-known to have a strong connection to CDS data; in structural credit models that we use in Chapter 4, this link is particularly evident and may be exploited for the joint valuation of credit and equity of a firm. Keiler and Eder (2013), who study spatial influences in CDS data, albeit in an SAR framework, also use equity data to incorporate spatial information in their model. Similarly to Fernandez (2011), we construct a distance based on equity return correlations (d_{E}) using the correlation distance ansatz (3.5). It corresponds to the Euclidean distance between the equity return vectors, thus the dimension $\mathfrak{d}$ of the coordinate space is very high, as it equals the length of the return time series. Hence, when fitting a valid variogram with `gstat` in R, it is prudent to consider only models corresponding to correlation functions in $\mathfrak{D}_\infty$, i.e. the Gaussian and Exponential variograms (3.4).
A minor drawback is that using d_{E} forces us to drop privately owned companies, which are not listed on stock markets, from our data set. A possible solution to this issue could be to use, as a substitute for each of these companies' equity data, an equally weighted index of equity return data of suitable peer companies. A list of peer companies for a chosen firm is supplied, e.g., by Reuters.

CDS spreads are further strongly associated with the credit quality of their reference entity, i.e. the underlying firm. Therefore we consider several performance figures obtained from the reference entities' balance sheets as covariates. As these are typically reported only once per year, we resort to Euclidean resp. Mahalanobis distances between the covariate vectors. The dimension $\mathfrak{d}$ of the resulting coordinate systems corresponds to the number of considered performance figures. However, as typically $\mathfrak{d} > 3$ the choice of valid variogram models is again limited.
We consider distances based on balance sheet data for changes in working capital (CWC), EBITDA, net debt, capital expenditures (CapEx), and total revenue, all divided by the

respective firm's market cap to minimize the influence of the firms' sizes, as well as a categorization by sector and country. Their correlation matrix is

$$\begin{array}{c} \\ \text{CWC} \\ \text{EBITDA} \\ \text{Net Debt} \\ \text{CapEx} \\ \text{Total Revenue} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{CWC} & \text{EBITDA} & \text{net debt} & \text{CapEx} & \text{total revenue} \end{array} \\ \begin{pmatrix} 1 & -0.3133 & -0.4027 & 0.0737 & 0.2060 \\ -0.3133 & 1 & 0.5873 & -0.9040 & 0.6511 \\ -0.4027 & 0.5873 & 1 & -0.5274 & 0.1809 \\ 0.0737 & -0.9040 & -0.5274 & 1 & -0.6697 \\ 0.2060 & 0.6511 & 0.1809 & -0.6697 & 1 \end{pmatrix} \end{array},$$

with all coordinates taken as ratios to market cap as indicated above. EBITDA and CapEx are strongly correlated, which introduces numerical problems in the computation of the Mahalanobis distance, as the (sector-extended) covariance matrix of coordinates is close to singular. Hence, for the Mahalanobis distance based on balance sheet data and sector information, EBITDA/market cap is excluded from the set of coordinates.[10] Sector and country information are encoded as 0-1 variables.
Due to the specific business concept of firms in the financial sector, their balance sheets differ fundamentally from those of firms outside the financial sector. Hence, it is necessary to remove all firms from the financial sector from our data set, as for a meaningful characterization of these, different balance sheet data is required.

To summarize, the considered distance measures are:

- Correlation distance constructed from equity return data (d_{E}),
- Euclidean distance, considering balance sheet data and sector information as coordinates ($d_{\mathrm{FRS,Eucl.}}$),
- Euclidean distance, considering balance sheet data, sector, and country information as coordinates ($d_{\mathrm{FRSC,Eucl.}}$),
- Mahalanobis distance, considering balance sheet data as coordinates ($d_{\mathrm{FR,M}}$),
- Mahalanobis distance, considering balance sheet data and sector information as coordinates ($d_{\mathrm{FRS,M}}$).

Figure 3.4 illustrates the relation between CDS log return correlations and the considered distance measures in our data set via hexbin plots. For a meaningful distance measure, a decline of CDS correlation with distance is expected, which is clearly visible in the hexbin plots for d_{E}, $d_{\mathrm{FR,M}}$, and $d_{\mathrm{FRS,M}}$, but less pronounced for $d_{\mathrm{FRS,Eucl.}}$ and $d_{\mathrm{FRSC,Eucl.}}$.

Another helpful diagnostic plot is to consider a heatmap of the distance matrices corresponding to d_{E}, $d_{\mathrm{FRS,Eucl.}}$, $d_{\mathrm{FRSC,Eucl.}}$, $d_{\mathrm{FR,M}}$, and $d_{\mathrm{FRS,M}}$, where the rows and columns are sorted in the same way as in the CDS correlation matrix after applying R's internal

[10] For all other distances, no numerical problems arise, and sample variograms with both the EBITDA and CapEx ratio included differ only marginally from those obtained when one of these ratios is excluded from the set of coordinates.

clustering approach (which essentially groups strongly correlated firms). Ideally, blocks of highly correlated firms in the CDS correlation matrix are visible as blocks with low distances in the distance matrices, cf. Figure 3.5.

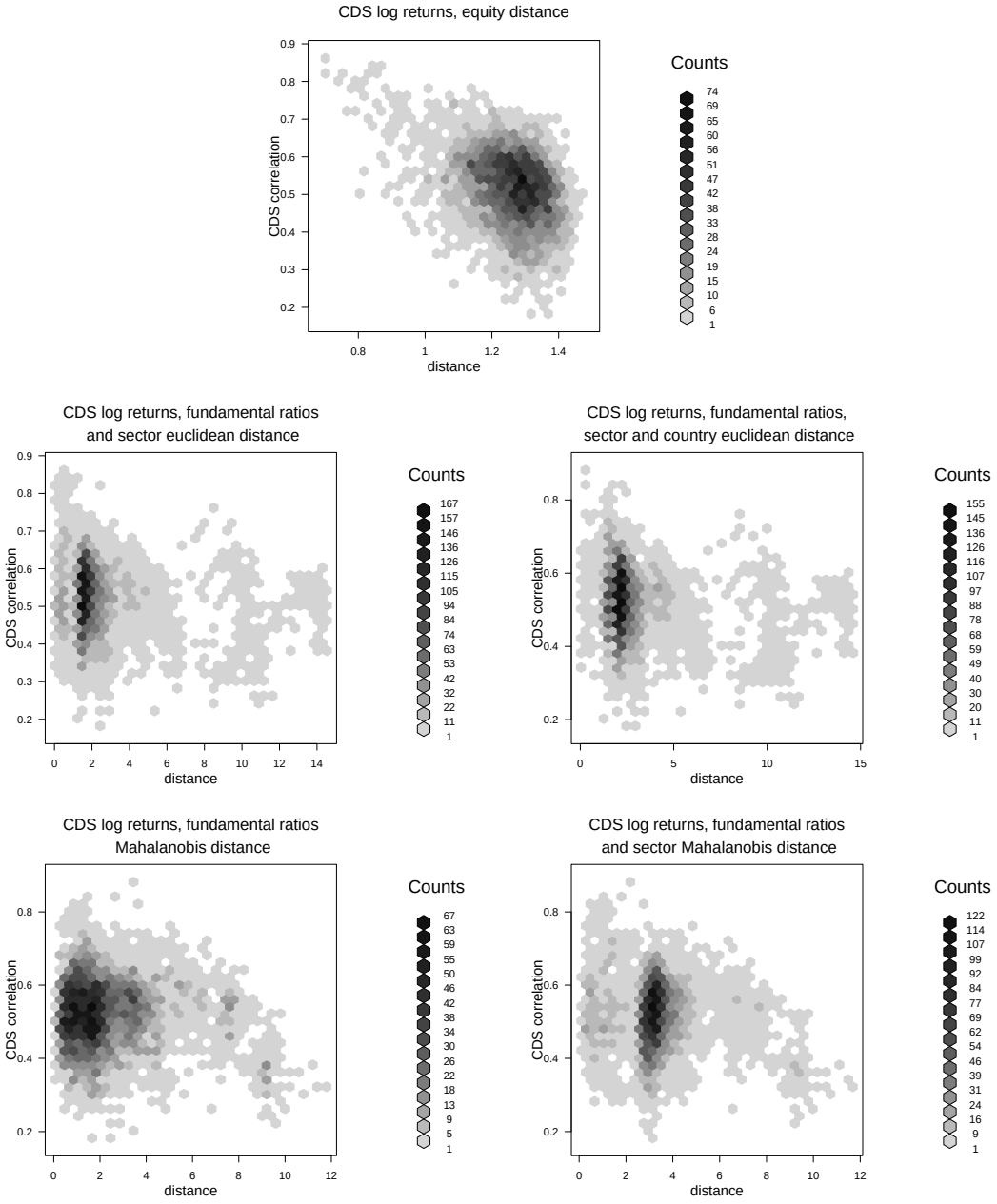

Figure 3.4: Hexbin plots of CDS log return correlations vs. the chosen distance measures d_{E}, $d_{\mathrm{FRS,Eucl.}}$, $d_{\mathrm{FRSC,Eucl.}}$, $d_{\mathrm{FR,M}}$ and $d_{\mathrm{FRS,M}}$. The darker the respective hexagonal field, the more observations it represents.

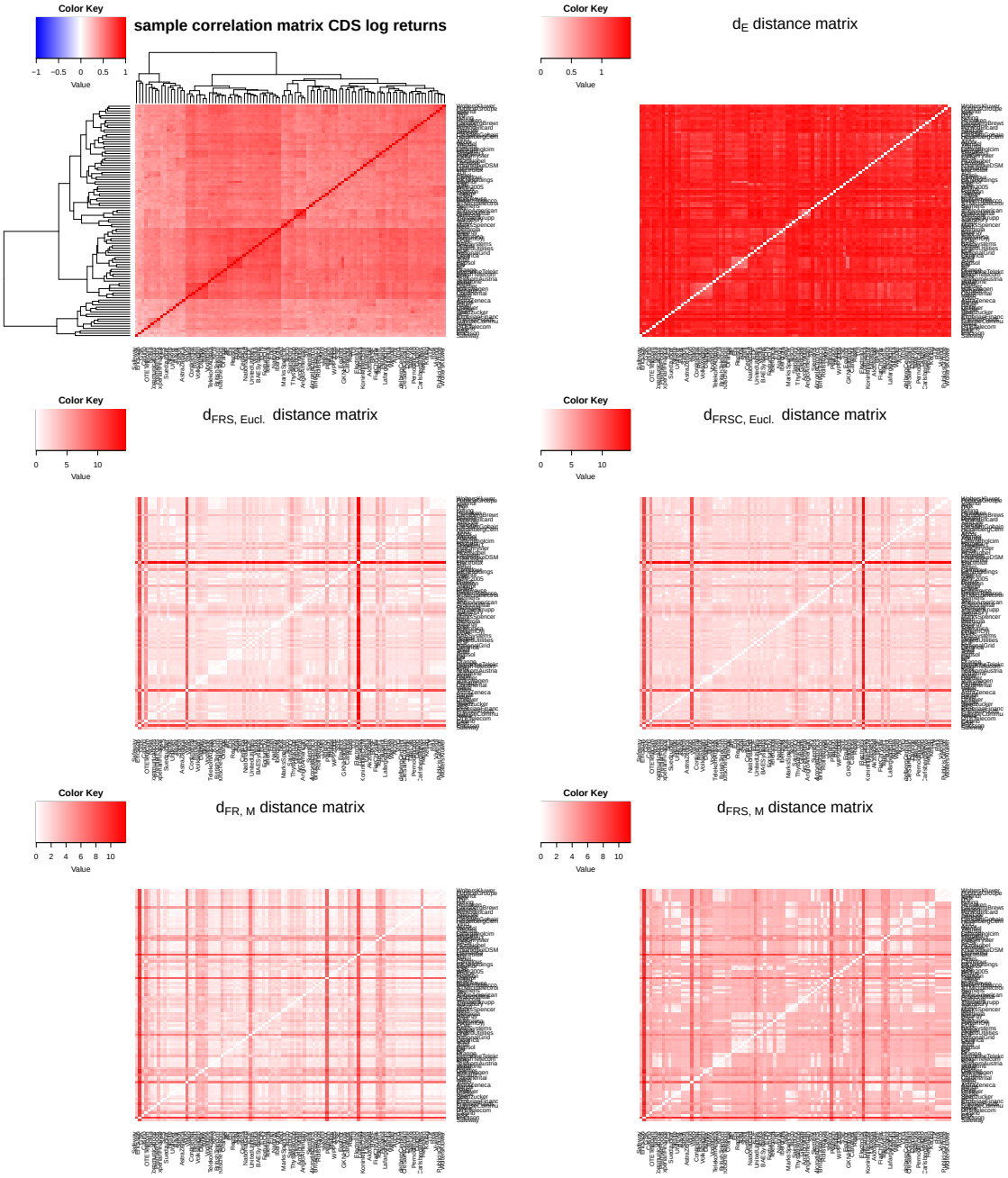

Figure 3.5: Heatmap of CDS correlation matrix (top left), grouped according to R's internal clustering, and distance matrices for d_{E}, $d_{\mathrm{FRS,Eucl.}}$, $d_{\mathrm{FRSC,Eucl.}}$, $d_{\mathrm{FR,M}}$, and $d_{\mathrm{FRS,M}}$, with the same ordering of firms.

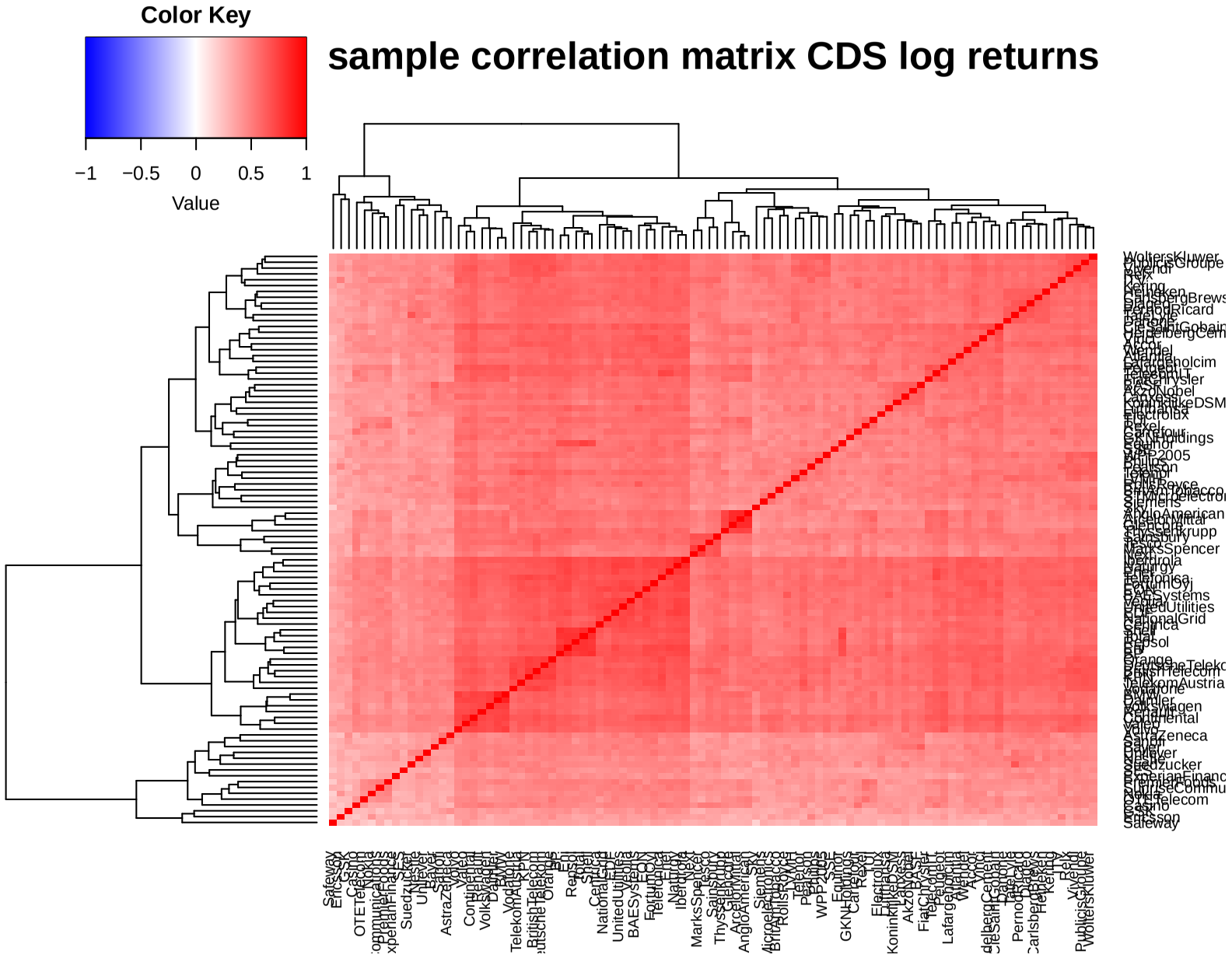

Figure 3.6: Heatmap of CDS log return correlation matrix, grouped according to R's internal clustering, i.e. according to blocks of strongly correlated assets. These blocks tend to correspond to firms operating in the same business sectors.

3.4.2 Data set

We illustrate the benefits of incorporating spatial information via geostatistical modeling for covariance matrix estimation and missing data imputation using a data set of 5-year CDS spreads, stock prices, and macroeconomic variables referring to the reference entities of the CDS listed in the Markit iTraxx Europe (ITRX IG; investment grade) and iTraxx Europe Crossover (ITRX XO; sub-investment grade) Indices Series 29. The ITRX IG (XO) indices contain the 125 (75) most liquidly traded (sub-) investment grade CDS and exist for maturities 3, 5, 7, and 10 years, with 5 years being the most liquidly traded maturity, similar to single-name CDS. The constituents are adjusted every year in March and September, and a new 'on-the-run' series is set up. Some of the reference entities are subsidiaries of publicly traded firms which are not traded themselves; in this case, the stock prices and macroeconomic variables of the public parent company are taken.

After removing firms with missing data, our data set comprises a complete record of CDS, stock, and macroeconomic data for 98 firms[11] in the time period July 25th, 2016 to July 24th, 2018. This already serves as an illustration of the data availability issue for CDS: 57 firms had to be removed from the data set due to missing CDS quotes. Further, private companies were excluded, as no stock price is available; apart from this, only two firms had to be removed from the data set due to missing equity data. Finally, also companies operating in the financial sector were excluded from the data set, as for those different macroeconomic variables are required.

For out-of-sample performance evaluation of the proposed approach we use data of Deutsche Post, Henkel, Merck, Air Liquide, Bouygues, Capgemini, and Sodexo.

Remark 3.4.1 ((Un-)Necessity of complete records)
Estimation of the sample variogram requires only pairwise complete observations of the variable of interest, as positive definiteness of the covariance (resp. correlation) matrix is ensured by fitting a valid model. Thus, missing CDS data is no knock-out criterion for the application of this method. We only remove time series with missing values to obtain a complete record for comparison in the missing data imputation application. For the coordinates, however, a complete record is required.

The geostatistical modeling assumptions require stationary time series; we therefore consider log returns of CDS par spread quotes. As we intend to focus on purely spatial dependence, temporal dependence must be taken care of in some way. We carry out our analyses directly on the log returns, as it is a quite common assumption among practitioners that log returns are approximately free of temporal dependence. However, the presence of temporal dependence in the series cannot be denied: Box tests show that most of the return series are autocorrelated, and heteroscedasticity effects are observed in more than half of the series. For comparison, we fit time series models to each

[11] Data obtained via Thomson Reuters Eikon, with CDS quotes from Markit (=MG).

log return series and repeat the analyses on the residuals. According to Cont and Kan (2011), an AR(1)-GARCH(1,1) model is well suited for capturing observed stylized facts of CDS spread return time series.[12] To comply with the assumption of a Gaussian field, and to be able to directly obtain correlation functions from the fitted variograms, the marginal distributions of each return and residual time series are transformed to the standard normal distribution.
Further, when preprocessing the data, we have to take care of the roll dates of the CDS contracts: As mentioned in Section 2.4.3, CDS maturities are fixed and standardized to 20 June or 20 December, and quotation changes on the roll dates 20 March and 20 September. At these roll dates the series exhibit noticeable jumps due to the changing maturity. We control for this by setting the log return on roll dates to zero.

Concerning the covariate data, the distance measure d_{E} is constructed from correlations of stock log returns observed in the same time period July 25th, 2016 to July 24th, 2018. As balance sheet data is typically reported in yearly intervals, we choose the reported values of 2017 as supplied by Thomson Reuters Eikon for the construction of the remaining distance measures. For firms who do not report at the end of December, we take the figure reported at the time closest to Dec. 31st, 2017.

3.4.3 Covariance or correlation matrix estimation

The estimation of covariance or correlation matrices is a problem often encountered in finance. It is especially challenging when the number of firms d is large, as the estimation of $d(d+1)/2$ parameters (variances and pairwise correlations) is required.

The most natural estimator is the sample covariance matrix of the data set, but this is problematic for several reasons: First, it is singular when the number of observations is small compared to the number of firms considered, $N < d$, cf. Ledoit and Wolf (2004a). Second, in the case of missing data, one has to make a choice between using only dates with a complete record, which might lead to the $N < d$ issue, or compute the sample covariance matrix using pairwise complete observations, which may result in an estimator that is not positive definite. Third, for d large, the sample covariance matrix typically contains a lot of noise compared to the true covariance matrix of the market. Therefore, estimators that impose more structure on the estimated covariance matrix are desirable. At the very least, positive definiteness of the estimator is a necessary requirement. There exists an abundance of literature on this topic, e.g. the series of papers by Ledoit and Wolf (2003, 2004a,b) focusing on shrinkage, Perreault *et al.* (2019) who simplify the covariance matrix by imposing an exchangeable block structure, principal component estimators as discussed, e.g., in Alexander (2002), factor model estimators

[12] We also fitted ARMA(p,q)-GARCH(1,1) with p,q optimally chosen, as AR(1)-GARCH(1,1) was not always the best-fitting model according to AIC/BIC, as well as ARMA(p,q) with p,q optimally chosen, as for about half of the return series homoscedasticity could not be rejected. However, the results of geostatistical model fitting obtained from the residuals of these models were very similar to those obtained from the residuals of the AR(1)-GARCH(1,1), and are thus not presented.

as described, e.g., in Fan *et al.* (2008), or Engle (2002)'s dynamic conditional correlation (DCC) model. This list is by no means exhaustive and just intends to give an overview of some of the proposed approaches. None of these approaches, however, is able to cope with missing data. In this case, a popular approach in the context of CDO modeling is to take the equity correlation matrix, cf. RiskMetrics Group (2007), which is inappropriate in the current market conditions, as we will see in the following.

Natural covariance resp. correlation matrix estimator from geostatistics

From Section 3.2 it can be seen that the geostatistical framework naturally encompasses an estimation procedure for large covariance resp. correlation matrices via the fitting of a valid variogram function. The resulting estimator has various appealing features: First, it is guaranteed to be positive definite, as the covariance resp. correlation function associated with the fitted variogram model is positive definite. Second, it is a parametric function of distance requiring only few parameters (3 at most in the parametric variogram functions considered here). Third, it can cope with missing data, i.e. given the coordinates, one can predict covariances resp. correlations of firms for which no data is available. Last but not least it is easy to compute, as it only requires one fit of the variogram function, which is then used for the estimation of all covariances resp. correlations.
It is worth noting that, when estimating covariances via the fitted variogram, all firms have the same variance, which is somewhat unrealistic. This issue can be resolved by standardizing the observations of the field to mean 0 and variance 1, thus obtaining a correlation function from the fitted variogram via Equation (3.1), and estimating the variances separately for each series.

The estimation procedure follows Steps 0.-3. described above: Having selected a financial distance measure, one collects the relevant coordinate data plus CDS spreads for a training set. In our case this is the ITRX universe, which consists of the most liquidly traded European CDS, and can hence be considered representative for estimating the correlation function governing the European CDS market. Then one calculates the sample variogram using Equation (3.3) and fits a valid variogram model, e.g. using R's `gstat` package, paying attention to the issues described in Section 3.3. Finally, using relation (3.1), one obtains the covariances as functions of the chosen distance.
In our use case, we standardize the field as indicated above, and constrain the nugget and sill parameters of the variogram such that $k(0) = 1$ and we obtain correlations as functions of distance. The fitted variogram models for our chosen distances are given in Table 3.1 and their fit to the sample variograms with respect to our chosen distances is displayed in Figure 3.7 for the CDS log return data set. Cross-validating the variogram fits as described in Remark 3.2.1 confirms that the chosen models fit the data quite well. We find that the fitted models for $d_{\mathrm{FRS,Eucl.}}$ and $d_{\mathrm{FRSC,Eucl.}}$, as well as the models for $d_{\mathrm{FR,M}}$ and $d_{\mathrm{FRS,M}}$ are very similar for both log returns and AR(1)-GARCH(1,1) residuals.

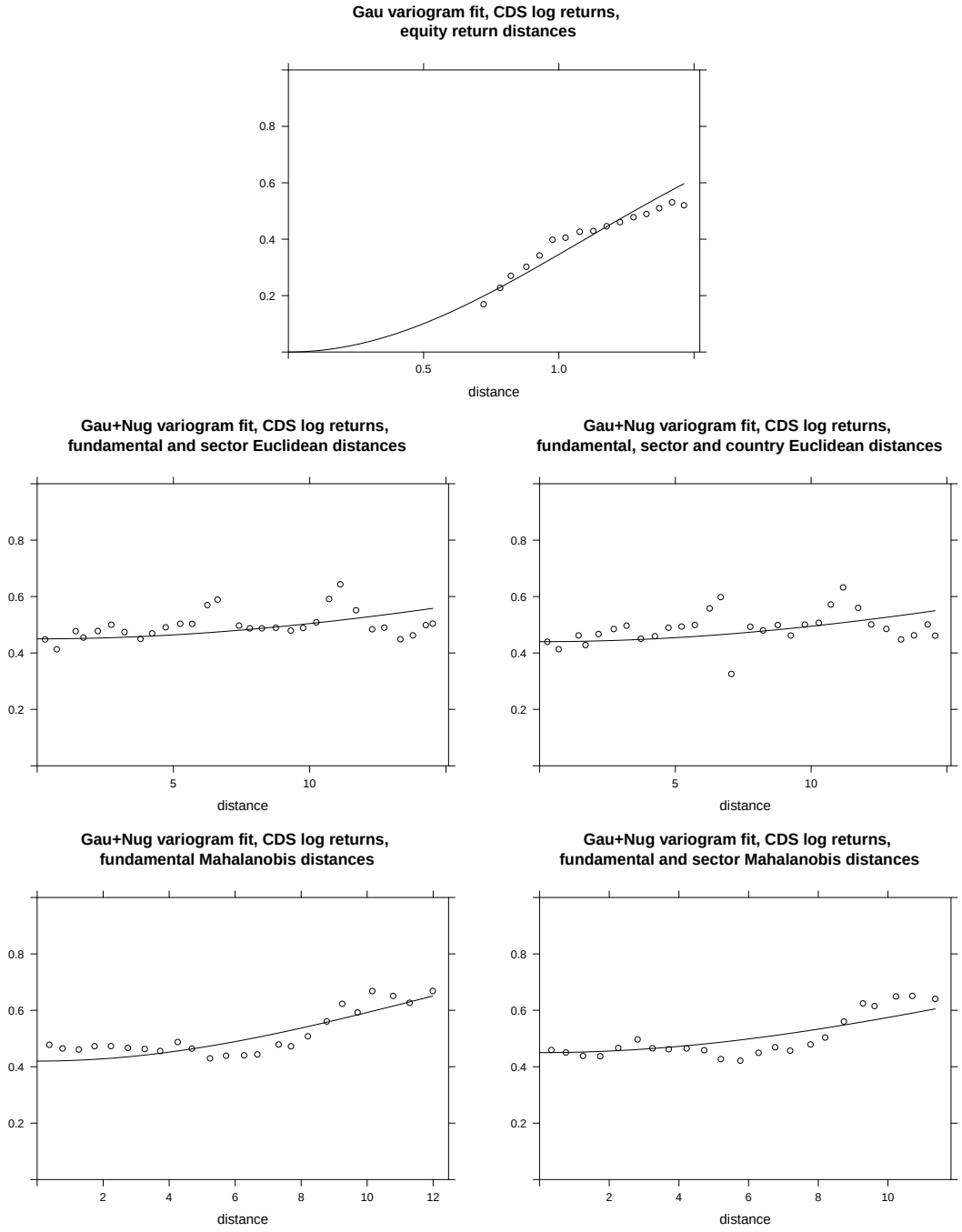

Figure 3.7: Fit of the valid variogram model to the respective sample variogram for the chosen financial distance measures d_{E} (top), $d_{\mathrm{FRS,Eucl.}}$ (middle left), $d_{\mathrm{FRSC,Eucl.}}$ (middle right), $d_{\mathrm{FR,M}}$ (bottom left), and $d_{\mathrm{FRS,M}}$ (bottom right).

distance	model type	nugget	(partial) sill	range
d_{E}, returns	Gaussian	0	1	1.5350
$d_{\mathrm{FRS,Eucl.}}$, returns	Gaussian	0.45	0.55	31.0424
$d_{\mathrm{FRSC,Eucl.}}$, returns	Gaussian	0.44	0.56	31.1891
$d_{\mathrm{FR,M}}$, returns	Gaussian	0.42	0.58	16.8436
$d_{\mathrm{FRS,M}}$, returns	Gaussian	0.45	0.55	19.7333
d_{E}, residuals	Gaussian	0	1	1.3868
$d_{\mathrm{FRS,Eucl.}}$, residuals	Gaussian	0.5	0.5	27.2024
$d_{\mathrm{FRSC,Eucl.}}$, residuals	Gaussian	0.5	0.5	29.3347
$d_{\mathrm{FR,M}}$, residuals	Gaussian	0.5	0.5	17.8945
$d_{\mathrm{FRS,M}}$, residuals	Gaussian	0.48	0.52	16.2328

Table 3.1: Fitted variogram models for CDS log returns and AR(1)-GARCH(1,1) residuals for each of our chosen distance measures.

Performance figures considered for the comparison of models

We compare the sample correlation matrices of our training set of 98 ITRX constituent firms (in-sample) and our training and test set including the 7 new firms listed in Section 3.4.2 (out-of-sample) to the model correlation matrices obtained from the fitted variograms for our chosen distance measures d_{E}, $d_{\mathrm{FRS,Eucl.}}$, $d_{\mathrm{FRSC,Eucl.}}$, $d_{\mathrm{FR,M}}$, and $d_{\mathrm{FRS,M}}$. For (normal-transformed) CDS log-returns, we include the (normal-transformed) equity correlation matrix for comparison. The corresponding results are identified by the subscript 'Equity' in the sequel. Both sample correlation matrices are well-defined here, as our training and test sets are without missing data with $N > d$.
To this end, we consider the following performance figures:

1. Matrix-valued loss functions: The Frobenius loss

$$\mathrm{trace}((C - \hat{C})^2),$$

where C and $\hat{C}$ are the sample and model covariance matrix, respectively, and the negative normal log-likelihood

$$\mathrm{trace}(C\hat{C}^{-1}) - \ln(\det(C\hat{C}^{-1})) - d,$$

as Neuberg and Glasserman (2019) find this is a more suitable loss function from a portfolio perspective.

2. Performance measures addressing the elements of the matrix: RMSE and MAPE for all pairwise correlations and for those of the newly added firms only:

$$\mathrm{RMSE}_{\mathrm{all}} = \sqrt{\frac{1}{d(d-1)/2} \sum_{k=1}^{d} \sum_{l=k+1}^{d} [C - \hat{C}]_{k,l}^2}, \quad C \in \mathbb{R}^{d \times d},$$

$$\mathrm{RMSE}_{\mathrm{new}} = \sqrt{\frac{1}{dd_{\mathrm{new}} + \frac{d_{\mathrm{new}}(d_{\mathrm{new}}-1)}{2}} \sum_{k=d+1}^{d+d_{\mathrm{new}}} \sum_{l=1}^{k-1} [C - \hat{C}]_{k,l}^2}, \; C \in \mathbb{R}^{(d+d_{\mathrm{new}})\times(d+d_{\mathrm{new}})},$$

$$\mathrm{MAPE}_{\mathrm{all}} = \frac{1}{d(d-1)/2} \sum_{k=1}^{d} \sum_{l=k+1}^{d} \frac{\|C_{k,l} - \hat{C}_{k,l}\|}{\|C_{k,l}\|}, \quad C \in \mathbb{R}^{d\times d},$$

$$\mathrm{MAPE}_{\mathrm{new}} = \frac{1}{dd_{\mathrm{new}} + \frac{d_{\mathrm{new}}(d_{\mathrm{new}}-1)}{2}} \sum_{k=d+1}^{d+d_{\mathrm{new}}} \sum_{l=1}^{k-1} \frac{\|C_{k,l} - \hat{C}_{k,l}\|}{\|C_{k,l}\|}, \; C \in \mathbb{R}^{(d+d_{\mathrm{new}})\times(d+d_{\mathrm{new}})}.$$

It is important to remark that MAPE has certain shortfalls as a measure of predictive accuracy, especially when used for model selection, cf. Gneiting (2011); Tofallis (2015). Nevertheless it remains quite popular among practitioners, which is why we include it in our comparison.

In-sample comparison

We find that for CDS log returns, simply taking the corresponding equity correlation matrix is the worst choice, cf. Table 3.2 and Table 3.3. Figure 3.8, displaying the histograms of the entrywise estimation errors for the different model correlation matrices, reveals that this is due to the fact that equity correlations in our data set are systematically much lower than CDS correlations. Figure 3.9 additionally shows which blocks of the CDS log return resp. residual correlation matrix are over- resp. underestimated in the different models. Considering both matrix-valued and element-wise performance figures, the correlation models parameterized by the equity correlation distance d_{E} and by the balance sheet ratios and sector-based Mahalanobis distance $d_{\mathrm{FRS,M}}$ perform best for CDS log returns. For CDS AR(1)-GARCH(1,1) residuals, the correlation model parameterized by $d_{\mathrm{FRSC,Eucl.}}$ performs best according to the Frobenius loss and pairwise correlations RMSE, whereas according to the negative normal log-likelihood and pairwise correlations MAPE, the models parameterized by $d_{\mathrm{FRS,M}}$ and $d_{\mathrm{FR,M}}$ perform best, respectively.

distance	Frobenius ret.	Frobenius resid.	neg. n. Llh. ret.	neg. n. Llh. resid.
Equity	890.1386	-	37.9848	-
d_{E}	**73.8306**	72.7075	27.5979	24.1303
$d_{\mathrm{FRS,Eucl.}}$	84.5127	67.4280	28.3136	25.3291
$d_{\mathrm{FRSC,Eucl.}}$	89.3108	**67.1137**	28.3793	25.3213
$d_{\mathrm{FR,M}}$	98.0467	67.2878	28.6524	25.2926
$d_{\mathrm{FRS,M}}$	79.8826	69.5488	**26.5629**	**23.6696**

Table 3.2: Results of the matrix-valued loss functions (Frobenius loss and negative normal log-likelihood) for CDS log returns and AR(1)-GARCH(1,1) residuals.

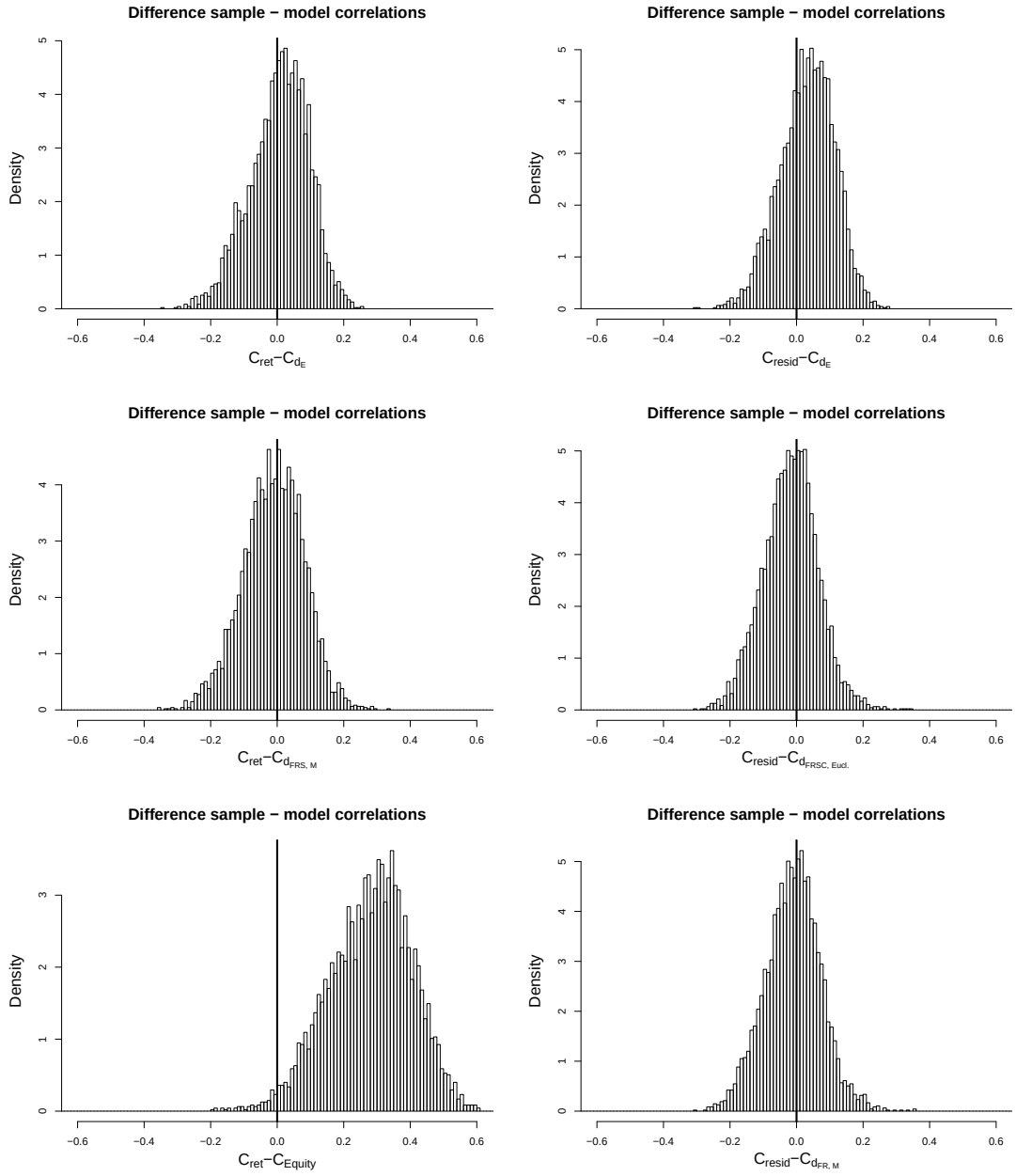

Figure 3.8: Histograms of the distribution of the entries of $\mathrm{Cor}_{\mathrm{sample}} - \mathrm{Cor}_{\mathrm{model}}$.
Left column: $\mathrm{Cor}_{\mathrm{sample}}$ is the CDS log return sample correlation matrix and $\mathrm{Cor}_{\mathrm{model}}$ is the respective model correlation matrix for d_{E} and $d_{\mathrm{FRS,M}}$, and the correlation matrix of (normal-transformed) equity correlation matrix in the bottom plot.
Right column: $\mathrm{Cor}_{\mathrm{sample}}$ is the CDS AR(1)-GARCH(1,1) residual sample correlation matrix and $\mathrm{Cor}_{\mathrm{model}}$ is the respective model correlation matrix for d_{E}, $d_{\mathrm{FRSC,Eucl.}}$, and $d_{\mathrm{FR,M}}$.

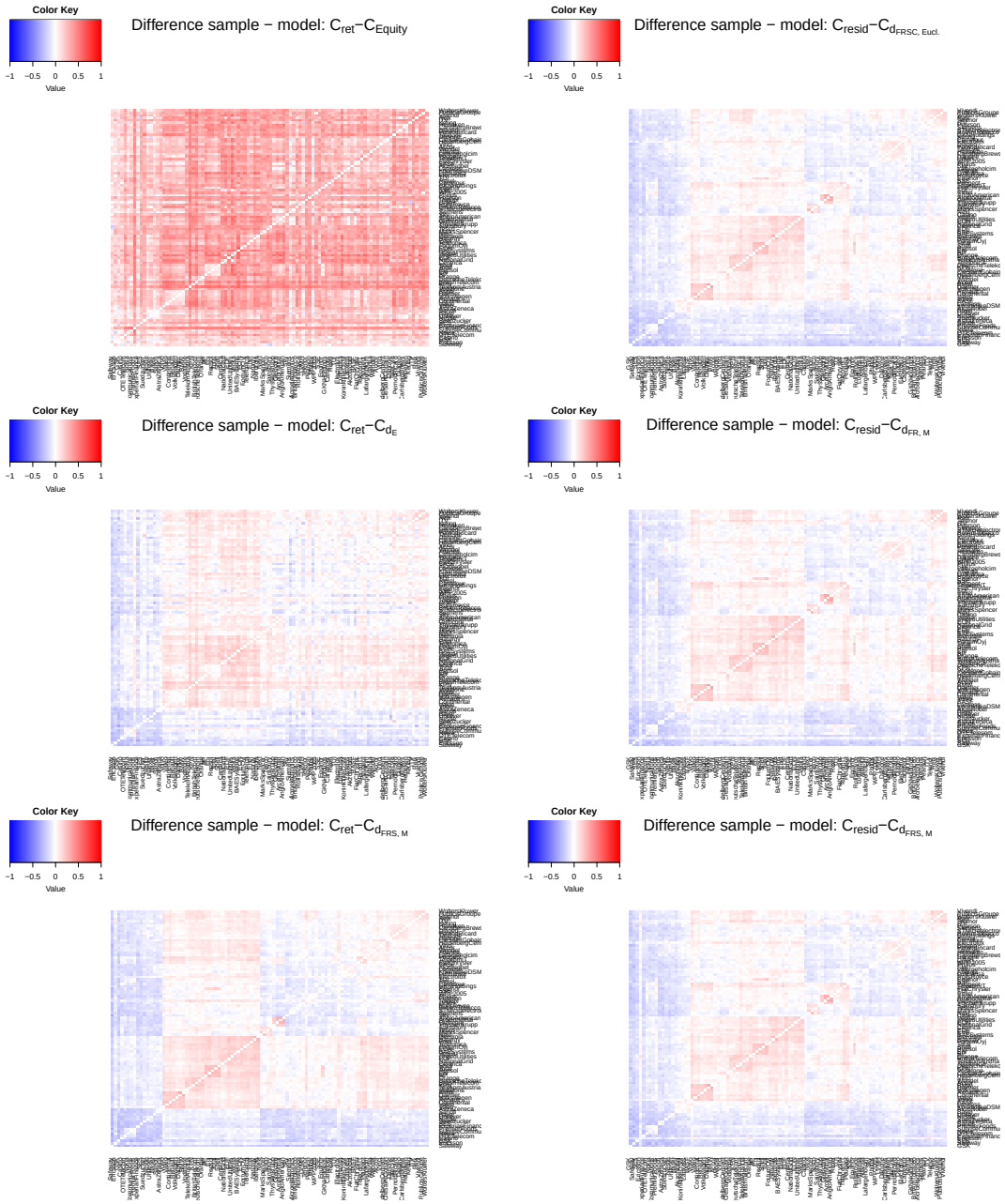

Figure 3.9: Heatmaps of the difference between sample and model correlation matrices for CDS log returns (left column) and residuals (right column). Blue (red) fields indicate that the model overestimates (underestimates) the respective pairwise correlation compared to the sample correlation matrix.

distance	RMSE returns	RMSE residuals	MAPE returns	MAPE residuals
Equity	0.3060	-	0.5400	-
d_{E}	**0.0881**	0.0875	**0.1464**	0.1523
$d_{\mathrm{FRS,Eucl.}}$	0.0943	0.0842	0.1620	0.1543
$d_{\mathrm{FRSC,Eucl.}}$	0.0969	**0.0840**	0.1683	0.1540
$d_{\mathrm{FR,M}}$	0.1016	0.0841	0.1769	**0.1519**
$d_{\mathrm{FRS,M}}$	0.0917	0.0855	0.1551	0.1570

Table 3.3: Results of element-wise performance measures for CDS log returns and AR(1)-GARCH(1,1) residuals.

Out-of-sample comparison

For the CDS log returns of our new firms, surprisingly the equity correlation matrix outperforms all distance-parameterized model correlation matrices concerning RMSE and MAPE of pairwise correlations. Apparently, for these firms, credit and equity correlations are more similar than for the ITRX constituents, and consequently the geostatistical models overestimate the CDS correlations, cf. Figure 3.10. The higher CDS correlations of the ITRX constituents may well be caused by being part of the index. The model based on the equity-correlation distance d_{E} performs best among the geostatistical models in both data sets, CDS log returns and AR(1)-GARCH(1,1) residuals, cf. Table 3.5. According to both matrix-values loss functions applied to the extended correlation matrices, the geostatistical models parameterized by d_{E} and $d_{\mathrm{FRS,M}}$ perform best for both CDS log returns and AR(1)-GARCH(1,1) residuals, cf. Table 3.4.

	Frobenius ret.	Frobenius resid.	neg. n. Llh. ret.	neg. n. Llh. resid.
Equity	926.2568	-	44.3533	-
d_{E}	**131.9362**	**112.7960**	33.9761	29.2042
$d_{\mathrm{FRS,Eucl.}}$	147.9159	119.4380	32.2188	28.9903
$d_{\mathrm{FRSC,Eucl.}}$	157.3199	118.9473	32.3884	28.9790
$d_{\mathrm{FR,M}}$	173.5536	117.4923	32.8869	28.9234
$d_{\mathrm{FRS,M}}$	138.3648	123.9517	**30.4045**	**27.4572**

Table 3.4: Results of the matrix-valued loss functions for CDS log returns and AR(1)-GARCH(1,1) residuals for the extended correlation matrix.

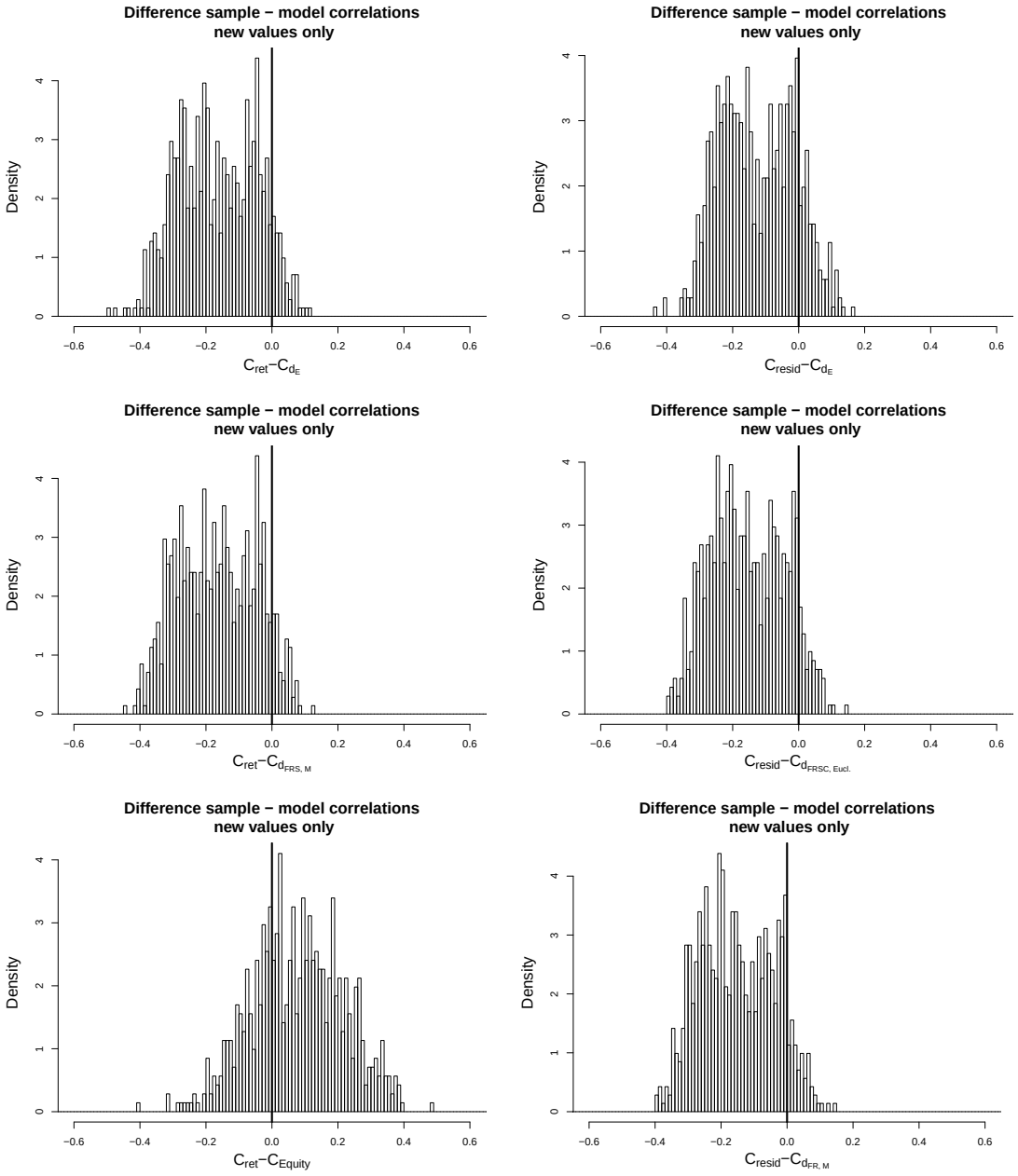

Figure 3.10: Histograms of the distribution of the new entries of $\mathrm{Cor}_{\mathrm{sample}} - \mathrm{Cor}_{\mathrm{model}}$. Left column: $\mathrm{Cor}_{\mathrm{sample}}$ is the CDS log return sample correlation matrix and $\mathrm{Cor}_{\mathrm{model}}$ is the respective model correlation matrix for d_{E} and $d_{\mathrm{FRS,M}}$, and the correlation matrix of (normal-transformed) equity correlation matrix in the bottom plot.
Right column: $\mathrm{Cor}_{\mathrm{sample}}$ is the CDS AR(1)-GARCH(1,1) residual sample correlation matrix and $\mathrm{Cor}_{\mathrm{model}}$ is the respective model correlation matrix for d_{E}, $d_{\mathrm{FRSC,Eucl.}}$, and $d_{\mathrm{FR,M}}$.

	RMSE returns	RMSE residuals	MAPE returns	MAPE residuals
Equity	**0.1598**	-	**0.3483**	-
d_{E}	0.2027	**0.1684**	0.6289	**0.5670**
$d_{\mathrm{FRS,Eucl.}}$	0.2118	0.1918	0.6716	0.6692
$d_{\mathrm{FRSC,Eucl.}}$	0.2193	0.1915	0.6974	0.6682
$d_{\mathrm{FR,M}}$	0.2311	0.1884	0.7352	0.6558
$d_{\mathrm{FRS,M}}$	0.2034	0.1961	0.6424	0.6841

Table 3.5: Results of element-wise performance measures for CDS log returns and AR(1)-GARCH(1,1) residuals for new firms only.

3.4.4 Interpolation of missing data

An issue often encountered when working with CDS data is poor data quality. Unlike in equity markets, quotes are not freely available, and often one only obtains data sets with either missing data or the same quote repeated over several days, which is clearly unrealistic. In general, this might also be caused by missing liquidity, but our data set comprises the most liquidly traded series by definition of the ITRX index family. Again, the geostatistical framework offers a nice solution to this problem: Executing Steps 0.-4. as described in Section 3.2, knowing the distance between firms and the covariance resp. correlation function of the data set, i.e. the considered market, one can easily interpolate missing values using the kriging technique.

Comparison with existing methods

We compare the performance of the kriging imputation with the performance of copula-based imputation methods proposed by Di Lascio *et al.* (2015) and Käärik and Käärik (2009, 2010).
Käärik and Käärik (2009, 2010) assume that the dependence structure of the data set is a Gaussian copula, and they impute the expected value given the observed data. Under the assumption of a Gaussian field, this is very similar to simple kriging as described in Section 3.2, the only difference being that in geostatistics, the copula is parameterized by distance among the components.
Di Lascio *et al.* (2015) extend this ansatz to a broader selection of copula classes, and instead of imputing the expected value of the conditional distribution given the observed values, they impute a value drawn randomly from this conditional distribution. The corresponding code is available in the R package `CoImp`.
In order to test our model, we deleted observations from our data sets using `CoImp`'s `MCAR` (missing completely at random) function and impute these missing values by simple kriging from fitted variograms using the different distances established in Section 3.3.1, as well as by using the copula-based imputation techniques introduced above. To evaluate

the different methods, we use the following performance figures:

$$\text{MAPE} = \frac{1}{M}\sum_{k=1}^{M}\left|\frac{z_k^{\text{imp}} - z_k^{\text{obs}}}{z_k^{\text{obs}}}\right|, \qquad \text{RMSE} = \sqrt{\frac{1}{M}\sum_{k=1}^{M}(z_k^{\text{imp}} - z_k^{\text{obs}})^2},$$

$$\text{MAE} = \frac{1}{M}\sum_{k=1}^{M}|z_k^{\text{imp}} - z_k^{\text{obs}}|,$$

where M is the number of missing observations and the superscripts 'imp' and 'obs' refer to imputed and observed values, respectively. As already stated, MAPE has several shortfalls, e.g. being undefined for $z_k^{\text{obs}} = 0$. However, it is a popular performance measure among practitioners and the performance measure of choice in Di Lascio *et al.* (2015). The missing values with $z_k^{\text{obs}} = 0$ (being less than 0.2% of all missing values) are excluded in the reported MAPE values in Table 3.6.

In our use case, `CoImp` identifies the Gaussian copula with unspecified dependence structure as the best-fitting dependence structure among the supplied models and the Gaussian copula with constant correlation structure as the second-best for both data sets, log returns and AR(1)-GARCH(1,1) residuals. Supplied copula families were Gaussian with constant correlation matrix, Gaussian, Clayton, Gumbel, Frank, and the t-copula with degree of freedom fixed to 4. (Imputation was performed with `CoImp` V. 0.3-1, which does not yet explicitly consider Gaussian copulas other than the constant correlation variant or t copulas, but the package's authors confirmed that the procedure works fine for these.)
We find that `CoImp` performs worst in both data sets according to all performance measures. This, however, is most likely due to the fact that the randomly drawn imputed value is not the optimal point forecast for any of the chosen performance measures, cf. Remark 3.4.2 below. The imputed values from the geostatistical approach slightly outperform the imputed values from the Gaussian copula approach of Käärik and Käärik (2009, 2010) according to all performance figures, with the forecasts based on $d_{\text{FRS,M}}$ and d_{E} performing best for the imputation of log returns, and d_{E} and $d_{\text{FRS,Eucl.}}$ performing best for the imputation of AR(1)-GARCH(1,1) residuals, cf. Table 3.6.

Remark 3.4.2 (On point forecasts)
As illustrated in Gneiting (2011), just comparing some point forecasts by some measures of predictive accuracy is essentially comparing apples to oranges. He argues for either disclosing the accuracy measure, so that forecasters can give the respective optimal point forecast from their predictive distribution (e.g. the expected value for RMSE), or explicitly asking for a certain functional of the predictive distribution as their point forecast (e.g. the expected value). In this spirit, comparing the performance measures of our forecast and the one made by `CoImp` has exactly this problem.

	MAPE (ret)	RMSE (ret)	MAE (ret)	MAPE (resid)	RMSE (resid)	MAE (resid)
Gauss. cop. (cc)	1.603722	0.728327	0.552596	2.502895	0.781716	0.600040
Gauss. cop. (un)	1.769959	0.730541	0.556732	2.537761	0.787327	0.603779
`CoImp` (cc)	2.304826	1.109832	0.887970	3.172209	1.150790	0.918862
`CoImp` (un)	2.389993	1.115477	0.894545	3.478642	1.143631	0.915300
d_{E}	1.646298	0.723373	**0.545422**	2.617390	**0.774820**	**0.592405**
$d_{\mathrm{FRS,Eucl.}}$	1.603087	0.727450	0.551777	**2.493205**	0.780758	0.599383
$d_{\mathrm{FRSC,Eucl.}}$	1.604094	0.727512	0.551711	2.496190	0.780921	0.599430
$d_{\mathrm{FR,M}}$	1.626555	0.727595	0.552468	2.505752	0.780601	0.599547
$d_{\mathrm{FRS,M}}$	**1.600154**	**0.723067**	0.548492	2.520846	0.775817	0.595213

Table 3.6: Performance figures for the imputed CDS log returns and AR(1)-GARCH(1,1) residuals from the Gaussian copula approach of Kääarik and Kääarik (2009, 2010), from `CoImp`, and from simple kriging using the different distances defined in Section 3.3.1. The labels (cc) and (un) refer to the constant correlation and unspecified dependence structure, respectively.

Discussion

The advantage of the distance-parameterized over the conventionally estimated Gaussian copulas is only marginal for our data set. However, this might be improved substantially for other financial distance measures, or in more diverse data sets.
Further, the assumption of a Gaussian field for financial data could be relaxed. The natural next step would be to apply a spatial copula approach as described in Gräler (2014), i.e. use a copula parameterized by distance to model the dependence structure, which can be seen as an approximation to a more complicated field by means of copulas, and also as an extension of the copula-based imputation methods presented in Di Lascio *et al.* (2015); Kääarik and Kääarik (2009, 2010). A related ansatz extending vine copulas[13] to incorporate spatial information is described in Erhardt *et al.* (2015).

3.5 Conclusion and outlook

We discussed the application of geostatistical modeling to financial data sets as well as the necessary adjustments in higher-dimensional coordinate systems. In the light of the presented results, we find that the geostatistical approach is a promising alternative for the joint modeling of dependent credit spreads concerning the estimation of the correlation matrix, especially considering the appealing properties of this approach in the presence of missing data, and the imputation of missing data.
Future research may generalize our results in several ways: In the context of other financial data sets, different distances could be explored. Considering the question of (geometric) anisotropy in higher dimensions, the promising metric learning approaches

[13] A vine copula is a d-dimensional copula constructed solely from bivariate copulas. For an introduction to vine copulas, see, e.g. Aas *et al.* (2009).

could be more thoroughly examined. Transcending the sometimes unrealistic assumption of a Gaussian field, one may work with spatially parameterized copulas to model the dependence structure. Finally, space-time models may be explored to be able to consider temporal and spatial dependence simultaneously.

Part II

New pricing approaches for credit products

4 Pricing single-name CDS options in a structural model with jumps

4.1 Motivation

CDS options (CDSO) grant their owner the right to enter into a credit default swap (CDS) at prespecified spread. They can provide insurance against rising or falling CDS spreads or allow for speculation on the evolution of CDS spreads. Further, CDSO offer the possibility to extend one's protection on an underlying instrument whose maturity might be extended contrary to investors' expectations, i.e. the instrument might not be called as expected. Although quotes for single-name CDSO, i.e. CDS options with a single firm as reference entity, are sparse nowadays, the valuation of these optionalities merits study, as they are embedded in certain structured products like cancellable single-name protection according to Martin (2012), and considering the increasing standardization of the underlying CDS contracts, a rise in trading volume of single-name CDSO is well possible.

CDSO exist in the form of *Payer* and *Receiver* CDSO, which grant the right to enter the CDS contract as protection buyer (hence *paying* coupons in exchange for default protection) and protection seller, respectively. Several contractual specifications are possible with regards to strike quotation and payment upon exercise, see Martin (2012), with a strike quotation in running spread and payment in terms of an upfront and fixed coupon being the current market standard. However, since the standard ISDA model furnishes a one-to-one conversion formula[1] for running spread to upfront+fixed coupon, we consider CDSO with both strike quotation and payment upon exercise in terms of running spread. Throughout this chapter, we focus exclusively on the pricing of Payer CDSO; Receiver CDSO prices can be derived in a similar fashion.

4.1.1 Standard approaches to CDSO valuation

The topic of CDSO valuation has already been explored in several directions: Approaches favored by practitioners are those that lead to Black-type formulas, analogously to equity options, see for example Schönbucher (2004) and Brigo (2005), who present

[1] The International Swaps and Derivatives Association (ISDA) supplies a standard model for CDS valuation, cf. `http://www.cdsmodel.com/cdsmodel/`. See also Mai (2014) for a short description.

different approaches that yield this formula. For an exhaustive overview of the different possible contractual terms of CDSO and Index CDSO[2] and the corresponding valuation methods with (or related to) Black-type formulas, see Martin (2012).
The main drawback of these approaches is the assumption that CDS spreads evolve according to a lognormal diffusion process. Market observations show that this is strongly questionable, as documented, e.g., by Cont and Kan (2011), who provide an extensive overview of stylized properties of market-observed CDS spreads. To further illustrate this, Figure 4.1 contrasts the logarithmic spread differences of several firms to a normal distribution. One observes that a normal distribution for CDS spread changes seems to underestimate the likelihood of large moves and to overestimate the probability of moderate spread changes. Nevertheless, the Black-type formula may be used for a quotation of implied volatilities of CDSO, analogously to equity options.

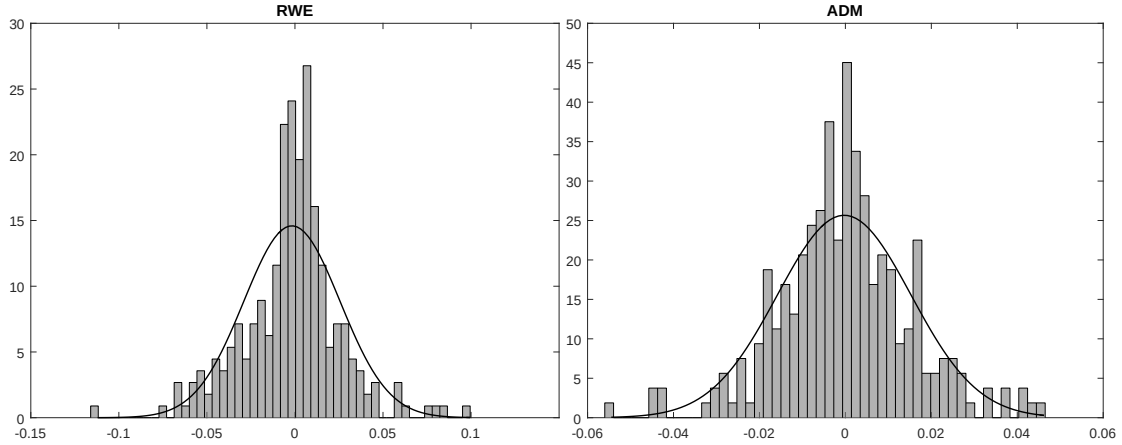

Figure 4.1: Empirical densities of observed CDS spread changes for RWE (left) and ADM (right) vs. the corresponding fitted normal densities with same mean and standard deviation. Clearly, the fitted normal distributions underestimate the probabilities of both large and very small moves, while overestimating the probability of moderate spread changes. (Source: Bloomberg.)

Apart from Black-type formulas, using intensity-based models[3] for the valuation of CDSO has dominated the academic literature so far, cf. e.g., Brigo and El-Bachir (2010), who present a quasi-analytic formula for CDSO in a shifted square root jump diffusion intensity model, and Bielecki *et al.* (2011a), who consider valuation and hedging of CDSO in a CIR default intensity model.

[2] An Index CDSO grants the right to enter into an Index CDS. These options are regularly traded in the market.

[3] Intensity-based models are a type of credit models where the default time is modeled as an exogenous random variable via a stochastic hazard rate.

4.1.2 Alternative approach via structural credit models

Choosing to model the spread evolution directly bereaves us of the opportunity to integrate CDSO in the bigger picture of a firm's capital structure. An alternative approach for the pricing of CDSO, which has mainly been overlooked in the literature, employs so-called *structural (credit-equity) models*. The class of structural models allows for unified credit and equity pricing, which as byproduct generates CDS spread dynamics and thus allows us to price CDSO alongside with not only CDS but also equity derivatives.

In the structural framework, it is assumed that the value of a firm's assets follows some stochastic process. All financial instruments related to this firm are then viewed as derivatives on the firm value. The time of default is typically defined as the first-passage time below a prespecified barrier. Early approaches include the seminal references Merton (1974) and Black and Cox (1976), where the firm value is assumed to follow a geometric Brownian motion. This assumption leads to predictable default times, an undesirable feature that results in a CDS curve that vanishes at the short end. To generate realistic default times, credit curves, and dynamics for the securities, jumps are added to the firm value process, see for example Zhou (2001) and Chen and Kou (2009). For an overview of structural credit-equity models, see also Hüttner (2014).

The main advantage of structural models is their firm economic interpretation, which, for models with realistic dynamics, typically comes at the cost of more complicated pricing formulas and calibration routines. The fact that they can handle any type of firm-related security makes them a powerful tool, e.g., for the pricing of hybrid securities or for detecting capital structure arbitrage opportunities.

To date, Jönsson and Schoutens (2008) is the only reference incorporating structural models for the valuation of CDSO. They consider firm value models driven by pure jump Lévy processes, which result in unrealistic CDS spread dynamics following a sawtooth trajectory, see Figure 4.2.

In the light of the drawbacks of the presented approaches, in this chapter we propose to use the structural model introduced in Chen and Kou (2009) for the valuation of CDSO. There, the firm value process follows a jump-diffusion process with double-exponential jumps. The major benefits of this model are that it supplies analytic formulas for the CDS price and spread, which facilitates CDSO pricing, and that it provides quite realistic dynamics for the CDS spread evolution, see Figure 4.2 for a comparison with real-world spread series and an exemplary spread series generated by one of Jönsson and Schoutens (2008)'s pure jump models.

For the valuation of CDSO, no analytical formulas are attainable, and one has to resort to Monte Carlo pricing, as the CDS price is a complicated function of the firm value. However, due to the fact that usually CDSO are European-style, we can exploit an efficient Monte Carlo algorithm based on Brownian bridges, see Metwally and Atiya (2002) and Ruf and Scherer (2011), for our purposes.

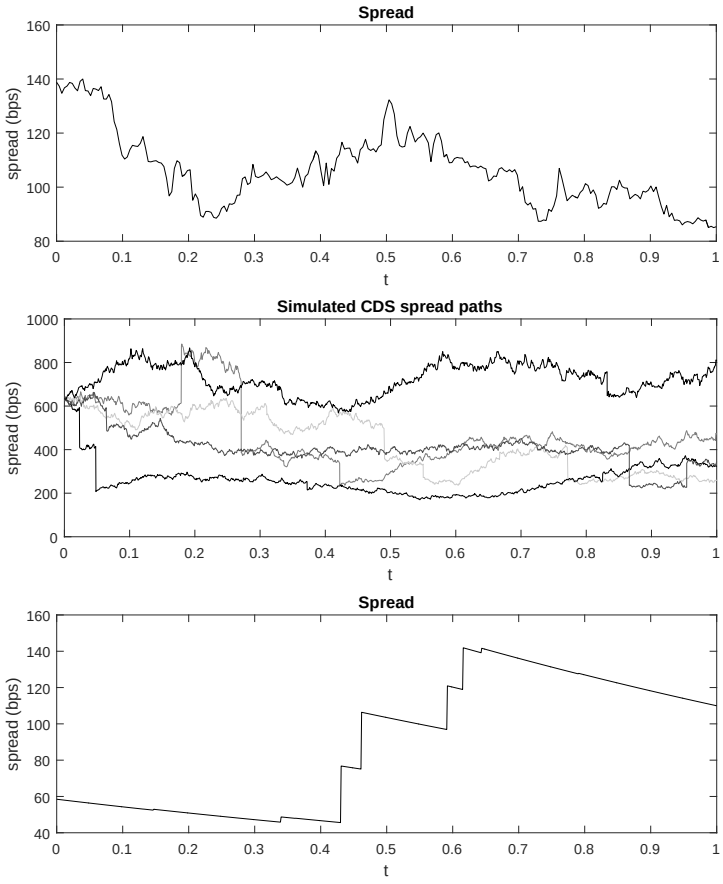

Figure 4.2: CDS spread time series of RWE (Source: Bloomberg) (top) vs exemplary spread paths generated in the Chen–Kou model (middle) and an exemplary spread path of the shifted Gamma pure jump model of Jönsson and Schoutens (2008) (bottom).

Most of the presented results are joint work with Matthias Scherer and have been published in the article Hüttner and Scherer (2016). The remainder of this chapter aims at a more detailed discussion and is structured as follows: Section 4.2 reviews the model-free valuation of CDS and states model-free valuation formulas for CDSO. Section 4.3 introduces the structural model by Chen and Kou (2009) and explains our Monte Carlo algorithm, and Section 4.4 conducts a sensitivity analysis of the CDSO price with respect to the model parameters and calculates model prices for CDSO with the presented algorithm in a real-world example.
The presented algorithm can be readily adapted to price other European optionalities in cases where the Chen–Kou model supplies an analytical pricing formula for the underlying asset. In Section 4.5 this is done for the valuation of callable bonds with one call date, which are closely linked to the concept of *extension risk* as studied, e.g., in De Spiegeleer and Schoutens (2014).

4.2 Model-free valuation of CDS options

In the following, we consider a (European-style Payer) CDSO with maturity $\hat{T}$, evaluated at a time $t \in [0, \hat{T}]$, as illustrated in Figure 4.3. At this time, one may enter into a CDS with maturity T at a prespecified strike running spread s^K.

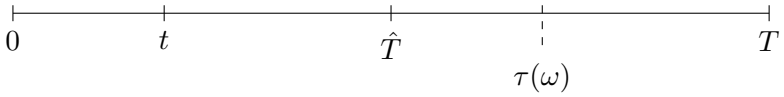

Figure 4.3: Timeline: $\hat{T}$ denotes the inception of the CDS contract, T its maturity. The time point t illustrates the valuation date of the CDSO, and the default may happen at a random time $\tau > t$.

First, we review the valuation of a CDS contract as sketched in the introduction in greater detail. A CDS is essentially an insurance contract between two parties against a credit event of a reference entity. Whereas in the market different credit events such as failure to pay or debt restructuring may trigger the compensation payment of a CDS, we restrict our modeling to a default being the only possible credit event. The random time where the reference entity defaults on its debt is denoted by τ. The protection buyer has to pay a standardized premium c, typically 100 or 500 bps, up to maturity of the contract or default, whichever comes earlier. In case of a default during the lifetime of the contract, they are compensated for losses by the protection seller. Premium payments are usually made quarterly in arrears, and the remaining difference between the two payment streams, the so-called upfront, has to be paid at inception of the contract, cf. Figure 4.4.

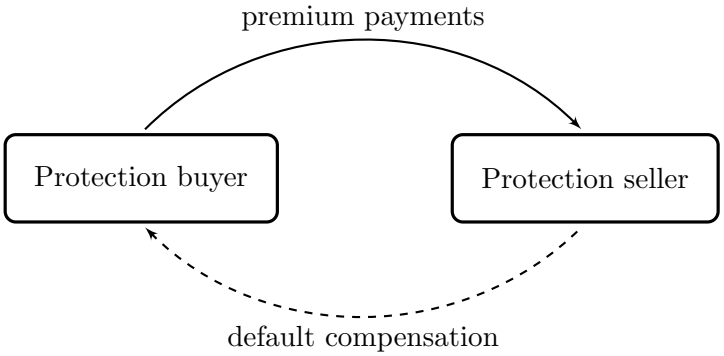

Figure 4.4: Schematic display of CDS payment streams. Depending on the sign of the difference of the two payment streams, the upfront has to be paid by the protection buyer or seller.

To ease calculations, we consider a simplified version of these payment streams based on continuous spread payments. The expected discounted values of the payments made by the protection buyer, the premium leg ($EDPL$), and the protection seller, the default leg ($EDDL$), are then given as follows:

$$\begin{aligned} EDPL_{\hat{T},T}(t) &= \mathbb{E}\left[\int_{\hat{T}}^{T} e^{-r(s-t)}\mathbb{1}_{\{\tau>s\}}\mathrm{d}s \,\middle|\, \mathcal{F}_t\right], \\ EDDL_{\hat{T},T}(t) &= (1-R)\mathbb{E}\left[e^{-r(\tau-t)}\mathbb{1}_{\{\hat{T}<\tau\leq T\}} \,\middle|\, \mathcal{F}_t\right], \end{aligned} \tag{4.1}$$

where $R \in [0,1]$ denotes the recovery rate of the reference entity[4] and is assumed to be deterministic, and $r \neq 0$ denotes the risk-free interest rate. Note that $t \in [0,\hat{T}]$, so the forward values of the legs are included in the above definition. In case of $\tau \leq \hat{T}$ the CDSO expires worthless. The value of a CDS at time $\hat{T}$ from the viewpoint of the protection buyer is then

$$CDS_{\hat{T},T}(t=\hat{T},c) = EDDL_{\hat{T},T}(\hat{T}) - c \cdot EDPL_{\hat{T},T}(\hat{T}). \tag{4.2}$$

CDS are often quoted in terms of the so-called running spread $s_{\hat{T},T}(t=\hat{T})$, which is the premium level c for which the contract has zero value at inception:

$$s_{\hat{T},T}(\hat{T}) = \frac{EDDL_{\hat{T},T}(\hat{T})}{EDPL_{\hat{T},T}(\hat{T})} = \frac{r(1-R)\mathbb{E}[e^{-r(\tau-\hat{T})}\mathbb{1}_{\{\hat{T}<\tau\leq T\}}|\mathcal{F}_{\hat{T}}]}{1-e^{-r(T-\hat{T})}\mathbb{Q}(\tau>T|\mathcal{F}_{\hat{T}}) - \mathbb{E}[e^{-r(\tau-\hat{T})}\mathbb{1}_{\{\hat{T}<\tau\leq T\}}|\mathcal{F}_{\hat{T}}]}, \tag{4.3}$$

which is obtained from (4.1) by integration by parts.

CDSO grant the right to enter into a CDS contract at some future time point $\hat{T}$, either as protection buyer (payer option) or as protection seller (receiver option). Single-name products are sparsely traded nowadays, but were advancing in volume before the crisis. Martin (2012) predicts a rising volume in these contracts in post-crisis times, considering the fact that the credit derivatives market is more regulated nowadays. CDSO are mostly traded European-style. As stated above, we consider options with the strike specified in terms of a running spread s^K where one exercises into a CDS with running spread s^K and no initial payment, which makes calculations easier to interpret than in the real-world setup where the underlying CDS is traded with fixed coupon and corresponding upfront payment[5]. In the risk-neutral valuation approach, the formula for a payer CDSO's value at time $t \leq \hat{T}$ with expiry $\hat{T}$ and CDS maturity T is

$$\begin{aligned} CDSO_{t,\hat{T},T}(s^K) &= \mathbb{E}\left[e^{-r(\hat{T}-t)}\mathbb{1}_{\{\tau>\hat{T}\}}CDS_{\hat{T},T}(\hat{T},s^K)^+ \,\middle|\, \mathcal{F}_t\right] \\ &= \mathbb{E}\left[e^{-r(\hat{T}-t)}\mathbb{1}_{\{\tau>\hat{T}\}}EDPL_{\hat{T},T}(\hat{T})(s_{\hat{T},T}(\hat{T})-s^K)^+ \,\middle|\, \mathcal{F}_t\right], \end{aligned} \tag{4.4}$$

[4] The recovery rate is the remaining value at default of a bond with nominal 1. For the sake of simplicity, we assume that every bond of a reference entity has the same recovery rate, and any bond of the reference entity may be delivered into the CDS upon default.

[5] A version of the CDSO payoff which takes underlying CDS with upfront/fixed coupon into account is $(u_{\hat{T},T}(s_{\hat{T},T}) - u^K(s^K))^+$, see also Martin (2012). Using the standard ISDA model, the two payoffs can be converted into each other.

where expectations are taken under a risk-neutral measure that is calibrated, e.g., to CDS or bond markets.

4.3 Valuation of CDS options in the Chen–Kou model

In the structural approach to credit risk, the evolution of a firm's value is modeled by a stochastic process $V = \{V_s\}_{s\geq 0}$. All financial instruments related to the firm are then interpreted as derivatives on the firm value. We use the model introduced by Chen and Kou (2009) for the valuation of CDS options, as it offers a good tradeoff between realistic price dynamics for the CDS spread and analytical tractability.

The firm value process is modeled as follows:

$$V_s = V_0 e^{X_s}, \qquad X_s = \mu s + \sigma W_s + \sum_{k=1}^{N_s} Y_k, \tag{4.5}$$
$$\mu = r - \delta - \frac{\sigma^2}{2} - \lambda\left(\frac{p\xi_1}{\xi_1 - 1} + \frac{(1-p)\xi_2}{\xi_2 + 1} - 1\right),$$

where r is the risk-free interest rate (assumed to be flat in order to achieve a quasi-closed form expression for the CDS price), δ refers to the firm's payout rate to its investors, $\sigma > 0$ is the volatility of the diffusion component, $W = \{W_s\}_{s\geq 0}$ is a standard Brownian motion, and $N = \{N_s\}_{s\geq 0}$ is a Poisson process with intensity $\lambda \geq 0$. The jump sizes $\{Y_k\}_{k\in\mathbb{N}}$, are double-exponentially distributed, i.e. with density

$$f_Y(x) = p\xi_1 e^{-\xi_1 x}\mathbf{1}_{\{x\geq 0\}} + (1-p)\xi_2 e^{\xi_2 x}\mathbf{1}_{\{x<0\}}, \quad p \in [0,1],\ \xi_1 > 1,\ \xi_2 > 0,$$

where p and $1/\xi_i$, $i \in \{1,2\}$, control the probability of an upward jump and the average jump sizes.
The default time τ is the first passage time below the barrier $V_B = \kappa V_0, 0 < \kappa < 1$, after the valuation date t:

$$\tau := \inf\{s \geq t : V_s \leq V_B\}.$$

A semi-explicit formula for the spot CDS price can be obtained: From (4.3) we see that $\mathbb{Q}(\tau > s), s \geq \hat{T}$, and $\mathbb{E}[e^{-r\tau}\mathbb{1}_{\{\hat{T}<\tau\leq T\}}]$ are the only expressions required for CDSO valuation. The Laplace transforms (denoted $\mathcal{LT}$ hereafter) of these quantities are known in closed form, see Kou and Wang (2003):

$$\mathcal{LT}[\mathbb{Q}(\tau > s)](u) = \frac{1}{u}\left(1 - \mathbb{E}\left[e^{-u\tau}\right]\right) = \frac{1}{u}\left(1 - w_1^u \kappa^{\beta_1^u} - w_2^u \kappa^{\beta_2^u}\right), \tag{4.6}$$
$$\mathcal{LT}[\mathbb{E}[e^{-r\tau}\mathbb{1}_{\{\tau\leq T\}}]](u) = \frac{1}{u}\left(w_1^{r+u}\kappa^{\beta_1^{r+u}} + w_2^{r+u}\kappa^{\beta_2^{r+u}}\right), \tag{4.7}$$
$$w_1^u = \frac{(\xi_2 - \beta_1^u)\beta_2^u}{\xi_2(\beta_2^u - \beta_1^u)}, \qquad w_2^u = 1 - w_1^u.$$

Without loss of generality we stated the valuation of the Laplace transforms for $\hat{T} = 0$, as this is just a simple time shift. The quantities β_1^u, β_2^u are the two negative roots of the equation $G(\beta) = u$, where $G(\cdot)$ is defined via

$$\mathbb{E}[e^{\beta X_s}] = e^{G(\beta)s}, \qquad G(\beta) = \mu\beta + \frac{\sigma^2\beta^2}{2} + \lambda\left(\frac{p\xi_1}{\xi_1 - \beta} + \frac{(1-p)\xi_2}{\xi_2 + \beta} - 1\right).$$

Fast computation of these roots is possible, as this problem corresponds to computing the roots of a quartic equation, cf. Kou and Wang (2003); Kou *et al.* (2005).

Assuming the model parameters σ, λ, p, ξ_1, and ξ_2 to be fixed at inception of the CDSO, the only unknown variable in the calculation of $CDS_{\hat{T},T}(\hat{T})$ in (4.4) is $\kappa(\hat{T}) = V_B/V_{\hat{T}}$, i.e. $V_{\hat{T}}$. Thus, for the pricing of CDSO required is the joint distribution of $V_{\hat{T}}$ and the running minimum of the firm value process, $\min_{t \le s \le \hat{T}} V_s$.

Kou and Wang (2003) state a semi-explicit but numerically challenging formula for the Laplace transform of the joint survival probability required for CDSO pricing, but testing with various parameter sets showed that it is highly unstable in certain cases. Hence, we opt to implement an efficient Monte Carlo simulation.

To apply a Monte Carlo pricing approach, one needs to model paths of V over the time interval $[t, \hat{T}]$ and to check if the default barrier V_B is crossed. In this case the option expires worthless. If the firm survives up to time $\hat{T}$, the value of the CDS can be calculated as a function of $V_{\hat{T}}$.
As our chosen firm value process follows a lognormal diffusion between the jump times, we can work with a log-transformed version of our problem:

$$X_s = \ln\left(\frac{V_s}{V_t}\right), \qquad d = \ln\left(\frac{V_B}{V_t}\right),$$

the default time τ remains the same. Now we can exploit the fact that the probability of a Brownian bridge between two fixed endpoints crossing some fixed barrier is explicitly known, see Metwally and Atiya (2002):

$$\mathbb{Q}\left(\min_{s\in[t_{i-1},t_i]} X_s > d \,\middle|\, X_{t_{i-1}}, X_{t_i-}\right) = \mathbf{1}_{\left\{\min\left\{X_{t_{i-1}}, X_{t_i-}\right\} > d\right\}}\left(1 - e^{\frac{-2(X_{t_i-}-d)(X_{t_{i-1}}-d)}{\sigma^2(t_i - t_{i-1})}}\right).$$

We reformulate the CDSO valuation formula (4.4) by conditioning on the number, location, and size of jumps in $[t, \hat{T}]$, which determine the relevant quantity for the CDS valuation $V_{\hat{T}}$:

$$\begin{aligned} CDSO_{t,\hat{T},T}(s^K) &= e^{-r(\hat{T}-t)} \\ &\quad \cdot \mathbb{E}\left[\mathbb{Q}(\tau > \hat{T} | N_{\hat{T}}, \{Y_k\}_{k=1,\dots,N_{\hat{T}}}, \text{jump times}) CDS_{\hat{T},T}(\hat{T}, s^K; V_{\hat{T}})^+\right], \end{aligned}$$

which leaves us with the task of checking for default at the jump times and on the Brownian bridges between two consecutive jumps. This enables us to adapt the Monte Carlo algorithm presented in Metwally and Atiya (2002); Ruf and Scherer (2011).

Algorithm 4.3.1 (Valuation of CDSO)
For each of the independent Monte Carlo runs $j \in \{1, \dots, n\}$:

1. Simulate the number of jumps in $[t, \hat{T}]$ of the firm value process, i.e.

$$N := N_{\hat{T}} - N_t \sim \mathrm{Poi}\big(\lambda(\hat{T} - t)\big),$$

the jump times $t_1, \dots, t_N \in [t, \hat{T}]$,

$$(t_1, \dots, t_N) \sim Sort\big(\mathcal{U}([t, \hat{T}], 1, N)\big),^{6}$$

and the jump sizes $X_{t_i} - X_{t_i^-} \sim \mathrm{DEXP}(p, \xi_1, \xi_2)$.

2. Extend the jump times to $\{t_0 = t, t_1, \dots, t_N, t_{N+1} = \hat{T}\}$ and simulate the Brownian increments between two consecutive time points in this set,

$$X_{t_{i+1}^-} - X_{t_i} \sim \mathcal{N}\big(\mu(t_{i+1} - t_i), \sigma^2(t_{i+1} - t_i)\big),$$

and calculate $X_{t_{i+1}^-}, X_{t_{i+1}}$.

3. Check if the barrier has been crossed at any of the $\{X_{t_i}, X_{t_i^-}\}_{i=0,\dots N+1}$. If so, set the payoff for this run to zero, as the CDSO expires worthless in case of default previous to $\hat{T}$.

4. If the barrier has not been crossed, calculate the conditional survival probability in $[t, \hat{T}]$, i.e.

$$CSP_j = \prod_{i=0}^{N} \mathbb{Q}\left(\min_{t_i \le s \le t_{i+1}} X_s > \ln\left(V_B/V_0\right) \middle| X_{t_i}, X_{t_{i+1}^-} \right).$$

Further calculate $V_{\hat{T}} = V_t \exp\{X_{\hat{T}}\}$ and $CDS_{\hat{T},T}(\hat{T}, s^K; V_{\hat{T}})$, and set the payoff for this run to $CSP_j \cdot CDS_{\hat{T},T}(\hat{T}, s^K; V_{\hat{T}})$.

Finally, $CDSO_{t,\hat{T},T}(s^K) \cong \frac{1}{n} \sum_{j=1}^{n} \text{payoff}_j$.

The algorithm can easily be applied to other European-style optionalities, for example to a callable bond with a single call date. This, however, is an unlikely feature of callable bonds; most of these instruments have several call dates or whole call periods, rendering the embedded option Bermudan- or American-style. These more complicated optionalities require themselves a numerical pricing approach which would have to be included within the Monte Carlo algorithm. Jönsson and Schoutens (2008) argue that, once an efficient method of generating paths of the underlying asset is available (i.e. CDS spreads in the context of CDSO), similar techniques as in the equity derivatives context can be applied, and propose to use least squares Monte Carlo techniques for the evaluation of American or Bermudan optionalities. Even though from Steps 1.-3., an

[6] See Sato (2007, Proposition 3.4) for a proof.

efficient way of generating paths of the firm value process V can be deduced, translating this to a spread path is computationally costly due to the involved Laplace inversions (depending on the number of Monte Carlo runs and nodes per path). Jönsson and Schoutens (2008) propose to speed up the pricing routine by precalculating the price of the underlying for a fine grid of possible firm values, simulating firm value paths, and interpolating on the precalculated grid to obtain paths for the underlying value.

The assessment of extension risk, however, which is closely related to the case of a callable bond with one call date, can be tackled with a simple modification of the presented algorithm, see Section 4.5.

4.4 Sensitivity analysis and an intuition for market prices

To better understand the behavior of CDSO prices in our model, we conduct a sensitivity analysis of the CDSO price with respect to the model's input parameters. The initial parameter choice is displayed in Table4.1.

V_0	κ	r	δ	σ	λ	p	ξ_1	ξ_2	R	$\hat{T}$	T	s^K
100	0.7	0.01	0	0.3	0.7	0.5	4	4	0.4	1	6	0.0233

Table 4.1: Base case model parameters for our example.

Sensitivity with respect to the diffusion parameters

As a first step, we consider the pure diffusion-based structural model, i.e. $\lambda = 0$. The sensitivity of the CDSO price with respect to changes in interest rate, diffusion volatility, option maturity, CDS maturity, and default barrier is displayed in Figure 4.5. Looking at the graph of CDSO price vs. CDS lifetime, one notes that this is merely a forward shift of the credit curve. The CDSO price is falling in the interest rate, but hump-shaped in the diffusion volatility. Most notably, the sensitivity of the CDSO price with respect to the option maturity depends on the values of σ and κ: For small values of the diffusion volatility and default barrier, it increases in $\hat{T}$, as in both cases default before $\hat{T}$ is unlikely. For higher values of σ or κ, it becomes increasingly likely that the option expires worthless, therefore the CDSO price decreases in $\hat{T}$ in these cases.

Sensitivity with respect to the jump parameters

Including jumps, CDSO prices are found to react similarly to changes in diffusion volatility and interest rate as in the diffusion-based model. The sensitivity with respect to the

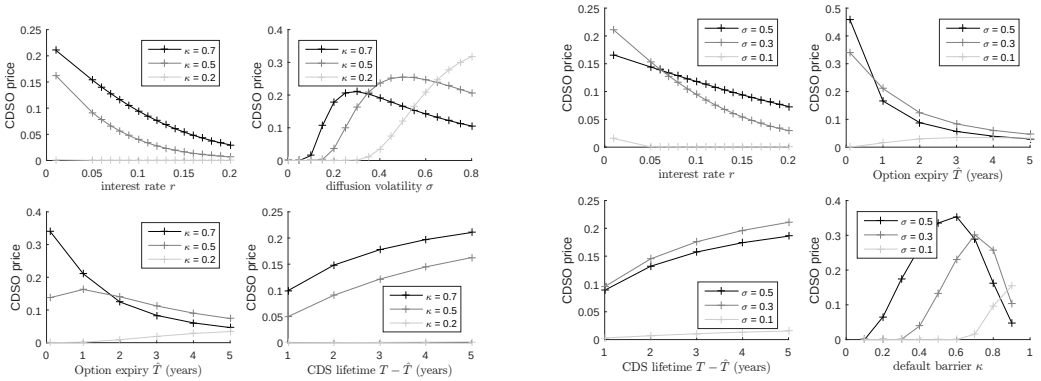

Figure 4.5: Sensitivity analysis of the CDSO price in a pure diffusion model with respect to the interest rate r, the diffusion volatility σ, the option and CDS maturity, and the default barrier parameter κ. Base parameters are given in Table 4.1.

jump parameters is visualized in Figure 4.6. One observes that the CDSO price reacts non-linear to changes in the jump parameters: it displays concavity in the jump parameters λ and p, and even changes curvature in ξ_1 and ξ_2.

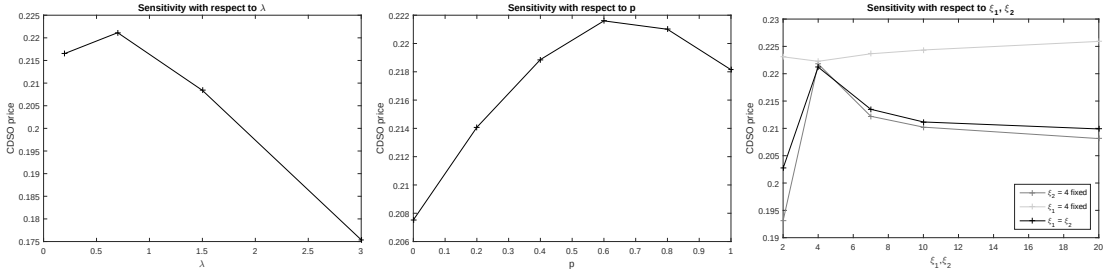

Figure 4.6: Sensitivity analysis of the CDSO price in the Chen–Kou model with respect to the jump parameters. The behavior with respect to r and σ is similar as in the pure diffusion case.

CDSO prices in a market example

In the following we briefly present an example with real-world data (source: Bloomberg), to give an intuition for CDSO prices in the market.
The Chen–Kou model is calibrated to the values of equity and debt (for which explicit formulas are known, see Chen and Kou (2009)) and the CDS curve, yielding the parameter set given in Table 4.2.

Figure 4.7 shows the fit of the model-implied credit curve to real-world CDS data. The model explains observed CDS spreads well for maturities 0.5 to 7 years, but has

$E(V_0)$	Shares	$D(V_0)$	B	c	R	r	δ
$3.2962 \cdot 10^9$	$3.0296 \cdot 10^8$	$5.1764 \cdot 10^9$	$5.3530 \cdot 10^9$	0.0732	0.4	0.01	0.0462

V_0	κ	σ	λ	p	ξ_1	ξ_2
$1.4376 \cdot 10^{10}$	0.7124	0.095	0.1	0.25	4	4

Table 4.2: Market observed data and parameter set obtained from a calibration to the CDS curve, debt value, and equity value. $E(V_0)$ and $D(V_0)$ denote the current value of equity and debt as observed on valuation date, respectively, B is the total nominal outstanding in debt, and c denotes the average coupon rate of the firm's bonds, weighted according to time to maturity and outstanding nominal. The interest rate r is chosen to reflect the current market situation, and for the payout ratio δ the proxy $cB/(E(V_0) + D(V_0))$ is taken. The remaining parameters except V_0 are calibrated to the CDS curve. Finally, V_0 is set so that the model equity value matches the observed market cap E_0.

difficulties in explaining the 10-year CDS spread. The debt value associated with the calibrated parameters is about 1.01 % (52 mn \$) higher than the debt value from bond prices. This is an acceptable error, as Bloomberg quotes for the total outstanding debt are 28 mn \$ higher than the nominal of the debt issues, indicating the company has some debt that is not traded in the market.

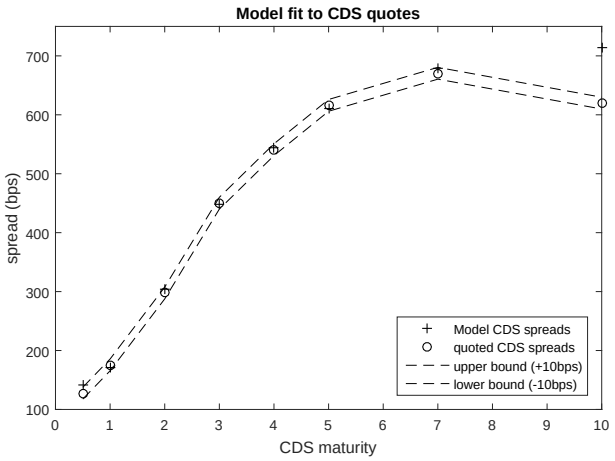

Figure 4.7: Model fit to observed credit curve.

Figure 4.8 shows the prices for a CDSO with expiry in $\hat{T} = 1$ year and a lifetime of the underlying CDS of 5 years for different strikes centered around the quote of the 5y spot CDS.

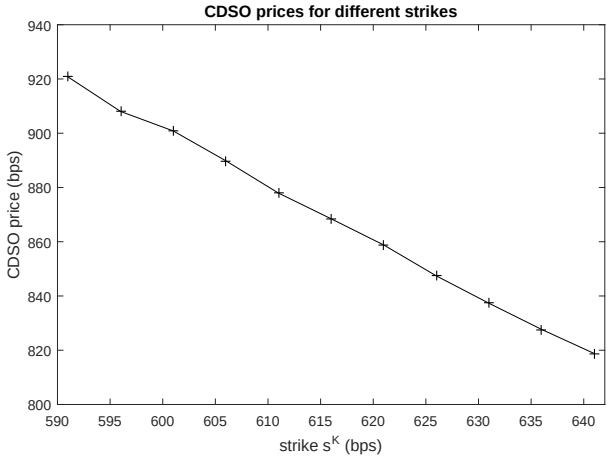

Figure 4.8: Model-generated CDSO prices for $\hat{T} = 1$ and $T = 6$.

4.5 Other optionalities and extension risk

As stated above, Algorithm 4.3.1 can easily be adapted to other European optionalities, as long as the Chen–Kou model supplies an analytical formula for the underlying asset.

Consider, for example, a callable bond[7] with maturity T and a single call date $\hat{T}$: It is well known that the value of a callable bond can be decomposed into the value of a similar bond without call feature minus the price of a call option on the bond. The price of a coupon-bond with coupon rate c is given as follows:

$$\begin{aligned} B_{t,T}(c) &= \mathbb{E}\left[e^{-r(T-t)}\mathbb{1}_{\{\tau>T\}} + e^{-r(\tau-t)}R\mathbb{1}_{\{t<\tau\leq T\}} \right. \\ &\qquad \left. + \sum_{t<t_i\leq T}(t_i - \max\{t_{i-1},t\})ce^{-r(t_i-t)}\mathbb{1}_{\{\tau>t_i\}}\,\middle|\,\mathcal{F}_t\right] \\ &= e^{-r(T-t)}\mathbb{Q}(\tau>T|\mathcal{F}_t) + R\mathbb{E}[e^{-r(\tau-t)}\mathbb{1}_{\{t<\tau\leq T\}}|\mathcal{F}_t] \\ &\quad + \sum_{t<t_i\leq T}(t_i - \max\{t_{i-1},t\})ce^{-r(t_i-t)}\mathbb{Q}(\tau>t_i|\mathcal{F}_t), \end{aligned}$$

where the t_i's denote the coupon payment dates. This formula relies again on the availability of an explicit expression for the survival probability, thus in the Chen–Kou model a quasi-analytic formula depending on V_t and the process parameters is available. With this quasi-analytic formula for the underlying, the embedded European option in a callable bond with a single call date can easily be evaluated with the above algorithm. As already stated, a single call date is a rather unrealistic feature for callable bonds. However, related to the case of callable bonds with a single call date is the

[7] Note that Brigo (2005) already established a link between CDSO and callable defaultable floaters in an intensity framework.

problem of *extension risk*, which was analyzed in the context of contingent convertibles (CoCos) by De Spiegeleer and Schoutens (2014). Basically, it refers to potential losses of bondholders that occur if a call feature is not exercised as expected, so the problem is not limited to contingent convertibles but is present in any callable bond. A prominent example is the 3.875 % January 2014 bond issued by Deutsche Bank in January 2004 (DE0003933511), which was commonly expected to be called at its first call date in 2009. Due to deteriorating market conditions, Deutsche Bank decided against calling it – a decision that inflicted a considerable loss on investors assuming a call at par to take place. See De Spiegeleer and Schoutens (2014) for more details on this example.

Newly issued bonds with call features may display a switch from fixed to floating rate coupons after the first call date. This is often true for aforementioned CoCos. Previously, there has often been a coupon step-up at the beginning of a call period, thus punishing issuers who did not redeem their bonds at the first possibility. Regulators forbade this practice of call incentives, which gave rise to this fixed-to-float switching practice: The floating rate coupon is some reference rate plus a fixed spread determined at contract inception, chosen such that at inception, the floating plus fixed spread coupon equals the fixed coupon. This way, the argument of call incentives is bypassed without actually removing it, as due to changes in the reference rate an incentive to call could build up until the first call date. Thus, the market still might consider the first call date as the "actual maturity date" of the bond, likewise in the example presented in the introduction. This illustrates the importance of assessing the probability of an extension event. Further, it suffices to only consider the bond's first call date, thus placing us directly in the case of a bond with one single call date.

We consider the problem of a firm that has issued a callable bond with maturity T and (first) call date $\hat{T}$. At $\hat{T}$, the firm can either redeem the bond and replace it with a new bond at market conditions, or keep the original bond running. Assuming the redemption price is par and, to completely avoid funding issues, the new bond is issued at par, the remaining free variable in the bond price formula is the coupon. Therefore we resort to comparing the step-up coupon rate of the original bond with the coupon rate of the new bond determined via the root search $B_{\hat{T},T}(c) \overset{!}{=} 1$.

To evaluate extension risk, we calculate the joint probability of survival and extension PSE, as the coupon of the new bond is only defined in case of survival up to $\hat{T}$.

Algorithm 4.5.1 (Joint probability of survival and extension)
For each run $j \in \{1, \dots, n\}$:

1. Simulate the number of jumps in $[t, \hat{T}]$ of the firm value process, i.e.

$$N := N_{\hat{T}} - N_t \sim \mathrm{Poi}\big(\lambda(\hat{T} - t), \big),$$

the jump times $t_1, \dots, t_N \in [t, \hat{T}]$,

$$(t_1, \dots, t_N) \sim Sort\big(\mathcal{U}([t, \hat{T}], 1, N)\big),$$

and the jump sizes $X_{t_i} - X_{t_i^-} \sim \mathrm{DEXP}(p, \xi_1, \xi_2)$.

2. Extend the jump times to $\{t_0 = t, t_1, \ldots, t_N, t_{N+1} = \hat{T}\}$ and simulate the Brownian increments between two consecutive time points in this set,

$$X_{t_{i+1}^-} - X_{t_i} \sim \mathcal{N}\big(\mu(t_{i+1} - t_i), \sigma^2(t_{i+1} - t_i)\big),$$

and calculate $X_{t_{i+1}^-}, X_{t_{i+1}}$.

3. Check if the barrier has been crossed at any of the $\{X_{t_i}, X_{t_i^-}\}_{i=0,\ldots,N+1}$. If so, set $PSE_j = 0$ and continue.

4. If the barrier has not been crossed, calculate the conditional survival probability up to $\hat{T}$, i.e.

$$CSP_j = \prod_{i=0}^{N} \mathbb{Q}\left(\min_{t_i \le s \le t_{i+1}} X_s > \ln(V_B/V_0) \middle| X_{t_i}, X_{t_{i+1}^-} \right).$$

Further calculate $V_{\hat{T}} = V_t \exp\{X_{\hat{T}}\}$ and c^{new} from $B_{\hat{T},T}(c^{new}) = 1$.

5. Check if $c^{new} > c^{call}$. If so, set $PSE_j = CSP_j$, otherwise set $PSE_j = 0$.

Finally, $PSE \cong \frac{1}{n} \sum_{j=1}^{n} PSE_j$.

This version is simplified to consider fixed coupons. For the case of floating rate coupons of the form $r^{ref} + x$, with r^{ref} a known reference interest rate and x fixed premium, just replace $c = r^{ref} + x$ in the above formulas and solve for x.

4.6 Conclusion and outlook

We illustrated the benefits and drawbacks of using a structural default model with jumps for the valuation of CDSO and related optionalities. Regarding CDSO, the main benefit of the class of structural models is that they open a door to the equity side of a firms balance sheet, providing the possibility to consider also equity instruments for hedging purposes, where so far only standard CDS have been considered (see Bielecki *et al.* (2011a)).

Considering further the choice of model, previous literature (e.g. Brigo (2005)) already found that structural models built only on diffusion processes might not be appropriate for the valuation of optionalities in defaultable assets, as default times are predictable. We further argue against the use of pure jump processes as used in Jönsson and Schoutens (2008), as they are not able to supply realistic dynamics for the CDS spread evolution. The jump-diffusion structural model introduced in Chen and Kou (2009) is identified to be an interesting choice of model in this context, as is supplies realistic CDS credit curves as well as realistic dynamics for the evolution of credit spreads. Additionally, it is to a large extent analytically tractable and offers quasi-closed formulas for survival probabilities and CDS prices. Unfortunately, employing the Chen–Kou

model introduces the drawback of a flat interest rate curve, as analytical pricing formulas rely on a fast routine for Laplace inversion which cannot deal with deterministic or stochastic rates.

An efficient Monte Carlo algorithm for the valuation of CDS options, adapted from Metwally and Atiya (2002); Ruf and Scherer (2011), was presented. It can be easily modified for European options on other assets for which the Chen–Kou model supplies an explicit pricing formula. This includes callable bonds with only one call date, which can also be exploited for the assessment of extension risk, where it makes sense to consider only the next call date of the corresponding bond.

For other option types, the presented approach is computationally costly: American or Bermudan options, as often embedded in callable bonds, require themselves numerical pricing routines that would slow down the Monte Carlo algorithm considerably. A closed-form pricing approach, based on the joint distribution of the process's value at option expiry and its running minimum, would be desirable in this context. Although Kou and Wang (2003) state an analytical formula of the Laplace transform of this distribution, the derivation of quasi-analytical formulas (using Laplace inversion) fails due to numerical instabilities related to the large parameter set in the Chen–Kou model and the complicated form of this Laplace transform.

5 Analytical lower bound for the price of a convertible bond

5.1 Motivation

This chapter focuses on analytical approximations for the price of a convertible bond, a hybrid debt instrument. In addition to the features of a regular bond, i.e. a claim against the issuing company that guarantees contractually specified coupon and redemption cash flows, provided the issuing company is solvent, a convertible bond provides its holder with the right to convert the bond into shares of the company in a prespecified period. The *conversion ratio*, the number of shares to be received per unit of bond nominal, is also prespecified in the bond prospectus. Typically, this ratio is increased in case of dividend payments, to prevent a dilution of convertible bond holders. Further, a convertible bond often includes a so-called *soft call* covenant, which protects the issuing company from having to give away its shares in a scenario when it is doing extremely well. Technically, the bond prospectus specifies a trigger level K, acting like a cap on parity. If the *parity* of the convertible bond, the current share price multiplied with the conversion ratio, breaches the trigger level, the issuer has the right to call the bond at par. A typical trigger level in the market equals $K = 1.3$ units of bond nominal, thus a rational holder of the bond will opt for conversion when parity breaches the trigger level in order to avoid a loss of $(K - 1)$ times the bond nominal. Hence, technically a soft call right is a right to enforce conversion rather than a right to call the bond at par. In rare cases, also regular put or call rights of the issuer may be present.

Both the involved American options and the necessity to simultaneously price credit- and equity-components make the pricing of convertible bonds a challenging task. Early approaches are based on a decomposition of the convertible bond into a regular bond and an equity call option. The next step in developing an accurate methodology for convertible bond pricing is Tsiveriotis and Fernandes (1998): Their model constitutes probably the first pricing approach that jointly takes into account credit- and equity-components. However, it still fails to respect important fundamental relationships between credit and equity, such as the advent of a severe drawdown of the stock price in case the company files for bankruptcy. The seemingly most popular approach among practitioners nowadays is based on so-called *defaultable Markov diffusion models*, also known as *1.5-factor models*. The idea is to model two stochastic driving factors, namely the stock price and the default intensity, but express the latter as a function of the former, so that there is essentially only one factor left. Mathematically, the driving stock price diffusion is sent to

a cemetery state (representing bankruptcy) once the integrated default intensity process breaches an exponential trigger level, which creates a feedback effect from the default intensity to the stock price diffusion. Evaluation algorithms for convertible bonds in these models, adapted from the PDE techniques known from the equity derivatives literature to the setup of a convertible bond, are presented in Ayache *et al.* (2003); Andersen and Buffum (2004). The series of papers Bielecki *et al.* (2008a,b, 2009, 2011b) provides a very detailed and technical overview on these probabilistic models and provides rigorous justification for the use of aforesaid PDE techniques. Prominent (semi-)parametric examples of such models include Carr and Linetsky (2006); Linetsky (2006); Carr and Madan (2010).

Due to the involved American optionalities, no analytical pricing formulas are available, and convertible bond pricing routines necessarily involve time-consuming numerical techniques such as finite differences for the numerical solution of a pricing PDE or tree approximations of the underlying stock price diffusion. We draw upon an idea well-known in the context of equity derivatives, namely the decomposition of the price of an American option into the price of a European option plus an *early exercise premium* (EEP), cf. Detemple and Tian (2002); Broadie and Detemple (2004). Often, the EEP is relatively small compared to the option price, so that the European option price serves as a good approximation of the price. For instance, the EEP is exactly zero in cases where early option exercise is well-known to be suboptimal, e.g. for an equity call option in case of no dividends and non-negative interest rates, or for an equity put option in case of non-positive, risk-free interest rates. Analogously, the price of a convertible bond can be decomposed into the sum of a convertible bond with European conversion option at maturity and an *early conversion premium* (ECP), which is often quite small for similar reasons as in the equity derivatives case. Consequently, our derivation of a lower bound for the convertible bond price is based on the idea of 'Europeanizing' the American conversion option of the holder.[1] We show, exemplarily in a special case of defaultable Markov diffusion model of Carr and Linetsky (2006), that the price of a convertible bond with such a European conversion option can be stated in closed form in 1.5-factor models that allow for analytical expressions for survival probabilities and European equity options in the case that the bond does not include a soft call covenant. When a soft call right for the issuer is present, an analytical price formula for the convertible bond with European conversion option can still be computed in a simple credit-equity model, constituting a limiting case of the considered 1.5-factor model, where the default time of the issuing company is modeled independently from the diffusion driving the stock price. This model, although lacking important properties of the more advance defaultable Markov diffusions, may still be useful in a pre-calibration routine of a more advanced model or when screening a large number of convertible bonds for investment purposes, due to its computational efficiency compared to the standard pricing routines.

[1] The rarely present regular call rights will be ignored hereafter due to a lack of relevance, and regular put rights will be taken into account in a simplified manner, cf. Remark 5.2.1 below.

The results presented in this chapter are based on joint work with Jan-Frederik Mai. A simplified discussion, limited to convertible bonds with matching bond and stock currency and soft call right executable from the valuation day on, is published in Hüttner and Mai (2018). This chapter builds upon these results, giving a more detailed presentation as well as presenting extensions of the formulas to convertible bonds where the stock is denominated in a different currency and the soft call may only be executed after a contractually given future date, two cases which are not uncommon in the market, as our real-world examples show.

The remainder of this chapter is organized as follows: We present the analytical proxy formulas for the price of a convertible bond with and without soft call within the simple credit-equity model, and without soft call in the more advanced defaultable Markov diffusion model of Carr and Linetsky (2006) in Section 5.2. In Section 5.3, the sharpness of the derived lower bounds is illustrated for real-world convertible bonds, one with and one without soft call. Section 5.4 summarizes and concludes.

5.2 Analytical pricing formulas

In the following, we first clarify the general assumptions made concerning bond covenants as well as the model cosmos introduced in Section 5.2.1, and derive a model-free lower bound based on 'Europeanizing' the conversion option, valid under these assumptions, in Section 5.2.2. As mentioned above, we derive the lower bound for convertible bonds with and without soft call exemplarily in two models, a special case of the jump-to-default-extended constant elasticity of variance (*JDCEV*) 1.5-factor model of Carr and Linetsky (2006), and a limiting case thereof, referred to as the *simple credit-equity model* in the sequel. These models are introduced in Section 5.2.3, and the lower bounds for the cases without and with soft call are derived in Sections 5.2.4 and 5.2.5, respectively.

5.2.1 General modeling assumptions

We denote by $\{S_t\}_{t\geq 0}$ the stock price process of a company that has issued a convertible bond with market (dirty[2]) price process $\{B_t\}_{t\geq 0}$. Usually convert and stock price are denominated in the same currency, but there exist some converts in the market, also referred to as quanto-converts[3], where this does not hold, and one has to include the (stochastic) exchange rate in the model. As the influence of the exchange rate on pricing is only of a secondary nature, we consider this in a simplified way by introducing a deterministic exchange rate. Let $r_S(\cdot), r_B(\cdot)$ denote the "risk-free" discounting rate in the equity and bond currency, respectively, which are both assumed to be deterministic functions. The exchange rate between the two currencies is assumed to be the deterministic function $FX(t) = FX(0)\exp\{-\int_0^t r_S(u) - r_B(u)\mathrm{d}u\}$, with $FX(0) > 0$

[2]The dirty price of a bond includes the accumulated interest since the bond's last coupon payment.
[3]For more information on quanto-converts, see e.g. Bernhart and Mai (2018).

the currently observed exchange rate. We denote the conversion ratio of the bond at time t by $\alpha(t)$, and the annualized bond coupon rate by c. We assume that $\alpha(t)$ is a deterministic function[4].

It is obvious that $B_t \geq \alpha(t)\, FX(t)\, S_t$ for all time points t at which the holder of the convertible bond is allowed to exchange into shares. In the sequel, we are going to refine this lower bound under the following additional assumptions, denoting by $\tau > 0$ the random future time point at which the issuer defaults on the bond:

(A1) **Recovery rate:** If τ occurs before bond maturity, all holders of the convertible bond are assumed to receive the fraction $R \in [0, 1)$ of their nominal at τ, and the *recovery rate* R is assumed to be a non-random constant.

(A2) **Jump-to-default:** The company's equity drops to zero at τ, i.e. $S_t = 0$ for all $t \geq \tau$.

(A3) **Forward contracts:** The marketplace offers equity forward contracts for all maturities. We denote by $F(t_1, t_2)$ the forward strike price for a forward contract on the stock struck at t_1 with maturity $t_2 \geq t_1$. Ignoring discrete cash dividends, we assume that $F(t_1, t_2) = S_{t_1} \exp\{\int_{t_1}^{t_2} r_S(t) - \delta \, \mathrm{d}t\}$ with a constant rate $\delta \geq 0$ accounting for proceeds from stock possession, either through dividends or stock lending.

We decompose the parameter δ into two non-negative components $\delta = \delta_r + \delta_d$. The part δ_r corresponds to proceeds from stock possession that are not shared by convertible bond holders, while the part δ_d corresponds to proceeds from stock possession that are passed through to convert holders by an appropriate adjustment of the conversion ratio. Consequently, we assume $\alpha(t) = \alpha(0) \exp\{\delta_d \, t\}$ with $\alpha(0)$ denoting the current conversion ratio at time $t = 0$. The parameter δ_r is henceforth called *repo rate* because its typical economic interpretation is in terms of a yield that can be consumed as stock owner by stock lending via repurchasement agreements. It is further typical, but not always the case, that the conversion ratio is increased when the company pays dividends in order not to dilute convertible bond holders, so that δ_d typically agrees with the dividend yield. However, in situations when parts of the dividend payments do not entail a conversion ratio adjustment, this respective part of the dividend yield needs to be subtracted from δ_d and added to the repo margin δ_r.

(A4) **Convert covenants:** We assume that the holder of the convert is allowed to exchange one unit of bond nominal into $\alpha(t)$ shares within a certain conversion period that we assume to include the maturity date T of the bond. Typically, the conversion period ends precisely at or only a few business days before maturity,

[4] In reality, the conversion ratio is typically increased when the underlying share pays a dividend in order to let convertible bond holders participate in the dividend also before conversion. Future dividend payments are unknown today in reality, but known within our model, which implies that also future conversion ratios are modeled deterministically. For advice on the inclusion of random future dividend payments into pricing models see Bernhart and Mai (2015).

so that this assumption is (at least approximately) satisfied in most practical cases. We furthermore assume that there is a soft call right that allows the issuing company to enforce conversion into shares at a time point when parity breaches a level K after a specified (future) date $\theta \in [0, T)$. A typical level in the marketplace is $K = 1.3$, measured in units of bond nominal. The case $K = \infty$ means that there is no soft call right. For future reference, we denote the soft call trigger time point by

$$\eta := \inf\{t > \theta \,:\, \alpha(t)\, FX(t)\, S_t \geq K\}, \quad \theta \in [0, T).$$

Concerning further notation, we denote the coupon payment dates of the bond by $0 \leq t_1 < \ldots < t_n < t_{n+1} = T$, i.e. t_n denotes the penultimate coupon payment date and $T > t_n$ is the last coupon date, i.e. at maturity T the bond pays its last coupon cash payment of size $c\,(T - t_n)$. Further, $\Delta t_k := t_k - t_{k-1}$ denotes the year fraction between the two coupon payment dates t_{k-1} and t_k, for $k = 1, \ldots, n$, with t_0 denoting the last coupon payment before the evaluation date.

Remark 5.2.1 (Regular call rights and discrete put right)
We assume that the issuer has no (regular) call right, except for the soft call right mentioned in (A4). Regular call rights are quite uncommon in the marketplace, so that this assumption is not a big loss of relevance, but instead simplifies our derivation massively. A discrete put right is sometimes present in practice, although not very often. Simply ignoring the put right yields a lower bound for the price, because it relies on taking away an optionality from the convert holder. More accurately, but equally simple to implement, one may assume the convert holder is forced to make the put decision right now at $t = 0$, thus effectively having to decide between two different convertible bonds without put option. For both of these fictitious convertible bonds (one is the original bond only without the put, the other is the convertible bond under the assumption that the put is exercised for sure) one may compute the lower bound derived below, and the maximum of both lower bounds constitutes a typically quite sharp lower bound for the original convertible bond (with put right). Consequently, we henceforth ignore regular call and discrete put rights.

5.2.2 A generic lower bound

We consider the following trading strategy: at $t = 0$ we buy the convert at a price of B_0, and we enter an equity forward contract to sell $\alpha(T)$ shares at the last conversion time point T for the price $F(0, T)$. In the meantime we do not consider to exchange our bond into shares, unless we are forced to do so by soft call in case $\eta \leq T$. Consequently, we purposely drop our conversion right at all time points different from T. Technically, this intentional abstinence transforms our convertible bond with American conversion right into another convertible bond with European conversion right at T, whose price process we denote by $\{\bar{B}_t\}_{t \geq 0}$. Obviously, $B_t \geq \bar{B}_t$ for all $t \in [0, T]$, and $B_T = \bar{B}_T$, unless converted prior to maturity.

The implementation of this trading strategy costs us B_0 at inception $t = 0$. At time T there are the following scenarios:

- $\eta \leq \min\{T, \tau\}$:
  The bond is exchanged into $\alpha(\eta)$ shares at η, which are held until T. The value at time T of a portfolio holding these $\alpha(\eta)$ shares from η to T is given by

  $$\alpha(\eta)\, FX(T)\, S_T\, e^{(\delta_r+\delta_d)\,(T-\eta)} = \alpha(T)\, FX(T)\, S_T\, e^{\delta_r\,(T-\eta)},$$

  assuming that received dividends are immediately reinvested into more shares. In case $\eta \leq \tau \leq T$, this position is worthless at maturity, as the stock price drops to zero at τ according to assumption (A2). Furthermore, $\alpha(T)$ shares are sold at maturity T of the forward contract at the price $FX(T)\, F(0,T)$. Additionally, we earn coupons from the bond until η.

- $\tau \leq \min\{T, \eta\}$:
  The bond defaults and drops to recovery value R at τ. Furthermore, the share price drops to zero so that we can buy $\alpha(T)$ shares at time T at zero cost and deliver them into the forward contract yielding the value $\alpha(T)\, FX(T)\, F(0,T)$. Additionally, we earn coupons from the bond until τ.
  Note that $\{\tau < \eta\}$ implies $\{\eta = \infty\}$ due to the definition of η and assumption (A2) on the stock price process: In case default happens first, the stock price jumps to zero, and conversion never happens as $\alpha(t)\, FX(t)\, S_t = 0 < K$ for all $t \geq \tau$.

- $T < \min\{\eta, \tau\}$:
  At maturity T, the bond ends up at $B_T = \max\{1 + c\,(T - t_n), \alpha(T)\, FX(T)\, S_T\}$, and the forward contract has value $\alpha(T)\, FX(T)\, (F(0,T) - S_T)$, leaving us with the payoff $(1 + c(T - t_n) - \alpha(T)\, FX(T)\, S_T)^+ + \alpha(T)\, FX(T)\, F(0,T)$. Additionally, we earn coupons from the bond at all t_k.

Summing up, at time T the value of our portfolio (in bond currency) - if we pursue the trading strategy outlined above - is given by

$$\begin{aligned}
&\sum_{0<t_k<\min\{\tau,\eta,T\}} c\,\Delta t_k\, e^{\int_{t_k}^{T} r_B(s)\,\mathrm{d}s} + R\, e^{\int_{\tau}^{T} r_B(s)\,\mathrm{d}s}\, \mathbb{1}_{\{\tau\leq T,\, \tau<\eta\}} + \alpha(T)\, FX(T)\, F(0,T) \\
&\quad + \left(1 + c\,(T - t_n) - \alpha(T)\, FX(T)\, S_T\right)^+ \mathbb{1}_{\{T<\min\{\eta,\tau\}\}} \\
&\quad + \mathbb{1}_{\{\eta\leq\min\{T,\tau\}\}}\, \alpha(T)\, FX(T)\, S_T \left(e^{\delta_r\,(T-\eta)} - 1\right).
\end{aligned}$$

By replication arguments, the value B_0 must be greater or equal to $\bar{B}_0$, which equals the discounted expected value of the last expression. This constitutes a generic lower bound for the convertible bond's price.

Lemma 5.2.2 (Generic lower bound for convertible bond price)
Under the aforementioned assumptions the price of the convertible bond B_0 is bounded

from below by

$$\begin{aligned}\bar{B}_0 = c \sum_{0<t_k<T} \Delta t_k \, e^{-\int_0^{t_k} r_B(s)\,\mathrm{d}s} \, \mathbb{P}(\tau > t_k, \eta > t_k) \\ + R\,\mathbb{E}\Big[e^{-\int_0^{\tau} r_B(s)\,\mathrm{d}s}\,\mathbb{1}_{\{\tau\le\min\{T,\eta\}\}}\Big] + e^{-\int_0^{T} r_B(s)\,\mathrm{d}s}\,\alpha(T)\,FX(T)\,F(0,T) \\ + e^{-\int_0^{T} r_B(s)\,\mathrm{d}s}\,\mathbb{E}\Big[\big(1+c(T-t_n)-\alpha(T)\,FX(T)\,S_T\big)^+\,\mathbb{1}_{\{T<\min\{\tau,\eta\}\}}\Big] \\ + e^{-\int_0^{T} r_B(s)\,\mathrm{d}s}\,\mathbb{E}\Big[\mathbb{1}_{\{\eta\le T<\tau\}}\,\alpha(T)\,FX(T)\,S_T\,\Big(e^{\delta_r\,(T-\eta)}-1\Big)\Big].\end{aligned} \tag{5.1}$$

The last summand in the formula for $\bar{B}_0$ in Lemma 5.2.2 equals exactly zero if $\delta_r = 0$ and/or if there is no soft call. Our goal is to compute the obtained lower bound of Lemma 5.2.2 analytically. To this end, we seek analytical evaluations of the following expressions in specific models:

(i) the probabilities $\mathbb{P}\big(\tau > t_k, \eta > t_k\big)$,

(ii) the recovery term $\mathbb{E}\Big[e^{-\int_0^{\tau} r_B(s)\,\mathrm{d}s}\,\mathbb{1}_{\{\tau\le\min\{T,\eta\}\}}\Big]$,

(iii) the put-like term $\mathbb{E}\Big[\Big(1+c(T-t_n)-\alpha(T)\,FX(T)\,S_T\Big)^+\,\mathbb{1}_{\{T<\min\{\tau,\eta\}\}}\Big]$,

(iv) the expectation $\mathbb{E}\Big[\mathbb{1}_{\{\eta\le T<\tau\}}\,\alpha(T)\,FX(T)\,S_T\,\Big(e^{\delta_r\,(T-\eta)}-1\Big)\Big]$.

In the simpler case that no soft call right is present (i.e. $\eta = \infty$), closed-form evaluations are possible in defaultable Markov diffusion models that allow for analytical solutions of survival probabilities and European equity options, such as, e.g., the models in Carr and Linetsky (2006); Linetsky (2006). As already mentioned, the apparently most critical term (iv) even vanishes completely in this case. Exemplarily, we demonstrate the respective formulas in a special case of the JDCEV credit-equity model of Carr and Linetsky (2006) in Section 5.2.4.

In the more difficult case that a soft call is present, i.e. $\mathbb{P}(\eta < \infty) > 0$, (numerically efficient) closed form evaluations of all four expressions can only be obtained in the simple credit-equity model under the additional assumption of a constant interest rate $r_B(t) \equiv r_B$, due to the complexity of the required expressions. However, we demonstrate below in a real-world example how useful even the simple credit-equity model can be for a 'first-shot' price indication. The presented proxy formulas for the simple credit-equity model are useful for at least two applications: First, they can be used to pre-calibrate certain parameters of a more advanced model (e.g. one of the models in Carr and Linetsky (2006); Linetsky (2006)), thus speeding up parameter calibration in the latter models. Second, they are useful to screen the marketplace for potential bond investment ideas based on certain quantitative criteria. Due to their computational efficiency in comparison with PDE pricing techniques, it is possible to compute (approximative) model bond prices for a large set of parameters within fractions of a second, and one can use these prices to implement quantitative criteria that help quickly narrow down a long

list of potential investments. The remaining interesting bonds can then be analyzed with a more advanced model by means of more time-consuming methods. For instance, it might be reasonable for portfolio management to postulate a certain 'admissibility interval' for the ratio between those model parameters in the simple credit-equity model that can match an observed market price.

5.2.3 Considered credit-equity models

Simple credit-equity model

Possibly the simplest convertible bond pricing model which respects the jump-to-default assumption (A2) defines $(\tau, \{S_t\}_{t\geq 0})$ as follows:

$$S_t = S_0\, e^{X_t}\, \mathbb{1}_{\{\tau>t\}}, \quad X_t = \int_0^t \underbrace{r_S(s) - \delta + \lambda - \frac{\sigma^2}{2}}_{=:\gamma(s)}\, \mathrm{d}s + \sigma W_t,$$

where $\{W_t\}_{t\geq 0}$ is a Brownian motion and τ an independent exponential random variable with mean $1/\lambda$ (with infinite mean in the limiting case $\lambda = 0$ being allowed and corresponding to the Black–Scholes model). The market filtration is defined as the conjunction of the natural filtration of the Brownian motion and of the default indicator process $\{\mathbb{1}_{\{\tau>t\}}\}_{t\geq 0}$.
Calculations in this model benefit tremendously from the independence between the default time τ and the Brownian motion W driving the stock price diffusion, making it possible to obtain analytical formulas even in the case with soft call.

JDCEV model

We further consider a three-parametric special case of the so-called *JDCEV model* (jump-to-default-extended (JD) constant elasticity of variance (CEV) model) introduced in Carr and Linetsky (2006) that is defined as follows: The pre-default stock price is

$$\frac{\mathrm{d}S_t}{S_t} = \big(r_S(t) - \delta + \lambda_t\big)\,\mathrm{d}t + \sigma\Big(\frac{S_t}{S_0}\Big)^{\beta}\,\mathrm{d}W_t, \quad t < \tau,$$

where $\sigma > 0$, $\beta < 0$, and the default intensity λ_t is expressed as a function h of the stock price:

$$\lambda_t = h(S_t) = \lambda_0 \left(\frac{S_t}{S_0}\right)^{2\beta},$$

i.e. the default time τ is the first time that the integral over the default intensity exceeds the value of an exponentially distributed random variable with unit mean ϵ, that is independent of the Brownian motion $W = \{W_t\}_{t\geq 0}$:

$$\tau = \inf\Big\{t \geq 0 : \int_0^t \lambda_s \,\mathrm{d}s > \epsilon\Big\}.$$

At time τ, the stock price jumps to zero and remains there for eternity in accordance with (A2), i.e. $S_t = 0$ for all $t \geq \tau$. As $\beta \nearrow 0$, this model converges to the simple credit-equity model, so it constitutes a proper enhancement thereof. However, the parameter choice $\beta = 0$ is not admissible and the limiting process $\beta \nearrow 0$ is unstable in all derived formulas, which is one reason why we treat the simple credit-equity model separately. The parameter $\beta < 0$ in the JDCEV model introduces a reciprocal relationship between stock price and default intensity, which is not present in the simple credit-equity model. In particular, if the stock price process loses value massively, the convertible bond price will do so as well, which is not the case in the simple credit-equity model.
The mathematical derivations of this model are based on relations to Bessel processes and may be considered an enhancement of the work by Delbaen and Shirakawa (2002), who consider the pricing of equity derivatives within the CEV model without jump to default.

5.2.4 Lower bound in the case without soft call

We first consider the case without soft call, i.e. $\eta = \infty$, where an evaluation of term (iv) is not required. In the simple model, terms (i)-(iii) can be computed straightforward, as can be seen in the following lemma:

Lemma 5.2.3 (Formulas without soft call, simple model)
The relevant terms are given as follows:

$$\begin{aligned}
&\text{(i)}\ \mathbb{P}(\tau > t_k, \eta > t_k) = \mathbb{P}(\tau > t_k) = e^{-\lambda\, t_k},\\
&\text{(ii)}\ \mathbb{E}\Big[e^{-\int_0^\tau r_B(s)\,\mathrm{d}s}\, \mathbb{1}_{\{\tau \leq \min\{T,\eta\}\}}\Big] = \int_0^T e^{-\int_0^t r_B(s)\,\mathrm{d}s}\, \lambda\, e^{-\lambda\, t}\,\mathrm{d}t,\\
&\text{(iii)}\ \mathbb{E}\Big[\Big(1 + c\,(T - t_n) - \alpha(T)\, FX(T)\, S_T\Big)^+ \mathbb{1}_{\{T<\tau\}}\Big]\\
&\qquad = e^{-\lambda\, T}\,\big(1 + c\,(T - t_n)\big)\Phi\left(\frac{k_1 - \int_0^T \tilde{\gamma}(s)\,\mathrm{d}s}{\sigma\,\sqrt{T}}\right)\\
&\qquad\quad - \alpha(0)\, FX(0)\, S_0\, e^{\int_0^T r_B(s) - \delta_r\,\mathrm{d}s}\,\Phi\left(\frac{k_1 - \int_0^T \tilde{\gamma}(s) + \sigma^2\,\mathrm{d}s}{\sigma\,\sqrt{T}}\right),
\end{aligned}$$

with

$$k_1 = \log\Big(\frac{1 + c\,(T - t_n)}{\alpha(0)\, FX(0)\, S_0}\Big), \qquad \tilde{\gamma}(t) := r_B(t) - \delta_r + \lambda - \frac{\sigma^2}{2}.$$

Proof
Terms (i) and (ii) are computed straightforward using the exponential distribution of τ, and term (iii) can easily be computed using standard Black–Scholes computations, exploiting the independence of τ and W. All terms are easy to evaluate numerically. □

For our special case of the JDCEV model, the desired formulas can be extracted from the results in Carr and Linetsky (2006):

Lemma 5.2.4 (Formulas without soft call, JDCEV model)
The relevant expressions are given by the following formulas:

(i) The survival probabilities $\mathbb{P}(\tau > t_k)$ are given by

$$\left(\frac{x^2}{\xi(t_k)}\right)^{\frac{1}{2|\beta|}} \mathcal{M}\left(\frac{1}{2\beta}, \nu, \frac{x^2}{\xi(t_k)}\right).$$

(ii) The recovery term $\mathbb{E}\left[e^{-\int_0^\tau r_B(t)\,\mathrm{d}t}\,\mathbb{1}_{\{\tau\le T\}}\right]$ is given by

$$\lambda_0 \int_0^T e^{-\int_0^t r_B(s)+2\beta(r_S(s)-\delta)\,\mathrm{d}s} \left(\frac{x^2}{\xi(t)}\right)^{1-\frac{1}{2\beta}} \mathcal{M}\left(\frac{1}{2\beta}-1, \nu, \frac{x^2}{\xi(t)}\right)\mathrm{d}t$$

(iii) The put-like term $\mathbb{E}\left[\left(1 + c(T-t_n) - \alpha(T)\,FX(T)\,S_T\right)^+ \mathbb{1}_{\{T<\tau\}}\right]$ equals

$$\begin{aligned}&(1 + c(T-t_n))\left(\frac{x^2}{\xi(T)}\right)^{\frac{1}{2|\beta|}} \Theta^-\left(\frac{1}{2\beta}, \frac{k^2}{\xi(T)}, \nu, \frac{x^2}{\xi(T)}\right)\\ &\quad - \alpha(T)\,FX(T) e^{\int_0^T r_S(s)-\delta\,\mathrm{d}s} S_0 \Theta^-\left(0, \frac{k^2}{\xi(T)}, \nu, \frac{x^2}{\xi(T)}\right).\end{aligned}$$

In these formulas, we have applied the following notations:

$$x = \frac{S_0^{|\beta|}}{|\beta|}, \qquad \xi(t) = \int_0^t \frac{\sigma^2}{S_0^{2\,\beta}} e^{2\beta\int_0^u r_S(v)-\delta\,\mathrm{d}v}\mathrm{d}u,$$

$$\nu = \frac{2\lambda_0}{|\beta|\sigma^2} + \frac{1}{|\beta|} + 2, \qquad k = \frac{(1 + c(T-t_n))^{|\beta|}}{(\alpha(T)\,FX(T))^{|\beta|}|\beta|} e^{\beta\int_0^T r_S(s)-\delta\,\mathrm{d}s},$$

and the functions $\mathcal{M}(p,\nu,y)$ and $\Theta^-(p,k,\nu,y)$ refer to the p-th (truncated) moment of a generic random variable X following a noncentral χ^2-distribution with ν degrees of freedom and noncentrality parameter y, given by

$$\mathcal{M}(p,\nu,y) = \mathbb{E}[X^p], \qquad \Theta^-(p,k,\nu,y) = \mathbb{E}[X^p\,\mathbb{1}_{\{X\le k\}}].$$

Proof
All formulas can directly be extracted from the results in Carr and Linetsky (2006). □

The efficient numerical evaluation of the involved (truncated) moments of a non-central χ^2-variable is explained in Carr and Linetsky (2006), the practical implementation being based on an idea of Benton and Krishnamoorthy (2003).

5.2.5 Lower bound in the case with soft call

When a soft call is present, the derivation of an analytical formula becomes dramatically more difficult. In the simple model, we require the distribution of a Brownian motion with time-dependent drift $\tilde{\gamma}(t)$ and its running maximum, as well as the distribution of its first passage time to a fixed level.
Theoretically, for piecewise constant $r_B(t)$, at least the prior distribution can be evaluated via the technique described in Wang and Pötzelberger (1997); Pötzelberger and Wang (2001); Wang and Pötzelberger (2007), which is shown exemplarily[5] for term (i) in the following lemma.

Lemma 5.2.5 (Term (i), simple model, piecewise constant $r_B(t)$)
In case of a piecewise constant interest rate[6], one can exploit results of Wang and Pötzelberger (1997, 2007) to obtain

$$\mathbb{P}(\tau > t, \eta > t) = e^{-\lambda t} \int_{-\infty}^{k_2} \mathbb{E}[g(W_{\nu_1}, \dots, W_{\nu_m} | d(\cdot; x))] f_{\mathcal{N}(\tilde{\Gamma}_\theta, \sigma^2\theta)}(x) \mathrm{d}x,$$

with $\nu_k + \theta$ denoting the times of discontinuity of the piecewise constant interest rate in $[\theta, T]$, $\tilde{\Gamma} = \int_0^\theta \tilde{\gamma}(u)\mathrm{d}u$, and the time-dependent barrier d and the function g are given as follows:

$$g(x_1, \dots, x_m | d) = \prod_{j=1}^{m} \mathbb{1}_{\{x_j < d(\nu_j)\}} \Big(1 - \exp\Big\{ - \frac{2(d(\nu_{j-1}) - x_{j-1})(d(\nu_j) - x_j)}{\nu_j - \nu_{j-1}} \Big\} \Big),$$
$$d(t; x) = \frac{1}{\sigma} \Big(k_2 - x - \int_0^t \tilde{\gamma}(u + \theta) \mathrm{d}u \Big).$$

Proof
By exploiting independence of τ and W, we rewrite

$$\begin{aligned}\mathbb{P}(\tau > t, \eta > t) &= \mathbb{P}(\tau > t) \mathbb{P}\Big(\max_{\theta < s \le t} \int_0^s \tilde{\gamma}(u)\mathrm{d}u + \sigma W_s < k_2, \int_0^\theta \tilde{\gamma}(u)\mathrm{d}u + \sigma W_\theta \le k_2 \Big) \\ &= e^{-\lambda t} \int_{-\infty}^{k_2} \mathbb{P}\Big(\max_{0 < s < t - \theta} \int_0^s \tilde{\gamma}(u + \theta)\mathrm{d}u + \sigma W_s < k_2 - x \Big) f_{\mathcal{N}(\tilde{\Gamma}_\theta, \sigma^2\theta)}(x) \mathrm{d}x.\end{aligned}$$

By Wang and Pötzelberger (2007, Corollary 16), which relates the boundary crossing probability of a geometric Brownian motion with time-dependent drift to the probability

[5] Term (ii) is calculated with the help of term (i) as in the case of a flat interest rate, cf. Lemma 5.2.7, and term (iii) can, with slight modifications of the arguments in Wang and Pötzelberger (1997, 2007), be computed in a similar fashion as term (i).

[6] It is a common approach among practitioners to obtain discount factors $\exp\{-\int_0^t r_\cdot(u)\mathrm{d}u\}$ from market quotes of liquidly traded interest rate swaps and linearly interpolate the logarithms of these discount factors, which corresponds to a piecewise linear interest rate, cf. 'raw' interpolation method in Hagan and West (2006).

of a standard Brownian motion crossing a piecewise continuous (and in our case even piecewise linear) boundary, the probability under the integral is given by

$$\begin{aligned}&\mathbb{P}\Big(\max_{0<s<t-\theta}\int_0^s \tilde{\gamma}(u+\theta)\mathrm{d}u + \sigma W_s < k_2 - x\Big)\\ &= \mathbb{P}\Bigg(W_s \leq \underbrace{\frac{1}{\sigma}\Big(k_2 - x - \int_0^s \tilde{\gamma}(u+\theta)\mathrm{d}u\Big)}_{=:d(s;x)} \;\forall\, s \leq t-\theta\Bigg).\end{aligned}$$

Finally, by Wang and Pötzelberger (1997, Theorem 1), which states the probability of a standard Brownian motion crossing a piecewise linear boundary, this expression equals

$$\mathbb{P}(W_s \leq d(s)\,\forall\, s \leq t-\theta) = \mathbb{E}[g(W_{\nu_1},\ldots,W_{\nu_m}|d)],$$

with with ν_k denoting the times of discontinuity of the piecewise linear boundary and g given as above. □

However, as can be seen from Lemma 5.2.5, the numerical evaluation still requires integration with respect to an m-dimensional normal density, where m depends on the number of discontinuities of $r_B(t)$ in $[\theta, T]$, which is too high in practice to yield an improvement over pricing the convertible bond via PDE methods right away.
Consequently, we have to assume that $r_B(t) \equiv r_B$ is constant[7], in which case the distribution of the maximum of a Brownian motion with drift $\tilde{\gamma}(t) \equiv \tilde{\gamma}$ is known in closed-form, cf. Musiela and Rutkowski (2009, Proposition A.18.2, Corollary A.18.2):

Lemma 5.2.6 (Distribution of a Brownian motion and its running maximum) Consider a Brownian motion $\{Y_t\}_{t\geq 0}$ with drift m and diffusion parameter s, i.e. $Y_t = mt + sW_t$, where W is a standard Brownian motion. The joint distribution of Y_T and its running maximum $\max_{0\leq t\leq T} Y_t$ is given by

$$\mathbb{P}\Big(Y_T \leq y_1, \max_{0\leq t\leq T} Y_t \leq y_2\Big) = \Phi\Big(\frac{y_1 - mT}{s\sqrt{T}}\Big) - e^{\frac{2my_2}{s^2}}\Phi\Big(\frac{y_1 - 2y_2 - mT}{s\sqrt{T}}\Big).$$

The distribution of the running maximum immediately follows to be

$$\mathbb{P}\Big(\max_{0\leq t\leq T} Y_t \leq y\Big) = \Phi\Big(\frac{y - mT}{s\sqrt{T}}\Big) - e^{\frac{2my}{s^2}}\Phi\Big(\frac{-y - mT}{s\sqrt{T}}\Big).$$

For future reference define

$$\begin{aligned}F(m,s,T,y_1,y_2) &:= \Phi\Big(\frac{\min\{y_1,y_2\} - m\,T}{s\sqrt{T}}\Big) - e^{\frac{2\,m\,y_2}{s^2}}\,\Phi\Big(\frac{\min\{y_1,y_2\} - 2\,y_2 - m\,T}{s\sqrt{T}}\Big)\\ F(m,s,T,y) &= F(m,s,T,y,y).\end{aligned} \tag{5.2}$$

[7] The presented formulas can easily be adjusted to consider $r_B(t)$ deterministic for $t \leq \theta$ and $r_B(t) \equiv r_B$ constant for $t > \theta$. One only has to pay attention that the correct drift term is considered in the restarted Brownian motion in the proofs, similarly as in Lemma 5.2.5. We omit this discussion for the sake of simplicity.

In practice, r_B can be chosen as some average value of $\{r_B(t)\}_{t\in[0,T]}$ or be optimized in order to yield the best lower bound on the observed valuation, and the resulting convertible bond lower bound must be viewed as an approximation. The resulting formulas are stated in the following lemma.

Lemma 5.2.7 (Formulas with soft call, simple model, case $r_B(t) \equiv r_B$)
Under the assumption that $r_B(t) \equiv r_B$, and hence $\tilde{\gamma}(t) \equiv \tilde{\gamma}$, the following formulas are valid:

(i) The probabilities $\mathbb{P}(\tau > t_k, \eta > t_k)$ are given by

$$\mathbb{P}(\tau > t, \eta > t) = \begin{cases} \mathbb{P}(\tau > t) = e^{-\lambda t} & \text{for } 0 \le t \le \theta \\ \mathbb{P}\left(\tau > t, \max_{\theta < s \le t} \alpha(s)\, FX(s)\, S_s < K\right) & \text{for } t > \theta, \end{cases}$$

where the latter probability is given by

$$e^{-\lambda t} \int_{-\infty}^{k_2} F(\tilde{\gamma}, \sigma, t-\theta, k_2 - x) f_{\mathcal{N}(\tilde{\gamma}\theta, \sigma^2\theta)}(x) \mathrm{d}x,$$

with $f_{\mathcal{N}(m,s^2)}$ denoting the density of a normally distributed random variable with mean m and variance s^2, F as defined in (5.2), and

$$k_2 = \log\Big(\frac{K}{\alpha(0)\, FX(0)\, S_0}\Big).$$

(ii) For the recovery part $\mathbb{E}\Big[e^{-r_B\,\tau}\, \mathbb{1}_{\{\tau \le \min\{T,\eta\}\}}\Big]$, we need to distinguish between the cases where default happens before and after θ, respectively:

$$\mathbb{E}\Big[e^{-r_B\tau}\, \mathbb{1}_{\{\tau \le T, \eta > \tau\}}\Big] = \mathbb{E}\Big[e^{-r_B\tau}\, \mathbb{1}_{\{\tau \le \theta, \eta > \tau\}}\Big] + \mathbb{E}\Big[e^{-r_B\tau}\, \mathbb{1}_{\{\theta < \tau \le T, \eta > \tau\}}\Big].$$

The first expectation equals the corresponding term in the case without soft call, as by definition $\eta > \theta \ge \tau$:

$$\mathbb{E}\Big[e^{-r_B\tau}\, \mathbb{1}_{\{\tau \le \theta, \eta > \tau\}}\Big] = \mathbb{E}\Big[e^{-r_B\tau}\, \mathbb{1}_{\{\tau \le \theta\}}\Big] = \int_0^\theta \lambda e^{-(r_B+\lambda)u} \mathrm{d}u.$$

The second expectation is given by

$$\mathbb{E}\Big[e^{-r_B\tau}\, \mathbb{1}_{\{\theta < \tau \le T, \eta > \tau\}}\Big] = \int_\theta^T \lambda e^{-r_B t}\, \mathbb{P}(\tau > t, \eta > t) \mathrm{d}t,$$

with the probability under the integral as given in part (i). This results in a double integral.

(iii) The put-like term $\mathbb{E}\Big[\Big(1+c(T-t_n)-\alpha(T)\,FX(T)\,S_T\Big)^+ \mathbb{1}_{\{T<\min\{\tau,\eta\}\}}\Big]$ is given by

$$\begin{aligned}
&\mathbb{E}\Big[\Big(1+c(T-t_n)-\alpha(T)\,FX(T)\,S_T\Big)^+ \mathbb{1}_{\{T<\min\{\tau,\eta\}\}}\Big]\\
&= (1+c(T-t_n))e^{-\lambda T}\int_{-\infty}^{k_2} F(\tilde{\gamma},\sigma,T-\theta,k_1-x,k_2-x) f_{\mathcal{N}(\tilde{\gamma}\theta,\sigma^2\theta)}(x)\mathrm{d}x\\
&\quad - \alpha(0)FX(0)S_0 e^{(r_B-\delta_r)T}\\
&\qquad \int_{-\infty}^{k_2} F(\tilde{\gamma}+\sigma^2,\sigma,T-\theta,k_1-x,k_2-x) f_{\mathcal{N}((\tilde{\gamma}+\sigma^2)\theta,\sigma^2\theta)}(x)\mathrm{d}x,
\end{aligned}$$

with k_2 as in part (i), k_1 as in Lemma 5.2.3 above, and F given by (5.2).

(iv) The expectation $\mathbb{E}\Big[\mathbb{1}_{\{\eta\le T<\tau\}}\,\alpha(T)\,FX(T)\,S_T\left(e^{\delta_r\,(T-\eta)}-1\right)\Big]$ is given by

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{1}_{\{\eta\le T<\tau\}}\,\alpha(T)\,FX(T)\,S_T\left(e^{\delta_r\,(T-\eta)}-1\right)\Big]\\
&= \alpha(0)\,FX(0)\,S_0 e^{(r_B-\delta_r)T}\left(\underbrace{\tilde{\mathbb{E}}\Big[\mathbb{1}_{\{\eta\le T\}}e^{\delta_r\,(T-\eta)}\Big|\tau>T\Big]}_{:=(B)} - \underbrace{\tilde{\mathbb{P}}(\eta\le T|\tau>T)}_{:=(A)}\right),
\end{aligned}$$

$$\begin{aligned}
(A) &= \Phi\Big(\frac{-k_2+(\tilde{\gamma}+\sigma^2)\theta}{\sigma\sqrt{\theta}}\Big)\\
&\quad + \int_{-\infty}^{k_2}(1-F(\tilde{\gamma}+\sigma^2,\sigma,T-\theta,k_2-x)) f_{\mathcal{N}((\tilde{\gamma}+\sigma^2)\theta,\sigma^2\theta)}(x)\mathrm{d}x,\\
(B) &= e^{\delta_r(T-\theta)}\Bigg(\Phi\Big(\frac{-k_2+(\tilde{\gamma}+\sigma^2)\theta}{\sigma\sqrt{\theta}}\Big)\\
&\quad + \int_{-\infty}^{k_2}\left[e^{\frac{l}{m}-b}\Phi\left(\frac{b(T-\theta)-l}{\sqrt{l(T-\theta)}}\right)+e^{\frac{l}{m}+b}\Phi\Big(-\frac{b(T-\theta)+l}{\sqrt{l(T-\theta)}}\Big)\right]\\
&\qquad \times f_{\mathcal{N}((\tilde{\gamma}+\sigma^2)\theta,\sigma^2\theta)}(x)\mathrm{d}x\Bigg),
\end{aligned}$$

with k_2 as in part (i) above, $l=(k_2-x)^2/\sigma^2$, $m=(k_2-x)/(\tilde{\gamma}+\sigma^2)$, and $b=\sqrt{l^2/m^2+2\,\delta_r\,l}$.

Proof

The derivations of terms (i), (ii), and (iii) rely heavily on the independence of τ and $\{W_t\}_{t\ge0}$, as well as on knowledge about the (joint) distribution of the Brownian motion with drift and its running maximum, cf. Lemma 5.2.6, and are obtained via integrating out the random future value of the stock price diffusion at θ and considering a Brownian motion restarted at θ in this value.

Specifically, the derivations proceed as follows:

(i) The probability $\mathbb{P}(\tau > t, \eta > t)$ for $t > \theta$ is calculated as follows:

$$\begin{aligned}
&\mathbb{P}\left(\tau > t, \max_{\theta<s\leq t} \alpha(s)\, FX(s)\, S_s < K\right) \\
&= \mathbb{P}(\tau > t)\mathbb{P}\Big(\max_{\theta<s\leq t} \alpha(0)\, FX(0)\, S_0 e^{\tilde{\gamma}s+\sigma W_s} < K\Big) \\
&= e^{-\lambda t}\Big[\mathbb{P}\Big(\max_{\theta<s\leq t} \tilde{\gamma}s + \sigma W_s < k_2, \tilde{\gamma}\theta + \sigma W_\theta \leq k_2\Big) \\
&\quad + \mathbb{P}\Big(\max_{\theta<s\leq t} \tilde{\gamma}s + \sigma W_s < k_2, \tilde{\gamma}\theta + \sigma W_\theta > k_2\Big)\Big],
\end{aligned}$$

exploiting the independence of τ and W. The second probability in the last line equals zero: It corresponds to the probability of the running maximum of a Brownian motion restarted at time θ in $\tilde{\gamma}\theta + \sigma W_\theta > k_2$ hitting the barrier k_2. Due to continuity of the paths of a Brownian motion, this running maximum cannot be below k_2. The final expression is then obtained by integrating out the random future value of the diffusion $\tilde{\gamma}t + \sigma W_t$ at $t = \theta$ and considering a restarted Brownian motion, i.e.

$$\begin{aligned}
&e^{-\lambda t}\,\mathbb{P}\Big(\max_{\theta<s\leq t} \tilde{\gamma}s + \sigma W_s < k_2, \tilde{\gamma}\theta + \sigma W_\theta \leq k_2\Big) \\
&= e^{-\lambda t}\,\mathbb{E}\Big[\mathbb{1}_{\{\tilde{\gamma}\theta+\sigma W_\theta \leq k_2\}}\mathbb{P}\Big(\max_{0<u\leq t-\theta} \tilde{\gamma}u + \sigma W'_u < k_2 - x \Big| \tilde{\gamma}\theta + \sigma W_\theta = x \leq k_2\Big)\Big] \\
&= e^{-\lambda t}\int_{-\infty}^{k_2} F(\tilde{\gamma}, \sigma, t-\theta, k_2 - x) f_{\mathcal{N}(\tilde{\gamma}\theta, \sigma^2\theta)}(x)\mathrm{d}x,
\end{aligned}$$

where W' denotes an independent copy of W.

(ii) The recovery term for $\theta < \tau \leq T$ is calculated as follows:

$$\begin{aligned}
&\mathbb{E}\Big[e^{-r_B\tau}\,\mathbb{1}_{\{\theta<\tau\leq T, \eta>\tau\}}\Big] = \mathbb{E}\Big[e^{-r_B\tau}\,\mathbb{1}_{\{\theta<\tau\leq T\}}\,\mathbb{P}\Big(\max_{\theta<s<\tau} \tilde{\gamma}s + \sigma W_s < k_2 | \theta < \tau \leq T\Big)\Big] \\
&= \int_\theta^T e^{-r_B t}\,\mathbb{P}\Big(\max_{\theta<s<t} \tilde{\gamma}s + \sigma W_s < k_2\Big)\,\lambda e^{-\lambda t}\mathrm{d}t = \int_\theta^T \lambda e^{-r_B t}\,\mathbb{P}(\tau > t, \eta > t)\mathrm{d}t.
\end{aligned}$$

(iii) The put-like term is rewritten as

$$\begin{aligned}
&\mathbb{E}\Big[\Big(1 + c(T-t_n) - \alpha(T)\, FX(T)\, S_T\Big)^+ \mathbb{1}_{\{T<\min\{\tau,\eta\}\}}\Big] \\
&= (1 + c(T-t_n))\mathbb{P}\Big(\tau > T, \eta > T, 1 + c(T-t_n) > \alpha(T)\, FX(T)\, S_T\Big) \\
&\quad - \mathbb{E}\Big[\alpha(T)\, FX(T)\, S_T\, \mathbb{1}_{\{\tau>T,\eta>T,1+c(T-t_n)>\alpha(T)\, FX(T)\, S_T\}}\Big].
\end{aligned}$$

Reformulating the term in the probability/indicator yields

$$\begin{aligned}
&\{\tau > T, \eta > T, 1 + c(T-t_n) > \alpha(T)\, FX(T)\, S_T\} \\
&= \{\tau > T, \max_{\theta<s\leq T} \tilde{\gamma}s + \sigma W_s < k_2, \tilde{\gamma}T + \sigma W_T < k_1\} \\
&= \{\tau > T, \max_{\theta<s\leq T} \tilde{\gamma}s + \sigma W_s < k_2, \tilde{\gamma}T + \sigma W_T < k_1, \tilde{\gamma}\theta + \sigma W_\theta \leq k_2\},
\end{aligned}$$

with a similar argument as in the proof of part (i). Again conditioning on the random diffusion value at θ and considering a restarted Brownian motion, we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\tau > T, \eta > T, 1 + c(T - t_n) > \alpha(T)\, FX(T)\, S_T\Big) \\
&= \mathbb{P}(\tau > T)\mathbb{P}\Big(\max_{\theta < s \le T} \tilde{\gamma} s + \sigma W_s < k_2, \tilde{\gamma} T + \sigma W_T < k_1, \tilde{\gamma}\theta + \sigma W_\theta \le k_2\Big) \\
&= e^{-\lambda T} \int_{-\infty}^{k_2} F(\tilde{\gamma}, \sigma, T - \theta, k_1 - x, k_2 - x) f_{\mathcal{N}(\tilde{\gamma}\theta, \sigma^2\theta)}(x) \mathrm{d}x,
\end{aligned}$$

and, after a change of measure to $\tilde{\mathbb{P}}$, with $\tilde{W}_t = W_t - \sigma t$ being a $\tilde{\mathbb{P}}$-Brownian motion, the second part is calculated analogously:

$$\begin{aligned}
&\mathbb{E}\Big[\alpha(T)\, FX(T)\, S_T\, \mathbb{1}_{\{\tau > T, \eta > T, 1 + c(T - t_n) > \alpha(T)\, FX(T)\, S_T\}}\Big] \\
&= \alpha(0) FX(0) S_0 e^{(r_B - \delta_r + \lambda)T} \tilde{\mathbb{P}}\Big(\tau > T, \eta > T, 1 + c(T - t_n) > \alpha(T) FX(T) S_T\Big) \\
&= \alpha(0) FX(0) S_0 e^{(r_B - \delta_r)T} \\
&\quad \times \int_{-\infty}^{k_2} F(\tilde{\gamma} + \sigma^2, \sigma, T - \theta, k_1 - x, k_2 - x) f_{\mathcal{N}((\tilde{\gamma} + \sigma^2)\theta, \sigma^2\theta)}(x) \mathrm{d}x.
\end{aligned}$$

The calculation of part (iv) additionally makes use of knowledge about the first hitting time of a fixed level of a Brownian motion with drift. We provide a sketch of the computation in the sequel. A change of measure to $\tilde{\mathbb{P}}$, with $\tilde{W}_t = W_t - \sigma t$ being a $\tilde{\mathbb{P}}$-Brownian motion, and exploiting the independence of τ and W yields

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{1}_{\{\eta \le T < \tau\}}\, \alpha(T)\, FX(T)\, S_T \left(e^{\delta_r (T - \eta)} - 1\right)\Big] \\
&= \alpha(0)\, FX(0)\, S_0 e^{(\tilde{\gamma} + \frac{\sigma^2}{2})T}\, \tilde{\mathbb{E}}\Big[\mathbb{1}_{\{\eta \le T, \tau > T\}} \left(e^{\delta_r (T - \eta)} - 1\right)\Big] \\
&= \alpha(0)\, FX(0)\, S_0 e^{(r_B - \delta_r + \lambda)T}\, \tilde{\mathbb{E}}\Big[\mathbb{1}_{\{\tau > T\}} \tilde{\mathbb{E}}\Big[\mathbb{1}_{\{\eta \le T\}} \left(e^{\delta_r (T - \eta)} - 1\right) \Big| \tau > T\Big]\Big] \\
&= \alpha(0)\, FX(0)\, S_0 e^{(r_B - \delta_r)T} \left(\underbrace{\tilde{\mathbb{E}}\Big[\mathbb{1}_{\{\eta \le T\}} e^{\delta_r (T - \eta)} \Big| \tau > T\Big]}_{:= (B)} - \underbrace{\tilde{\mathbb{P}}(\eta \le T | \tau > T)}_{:= (A)} \right).
\end{aligned}$$

To calculate the probability (A), we rewrite

$$\begin{aligned}
&\{\eta \le T | \tau > T\} = \Big\{ \max_{\theta < s \le T} \alpha(s) FX(s) S_s \ge K \Big| \tau > T \Big\} = \Big\{ \max_{\theta < s \le T} \tilde{\gamma} s + \sigma W_s \ge k_2 \Big\} \\
&= \underbrace{\Big\{ \max_{\theta < s \le T} \tilde{\gamma} s + \sigma W_s \ge k_2, \tilde{\gamma}\theta + \sigma W_\theta > k_2 \Big\}}_{= \{\tilde{\gamma}\theta + \sigma W_\theta > k_2\}} \cup \Big\{ \max_{\theta < s \le T} \tilde{\gamma} s + \sigma W_s \ge k_2, \tilde{\gamma}\theta + \sigma W_\theta \le k_2 \Big\},
\end{aligned}$$

thus obtaining

$$\tilde{\mathbb{P}}(\eta \le T|\tau > T) = \tilde{\mathbb{P}}(\tilde{\gamma}\theta + \sigma W_\theta > k_2) + \tilde{\mathbb{P}}\Big(\max_{\theta<s\le T} \tilde{\gamma}s + \sigma W_s \ge k_2, \tilde{\gamma}\theta + \sigma W_\theta \le k_2\Big)$$
$$= \Phi\Big(\frac{-k_2 + (\tilde{\gamma}+\sigma^2)\theta}{\sigma\sqrt{\theta}}\Big) + \int_{-\infty}^{k_2} (1 - F(\tilde{\gamma}+\sigma^2, \sigma, T-\theta, k_2 - x)) f_{\mathcal{N}((\tilde{\gamma}+\sigma^2)\theta, \sigma^2\theta)}(x)\mathrm{d}x,$$

with the latter probability calculated similarly as in part (i).
The remaining expectation term (B) is computed as follows:

$$\tilde{\mathbb{E}}\Big[\mathbb{1}_{\{\eta\le T\}} e^{\delta_r(T-\eta)}\Big|\tau > T\Big] = \tilde{\mathbb{E}}\Big[\mathbb{1}_{\{\max_{\theta<s\le T}(\tilde{\gamma}+\sigma^2)s+\sigma\tilde{W}_s\ge k_2\}} e^{\delta_r(T-\eta)}\Big|\tau > T\Big]$$
$$= \tilde{\mathbb{E}}\Big[\underbrace{\mathbb{1}_{\{\max_{\theta<s\le T}(\tilde{\gamma}+\sigma^2)s+\sigma\tilde{W}_s\ge k_2\}}}_{=1 \text{ a.s., and } \eta=\theta \text{ a.s.}} e^{\delta_r(T-\eta)} \mathbb{1}_{\{(\tilde{\gamma}+\sigma^2)\theta+\sigma\tilde{W}_\theta > k_2\}}\Big|\tau > T\Big]$$
$$+ \tilde{\mathbb{E}}\Big[\mathbb{1}_{\{\max_{\theta<s\le T}(\tilde{\gamma}+\sigma^2)s+\sigma\tilde{W}_s\ge k_2\}} e^{\delta_r(T-\eta)} \mathbb{1}_{\{(\tilde{\gamma}+\sigma^2)\theta+\sigma\tilde{W}_\theta \le k_2\}}\Big|\tau > T\Big]$$
$$= e^{\delta_r(T-\theta)}\Phi\Big(\frac{-k_2 + (\tilde{\gamma}+\sigma^2)\theta}{\sigma\sqrt{\theta}}\Big)$$
$$+ \int_{-\infty}^{k_2} \tilde{\mathbb{E}}\Big[\mathbb{1}_{\{\max_{0<u\le T-\theta}(\tilde{\gamma}+\sigma^2)u+\sigma\tilde{W'}_u\ge k_2-x\}} e^{\delta_r(T-\theta-\tilde{\eta}_x)}\Big|\tau > T\Big] f_{\mathcal{N}((\tilde{\gamma}+\sigma^2)\theta, \sigma^2\theta)}(x)\mathrm{d}x,$$

where $\tilde{\eta}_x|\tau > T$ for $\eta \le T$ corresponds to the first passage time of the restarted Brownian motion $\tilde{W'}$ to the level $k_2 - x$:

$$\begin{aligned}
&\eta|\tau > T = \inf\{t \in (\theta, \tau) : \tilde{\gamma}t + \sigma W_t \ge k_2\} \\
&\eta|\{\tau > T, \tilde{\gamma}\theta + \sigma W_\theta = x\} \\
&= \inf\{t \in (\theta, \tau) : \tilde{\gamma}(t-\theta) + \sigma(W_t - W_\theta) \ge k_2 - x\}|\{\tilde{\gamma}\theta + \sigma W_\theta = x\} \\
&\overset{d}{=} \inf\{u \in (0, \tau-\theta) : \tilde{\gamma}u + \sigma W'_u \ge k_2 - x\} := \tilde{\eta}_x|\tau > T.
\end{aligned}$$

The inner expectation term can then be calculated using the fact that $\tilde{\eta}_x|\tau > T$ has the same distribution (under $\tilde{\mathbb{P}}$) as the first passage time of the restarted Brownian motion with drift $\tilde{\gamma} + \sigma^2$ and diffusion parameter σ to the level $k_2 - x$. This is an Inverse Gaussian distribution, cf. Musiela and Rutkowski (2009, Proposition A18.1), with density and parameters

$$f_{\tilde{\eta}_x|\tau>T}(z) = \sqrt{\frac{l}{2\pi z^3}} \exp\Big\{-\frac{l}{2m^2 z}(z-m)^2\Big\}, \quad l = \Big(\frac{k_2 - x}{\sigma}\Big)^2, \quad m = \frac{k_2 - x}{\tilde{\gamma} + \sigma^2}.$$

Therefore,

$$\tilde{\mathbb{E}}\Big[\mathbb{1}_{\{\max_{0<u\le T-\theta}(\tilde{\gamma}+\sigma^2)u+\sigma\tilde{W'}_u\ge k_2-x\}} e^{\delta_r(T-\theta-\tilde{\eta}_x)}\Big|\tau > T\Big]$$
$$= \tilde{\mathbb{E}}\Big[\mathbb{1}_{\{\tilde{\eta}_x\le T-\theta\}} e^{\delta_r(T-\theta-\tilde{\eta}_x)}\Big|\tau > T\Big] = e^{\delta_r(T-\theta)} \int_0^{T-\theta} e^{-\delta_r z} f_{\tilde{\eta}_x|\tau>T}(z)\mathrm{d}z.$$

Introducing $c = \delta_r + l/(2m^2)$ and $b = \sqrt{2lc}$, the integral term can then be calculated explicitly:

$$\int_0^{T-\theta} e^{-\delta_r z} f_{\tilde{\eta}_x|\tau>T}(z)\mathrm{d}z = e^{\frac{l}{m}} \int_0^{T-\theta} \sqrt{\frac{l}{2\pi z^3}} \exp\Big\{ - \underbrace{z\Big(\frac{l}{2m^2} + \delta_r\Big)}_{=:c} - \frac{l}{2z}\Big\}\mathrm{d}z$$

$$\underbrace{\qquad\qquad}_{=:y,\text{ change of variables}}$$

$$= e^{\frac{l}{m}} \int_0^{c(T-\theta)} \sqrt{\frac{lc}{2\pi y^3}} e^{-y-\frac{lc}{2y}}\mathrm{d}y = e^{\frac{l}{m}-b} \int_0^{\frac{b^2(T-\theta)}{2l}} \sqrt{\frac{b^2/2}{2\pi y^3}} e^{-\frac{b^2/2}{2\cdot b^2/4\cdot y}(y-\frac{b}{2})^2}\mathrm{d}y$$

$$= e^{\frac{l}{m}-b}\Phi\bigg(\frac{b(T-\theta)-l}{\sqrt{l(T-\theta)}}\bigg) + e^{\frac{l}{m}+b}\Phi\bigg(-\frac{b(T-\theta)+l}{\sqrt{l(T-\theta)}}\bigg).$$

In the penultimate equality one integrates over an Inverse Gaussian density with parameters $b^2/2$ and $b/2$, and exploits the known closed form of its distribution function. □

In the JDCEV model, the dependence between credit and equity induces a complicated dependence structure between η and τ, which makes it impossible for us to arrive at analytical formulas for the desired quantities. Already in the calculation of the simplest term (i), one requires the joint distribution of a Bessel process, its running minimum, and (a complicated function of) its running maximum, for which, to the best of our knowledge, no closed-form results are available.

5.3 Examples

In the following, we illustrate the performance of the presented approximation formulas by considering two real world examples where the bonds are denominated in different currencies as their corresponding stocks, one with and one without soft call. In both cases, the lower bound is very sharp, see Figures 5.1 and 5.3. Only in rarely encountered market conditions, or for extreme (and unrealistic) parameter choices, the lower bound differs significantly from the price computed via the more involved PDE method, see Figures 5.2 and 5.4.
Similar observations are made in Hüttner and Mai (2018) for two different convertible bonds, which are denominated in the same currency as their underlying equity, and with $\theta = 0$ in the case with soft call.

5.3.1 An example without soft call

We consider a convertible bond without soft call, with a (dirty) price of $B_0 = 105.547\%$. The bond is denominated in a different currency than the underlying equity, with a current exchange rate of $FX(0) = 1.1500$, and an equity value of $S_0 = 47.60$. The

conversion ratio is $\alpha(0) = 0.0171622555$, the coupon rate is $c = 0.5\%$, and maturity is $T = 3.8959$. We furthermore assume a recovery rate of $R = 0.3$, cautiously lowering the common market assumption of 0.4, cf. Cont and Kan (2011). The company currently pays dividends of 5%, which are currently not passed on to convertible bond holders through an adjustment of the conversion ratio. This is a very atypical feature in a convertible bond. A reasonable assumption on the repo margin parameter is 1%, so $\delta_r = 0.06$ and $\delta_d = 0$. The interest rates $r_B(t)$ and $r_S(t)$ are bootstrapped from observed market prices for liquidly traded interest rate swaps.
This bond is convertible only in the time period $t \in [3.6466, 3.7863]$, so we expect the lower bound to be very sharp, even though the ECP should be non-negligible, as the proceeds from stock possession which are not passed through to convertible bond holders significantly outweigh the small earnings from coupon payments.
Since no soft call is present, we are able to use the more advanced JDCEV model to evaluate an analytical lower bound for the price. The assumed model parameters are $\beta = -0.3$, $\sigma = 32.35\%$, and $\lambda = 30$ bps, which are chosen as a result of a calibration to the observed market price as well as a trader's expertise. Figure 5.1 visualizes the sensitivities of the model price, evaluated via the PDE method of Andersen and Buffum (2004), with respect to changes in the underlying stock price, as well as the lower bound computed with help of the formulas in Lemma 5.2.4.

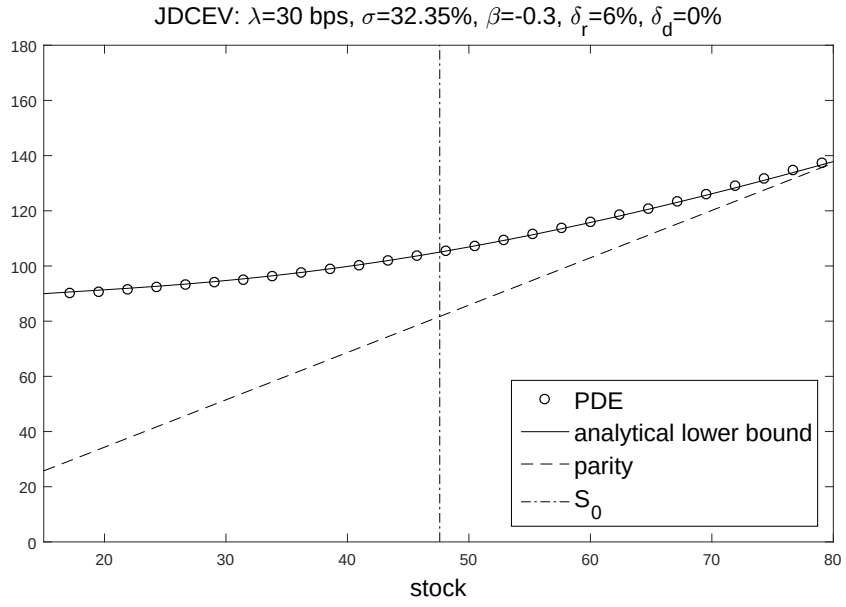

Figure 5.1: Model price and lower bound of the convertible bond without soft call in dependence on the underlying stock price S_0.

The computational gain is huge when computing these price sensitivities: Computing the 41 exact model prices via the PDE methods takes roughly 100 times as long as

computing the corresponding 41 lower bounds[8].
For illustration purposes, we repeat the analysis for the same bond for the case that conversion is always possible. Using the same parameter set as above, the price of the bond is now[9] $B_0 = 108.16\%$. Indeed, we now find that the ECP has a non-negligible value and the lower bound differs from the model price for high values of S_0, see Figure 5.2.

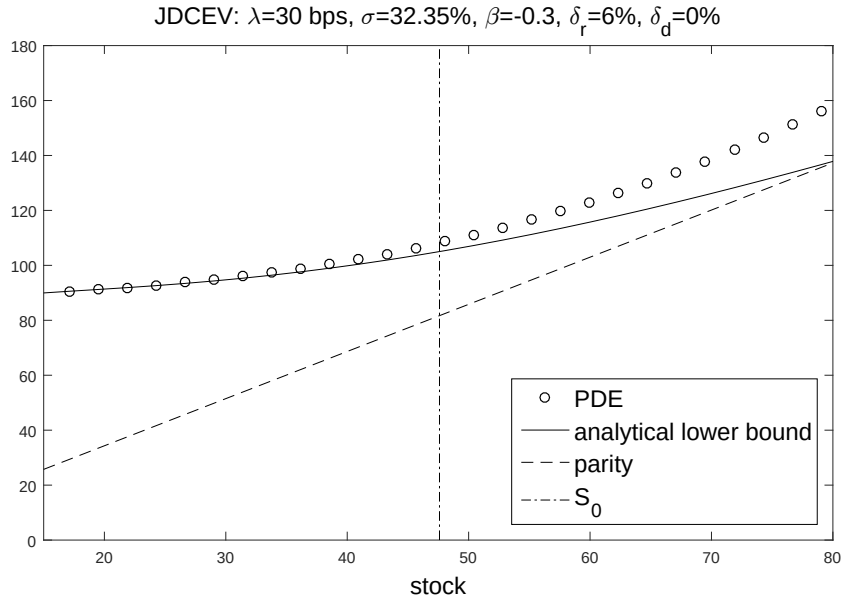

Figure 5.2: Model price and lower bound of the same convertible bond without soft call as in Figure 5.1 in dependence on the underlying stock price S_0, assuming conversion is always allowed.

5.3.2 An example with soft call

We consider a convertible bond with soft call, with a (dirty) price of $B_0 = 87.8782\%$. Again, the bond is denominated in a different currency than the underlying equity, with a current exchange rate of $FX(0) = 0.0722$, and an equity value of $S_0 = 32.30$. The coupon rate is $c = 3.25\%$, the conversion ratio is $\alpha(0) = 0.257023158$, and maturity is $T = 3.5068$. We furthermore assume a recovery rate of $R = 0.20$, lowering the common market assumption of 0.4 to reflect the credit quality of the issuer. The company currently pays no dividends, hence $\delta_d = 0$, and a reasonable assumption on

[8]The code is implemented in Matlab and executed on a standard computer. The PDE prices were computed separately for each initial stock price, to illustrate the computational gain in a situation where one computes prices for different convertible bonds.
[9]The price was determined using the JDCEV model (PDE).

the repo margin parameter is $\delta_r = 1\%$. There is a soft call at trigger level $K = 1.3$, executable from the soft call trigger date $\theta = 2.5315$ on, hence we must resort to the simple credit-equity model with the additional assumption of a constant interest rate. As in the previous example, the interest rates $r_B(t)$ and $r_S(t)$ are bootstrapped from observed market prices for liquidly traded interest rate swaps, and we choose the constant parameter $r_B := \frac{1}{T-\theta} \int_\theta^T r_B(t)\,\mathrm{d}t$ for evaluation of the lower bound. In addition, we evaluate the price of the convertible bond numerically both with the JDCEV model and with the simple credit-equity model, with similar parameters. The parameters are depicted in Figure 5.3 and only differ in order for both models to exactly match the currently observed market price. Again, the chosen parameters are a result of a trader's expertise as well as a calibration to the observed market price. Figure 5.3 visualizes the sensitivities of the model prices, evaluated via the PDE method of Andersen and Buffum (2004), with respect to changes in the underlying stock price, as well as the lower bound computed with help of the formulas in Lemma 5.2.7.

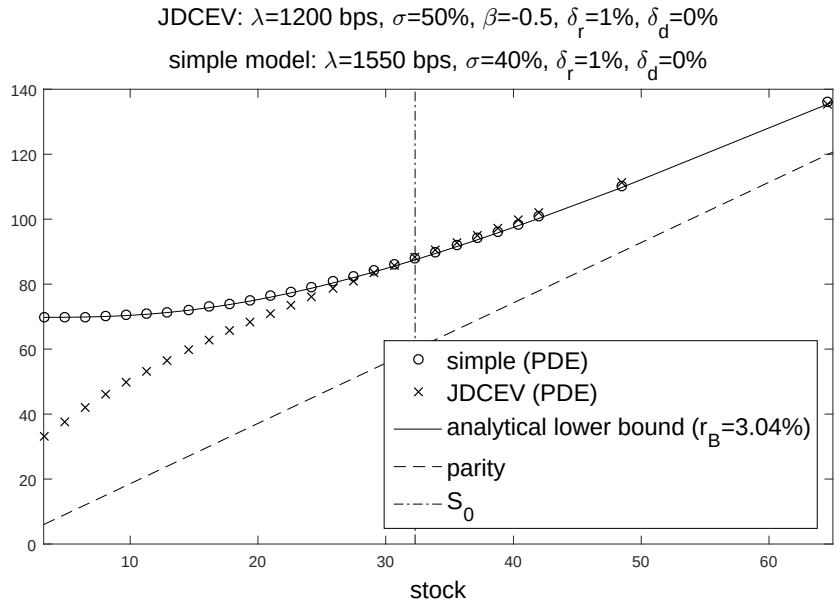

Figure 5.3: Model prices and lower bound of the convertible bond with soft call in dependence on the underlying stock price S_0.

Remark 5.3.1 (On the lower bound in the case with soft call)

1. The difference between the price of the simple model and the JDCEV model for small stock prices stems from the fact that a decline of the share price does not induce an increase of default likelihood in the simple credit-equity model. This is the main weakness of the simple credit-equity model, which necessitates the use of a more advanced model, such as the JDCEV model, for distressed bonds. However, for higher stock prices, in particular near the current stock price, both models imply almost identical prices and price sensitivities, in particular both

models yield almost identical convert deltas.

2. The lower bound now depends on the assumption of a constant interest rate r, while the PDE evaluations do not. This implies a little flaw that is probably negligible in the present example, but in theory could lead to imprecision, e.g. it might theoretically even occur that the lower bound turns out to be larger than the true model price (depending on how r is chosen).

3. Again, the lower bound is very sharp, which means in economic terms that the value of the ECP is practically zero. A possible explanation for this is that the bond coupon of 3.25% outweighs the proceeds from stock lending that are modeled by $\delta_r = 1\%$ (additionally considering the low exchange rate), so that it is not optimal to exercise the conversion right early.

As stated above, in practice a situation in which the ECP is significantly larger than zero is hardly ever met. In order to encounter such a case, one would need a very large parameter δ_r, i.e. the proceeds from stock possession not shared by convertible bond holders, the repo rate and/or the amount of the dividend not passed though to bond holders, should outweigh the proceeds from holding the bond. However, dividends are typically fully passed through to bond holders, making the bond discussed in the previous example highly exceptional, and high repo margins are only typical for highly illiquid, or highly distressed, stock prices, and in these situations it is rather typical that the bond coupon rates are large as well and no dividends are paid.
For the sake of completeness Figure 5.4 depicts the same convertible bond as Figure 5.3, only with two parameters changed: the coupon rate is manually lowered to 0.5%, while the repo margin is increased to $\delta_r = 12\%$, so that the bond price equals now 70.50%[10]. These modifications lead to a highly unrealistic situation, which only suits as an educational example to demonstrate another possible scenario where the derived lower bound to becomes less sharp. Indeed, for high stock prices the ECP now has a non-negligible value.

5.4 Conclusion

Based on the idea of "Europeanizing" the American conversion right in a convertible bond, it was demonstrated how to derive sharp, analytical lower bounds for the bond's model price. In the absence of a soft call covenant, an analytical bound can be derived for defaultable Markov diffusion models allowing for closed-form expressions for survival probabilities and European equity options, as exemplarily shown within the JDCEV model of Carr and Linetsky (2006). In presence of a soft call, an analytical bound is only available within a simple credit-equity model under the assumption of a constant

[10]The bond price was determined using the JDCEV model (PDE). For the simple model to match this price, the parameters had to be changed slightly to $\sigma = 37.24\%$ and $\lambda = 1650$ bps.

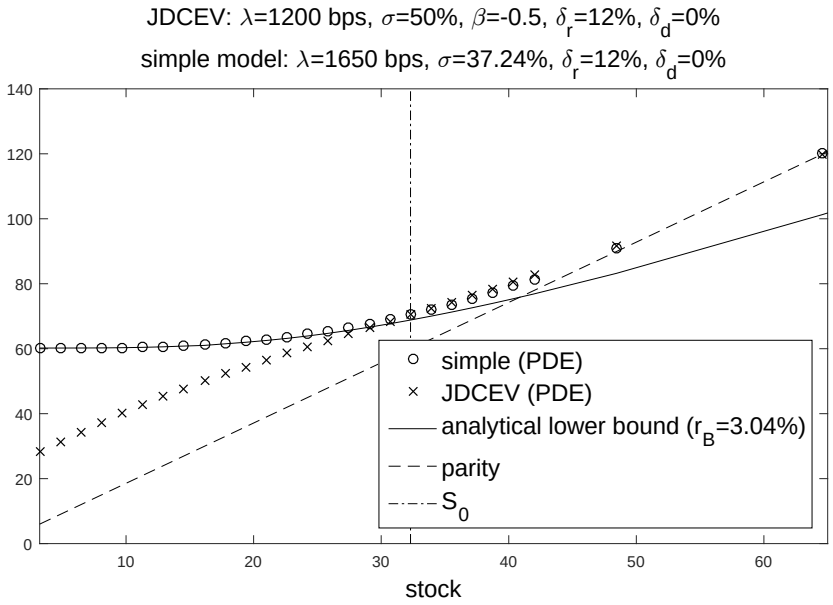

Figure 5.4: Model prices and lower bound of the same convertible bond (with soft call) as in Figure 5.3 in dependence on the underlying stock price S_0, assuming a repo margin of $\delta_r = 12\%$ and coupon rate of $c = 0.5\%$.

risk-free interest rate, which nevertheless performs very well in most cases of practical interest. The sharpness of the analytical bounds depends on the value of the early conversion premium, which itself is an increasing function in the assumed repo margin under the realistic assumption that dividends are passed through to bond holders. This has been demonstrated by means of two examples that were inspired by practical use cases.

6 Conclusion

In this thesis we contributed new results in two areas related to credit risk, namely in the simulation and modeling of correlations and in the pricing of two credit-risky products.

In the first part, we developed a simulation algorithm for Perron–Frobenius correlation matrices, which additionally allows to manipulate the eigenvalue distribution of the generated matrices. It was shown that this algorithm is able to generate all Perron–Frobenius correlation matrices, and from the employed construction principle it was shown that the proportion of such correlation matrices in the set of all correlation matrices is $1/2^{d-1}$ in dimension d. Choosing a realistic eigenvalue structure, large correlation matrices simulated from this algorithm tend to exhibit all observed stylized facts of financial correlation matrices. The algorithm has been used, alongside other algorithms generating correlation matrices with different properties, to show that the empirically observed relation between graph-based portfolio selection techniques and the classical Markowitz ansatz is not due to an inherent connection between the two approaches, but instead originates in the special structure of financial correlation matrices.

We further adapted an approach from geostatistics based on Gaussian random fields for the use with financial data. The field's covariance is modeled via a low-parametric function of the distance between observations, hence allowing for a parsimonious, yet easily extendible modeling of the dependence structure. Special focus was laid on the design of financial distance measures and on the necessary adjustments of the method when used in the high-dimensional coordinate systems typically required for the modeling of financial data sets. Further, we have showcased the benefits of this method for the estimation and parametrization of large financial correlation matrices and for the imputation of missing data in large data sets of financial return series.

In the second part, focused on the pricing of credit-risky products, we developed a new approach for the valuation of CDS options based on an efficient Monte Carlo scheme in a structural model with jumps, which furnishes realistic CDS spread paths. An indication for market prices has been given, as well as a sensitivity analysis of prices with respect to the parameters of the chosen model.
Finally, a sharp lower bound for the price of a convertible bond with and without soft call right was derived in two defaultable Markov diffusion models, and its sharpness was illustrated in two real-world examples.

Bibliography

Aas, K., Czado, C., Frigessi, A. and Bakken, H., Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 2009, **44**, 182–198.

Abrahamsen, P., A review of Gaussian random fields and correlation functions. Technical report 917, Norwegian Computing Center, 1997 `http://publications.nr.no/directdownload/rask/old/917_Rapport.pdf`.

Absil, P.A., Mahony, R. and Sepulchre, R., *Optimization algorithms on matrix manifolds*, 2008 (Princeton University Press: Princeton).

Alexander, C., Principal component models for generating large GARCH covariance matrices. *Economic Notes*, 2002, **31**, 337–359.

Andersen, L. and Buffum, D., Calibration and implementation of convertible bond models. *Journal of Computational Finance*, 2004, **7**, 1–34.

Arbia, G. and Di Marcantonio, M., Forecasting interest rates using geostatistical techniques. *Econometrics*, 2015, **3**, 733–760 Available at `www.mdpi.com/journal/econometrics`.

Arnold, M., Stahlberg, S. and Wied, D., Modelling different kinds of spatial dependence in stock returns. *Empirical Economics*, 2013, **44**, 761–774.

Asgharian, H., Hess, W. and Liu, L., A spatial analysis of international stock market linkages. *Journal of Banking and Finance*, 2013, **37**, 4738–4754.

Ayache, E., Forsyth, P. and Vetzal, K., The valuation of convertible bonds with credit risk. *Journal of Derivatives*, 2003, **11**, 9–29.

Barabási, A.L. and Albert, R., Emergence of scaling in random networks. *Science*, 1999, **286**, 509–512.

Bendel, R. and Mickey, M., Population correlation matrices for sampling experiments. *Communications in Statistics - Simulation and Computation*, 1978, **7**, 163–182.

Benton, D. and Krishnamoorthy, K., Computing discrete mixtures of continuous distributions: noncentral chisquare, noncentral t and the distribution of the square of the sample multiple correlation coefficient. *Computational Statistics and Data Analysis*, 2003, **43**, 249–267.

Bernhart, G. and Mai, J.F., Consistent modeling of discrete cash dividends. *Journal of Derivatives*, 2015, **22**, 9–19.

Bernhart, G. and Mai, J.F., Notizen zur Modellierung von Quanto-Converts. Technical report, XAIA Investment, 2018 Available at `https://www.xaia.com/uploads/media/QuantoConvert_01.pdf`.

Bielecki, T., Jeanblanc, M. and Rutkowski, M., Hedging of a credit default swaption in the CIR default intensity model. *Finance and Stochastics*, 2011a, **15**, 541–572.

Bielecki, T., Crépey, S., Jeanblanc, M. and Rutkowski, M., Arbitrage pricing of defaultable game options with applications to convertible bonds. *Quantitative Finance*, 2008a, **8**, 795–810.

Bielecki, T., Crépey, S., Jeanblanc, M. and Rutkowski, M., Defaultable options in a Markovian intensity model of credit risk. *Mathematical Finance*, 2008b, **18**, 493–518.

Bielecki, T., Crépey, S., Jeanblanc, M. and Rutkowski, M., Defaultable game options in a hazard process model. *Journal of Applied Mathematics and Stochastic Analysis*, 2009, **2009**, 33 pages Article ID 695798.

Bielecki, T., Crépey, S., Jeanblanc, M. and Rutkowski, M., Convertible bonds in a defaultable diffusion model. In *Stochastic Analysis with Financial Applications*, edited by A. Kohatsu-Higa, N. Privault and S. Sheu, pp. 255–298, 2011b (Springer: Hongkong).

Black, F. and Cox, J., Valuing corporate securities: some effects of bond indenture provisions. *Journal of Finance*, 1976, **31**, 351–367.

Blasques, F., Koopman, S., Lucas, A. and Schaumburg, J., Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics*, 2016, **195**, 211–223.

Bonanno, G., Caldarelli, G., Lillo, F. and Mantegna, R., Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 2003, **68**, 046130.

Bouchaud, J.P. and Potters, M., Financial applications of random matrix theory: a short review. In *The Oxford Handbook of Random Matrix Theory*, edited by G. Akeman, J. Baik and D.F. P., pp. 824–850, 2011 (Oxford University Press: Oxford).

Boyle, P., Feng, S., Melkuev, D. and Zhang, J., Correlation matrices with the Perron–Frobenius property. Working paper, available at SSRN: `https://ssrn.com/abstract=2493844`, 2014.

Boyle, P. and N'Diaye, T., Correlation matrices with the Perron–Frobenius property. *Electronic Journal of Linear Algebra*, 2018, **34**, 240–268.

Boyle, P., Feng, S., Melkuev, D., Yang, S. and Zhang, J., Short positions in the first principal component portfolio. *North American Actuarial Journal*, 2018, **22**, 223–251.

Brechmann, E., Hendrich, K. and Czado, C., Conditional copula simulation for systemic risk stress testing. *Insurance: Mathematics and Economics*, 2013, **53**, 722–732.

Brigo, D., Market models for CDS options and callable floaters. *Risk*, 2005, **01/2005**, 89–94.

Brigo, D. and El-Bachir, N., An exact formula for default swaptions' pricing in the SSRJD stochastic intensity model. *Mathematical Finance*, 2010, **20**, 365–382.

Broadie, M. and Detemple, J., Option pricing: valuation models and applications. *Management Science*, 2004, **50**, 1145–1177.

Bun, J., Bouchaud, J.P. and Potters, M., Cleaning large correlation matrices: tools from random matrix theory. *Physics Reports*, 2017, **666**, 1–109.

Burtschell, X., Gregory, J. and Laurent, J.P., A comparative analysis of CDO pricing models under the factor copula framework. *Journal of Derivatives*, 2009, **16**, 9–37.

Carr, P. and Linetsky, V., A jump to default extended CEV model: an application of Bessel processes. *Finance and Stochastics*, 2006, **10**, 303–330.

Carr, P. and Madan, D., Local volatility enhanced by a jump to default. *SIAM Journal on Financial Mathematics*, 2010, **1**, 2–15.

Chen, N. and Kou, S., Credit spreads, optimal capital structure, and implied volatility with endogenous default and jump risk. *Mathematical Finance*, 2009, **19**, 343–378.

Cherry, S., An evaluation of a non-parametric method of estimating semi-variograms of isotropic spatial processes. *Journal of Applied Statistics*, 1996, **23**, 435–449.

Christakos, G., On the problem of permissible covariance and variogram models. *Water Resources Research*, 1984, **20**, 251–265.

Christakos, G. and Papanicolaou, V., Norm-dependent covariance permissibility of weakly homogeneous spatial random fields and its consequences in spatial statistics. *Stochastic Environmental Research and Risk Assessment*, 2000, **14**, 471–478.

Cizeau, P., Potters, M. and Bouchaud, J.P., Correlation structure of extreme stock returns. *Quantitative Finance*, 2001, **1**, 217–222.

Clauset, A., Shalizi, C. and Newman, M., Power-law distributions in empirical data. *SIAM Review*, 2009, **51**, 661–703.

Cont, R. and Kan, Y., Statistical modelling of credit default swap portfolios. Available at: `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1771862`, 2011.

Cressie, N., The origins of kriging. *Mathematical Geology*, 1990, **22**, 239–252.

Cressie, N., *Statistics for spatial data*, 1993, Wiley.

Davies, P. and Higham, N., Numerically stable generation of correlation matrices and their factors. *BIT*, 2000, **40**, 640–651.

De Spiegeleer, J. and Schoutens, W., CoCo bonds with extension risk. *Wilmott Magazine*, 2014, **71**, 78–91.

Delbaen, F. and Shirakawa, H., A note on option pricing for the constant elasticity of variance model. *Asia-Pacific Financial Markets*, 2002, **9**, 85–99.

Detemple, J. and Tian, W., The valuation of American options for a class of diffusion processes. *Management Science*, 2002, **48**, 917–937.

Di Lascio, F., Giannerini, S. and Reale, A., Exploring copulas for the imputation of complex dependent data. *Statistical Methods and Applications*, 2015, **24**, 159–175.

Engle, R., Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 2002, **20**, 339–350.

Erhardt, T., Czado, C. and Schepsmeier, U., R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 2015, **71**, 323–332.

Fan, J., Fan, Y. and Lv, J., High-dimensional covariance matrix estimation using a factor model. *Journal of Econometrics*, 2008, **147**, 186–197.

Fang, K.T., Kotz, S. and Ng, K.W., *Symmetric multivariate and related distributions*, 1990 (Chapman and Hall: London).

Fernandez, V., Spatial linkages in international financial markets. *Quantitative Finance*, 2011, **11**, 237–245.

Fernández-Avilés, G., Montero, J.M. and Orlov, A., Spatial modelling of stock market comovements. *Finance Research Letters*, 2012, **9**, 202–212.

Geidosch, M. and Fischer, M., Application of vine copulas to credit portfolio risk modeling. *Journal of Risk and Financial Management*, 2016, **9**, 4:1–15.

Ghosh, S. and Henderson, S., Behavior of the NORTA method for correlated random vector generation as the dimension increases. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 2003, **13**, 276–294.

Gneiting, T., Making and evaluating point forecasts. *Journal of the American Statistical Association*, 2011, **106**, 746–762.

Gräler, B., Modelling skewed spatial random fields through the spatial vine copula. *Spatial Statistics*, 2014, **10**, 87–102.

Hagan, P. and West, G., Interpolation methods for curve construction. *Applied Mathematical Finance*, 2006, **12**, 89–129.

Hardin, J., Garcia, S. and Golan, D., A method for generating realistic correlation matrices. *The Annals of Applied Statistics*, 2013, **7**, 1733–1762.

Hirschberger, M., Qi, Y. and Steuer, R., Randomly generating portfolio-selection covariance matrices with specified distributional characteristics. *European Journal of Operational Research*, 2007, **177**, 1610–1625.

Holmes, R., On random correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 1991, **12**, 239–272.

Hüttner, A., An introduction to structural credit-equity models. Technical report, XAIA Investment, 2014 Available at `https://www.xaia.com/uploads/media/IntroStructural_01.pdf`.

Hüttner, A. and Mai, J.F., Sharp analytical lower bounds for the price of a convertible bond. *Journal of Derivatives*, 2018, **26**, 7–18.

Hüttner, A. and Mai, J.F., Simulating realistic correlation matrices for financial applications: correlation matrices with the Perron–Frobenius property. *Journal of Statistical Computation and Simulation*, 2019, **89**, 315–336.

Hüttner, A., Mai, J.F. and Mineo, S., Portfolio selection based on graphs: does it align with Markowitz-optimal portfolios?. *Dependence Modeling*, 2018, **6**, 63–87.

Hüttner, A. and Scherer, M., A note on the valuation of CDS options and extension risk in a structural model with jumps. *International Journal of Financial Engineering*, 2016, **3**, 1650011–1–16.

Hüttner, A., Scherer, M. and Gräler, B., Geostatistical modeling of financial data: estimation of large covariance matrices and imputation of missing data. Working paper, Technical University of Munich, Chair of Mathematical Finance, 2019.

Jagannathan, R. and Ma, T., Risk reduction in large portfolios: why imposing the wrong constraints helps. *The Journal of Finance*, 2003, **58**, 1651–1683.

Joe, H., Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 2006, **97**, 2177–2189.

Johnson, C.R. and Tarazaga, P., On matrices with Perron–Frobenius properties and some negative entries. *Positivity*, 2004, **8**, 327–338.

Jönsson, H. and Schoutens, W., Single name credit default swaptions meet single sided jump models. *Review of Derivatives Research*, 2008, **11**, 153–169.

Käärik, E. and Käärik, M., Modeling dropouts by conditional distribution, a copula-based approach. *Journal of Statistical Planning and Inference*, 2009, **139**, 3830–3835.

Käärik, M. and Käärik, E., Imputation by Gaussian copula model with an application to incomplete customer satisfaction data. In *Proceedings of the Proceedings of COMPSTAT 2010*, pp. 485–492, 2010, Springer.

Kaya, H., Eccentricity in asset management. *Journal of Network Theory in Finance*, 2015, **1**, 1–32.

Kazakov, M. and Kalyagin, V., Spectral properties of financial correlation matrices. In *Proceedings of the Models, Algorithms and Technologies for Network Analysis*, edited by V. Kalyagin, P. Koldanov and P. Pardalos, pp. 135–156, 2016, Springer International Publishing.

Keiler, S. and Eder, A., CDS spreads and systemic risk - A spatial econometric approach. Technical report 01/2013, Bundesbank Discussion Paper, 2013 Available at `https://www.bundesbank.de/Redaktion/EN/Downloads/Publications/Discussion_Paper_1/2013/2013_01_28_dkp_01.pdf?__blob=publicationFile`, downloaded 07.12.2015.

Kennedy, D., The term structure of interest rates as a Gaussian random field. *Mathematical Finance*, 1994, **4**, 247–258.

Kennedy, D., Characterizing Gaussian models of the term structure of interest rates. *Mathematical Finance*, 1997, **7**, 107–116.

Kou, S., Peng, X. and Zhong, H., Asset pricing with spatial interaction. *Management Science*, 2016, **64**, 1975–2471.

Kou, S., Petrella, G. and Wang, H., Pricing path-dependent options with jump risk via Laplace transforms. *The Kyoto Economic Review*, 2005, **74**, 1–23.

Kou, S. and Wang, H., First passage times of a jump diffusion process. *Advances of Applied Probability*, 2003, **35**, 504–531.

Kulis, B., Metric learning: a survey. *Foundations and Trends in Machine Learning*, 2012, **5**, 287–364.

Laloux, L., Cizeau, P., Bouchaud, J.P. and Potters, M., Noise dressing of financial correlation matrices. *Physical Review Letters*, 1999, **83**, 1467–1470.

Ledoit, O. and Wolf, M., Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 2003, **10**, 603–621.

Ledoit, O. and Wolf, M., Honey, I shrunk the covariance matrix. *Journal of Portfolio Management*, 2004a, **31**, 110–119.

Ledoit, O. and Wolf, M., A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 2004b, **88**, 365–411.

Lemieux, V., Rahmdel, P.S., Walker, R., Wong, B.L.W. and Flood, M., Clustering techniques and their effect on portfolio formation and risk analysis. In *Proceedings of the Proceedings of the International Workshop on Data Science for Macro-Modeling*, pp. 1–6, 2014, ACM.

Lewandowski, D., Kurowicka, D. and Joe, H., Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 2009, **100**, 1989–2001.

Li, D., Sun, X. and Wang, J., Optimal lot solution to cardinality constrained mean-variance formulation for portfolio selection. *Mathematical Finance*, 2006, **16**, 83–101.

Linetsky, V., Pricing equity derivatives subject to bankruptcy. *Mathematical Finance*, 2006, **16**, 255–282.

Mai, J.F., Pricing single-name CDS options: a review of standard approaches. Technical report, XAIA Investment, 2014 Available at `http://www.xaia.com/uploads/media/XAIACDSOptions.pdf`.

Mai, J.F. and Scherer, M., Lévy-frailty copulas. *Journal of Multivariate Analysis*, 2009, **100**, 1567–1585.

Mantegna, R. and Stanley, H., *An introduction to econophysics*, 2000, Cambridge University Press.

Mantegna, R., Hierarchical structure in financial markets. *European Physics Journal B*, 1999, **11**, 193–197.

Markowitz, H., Portfolio selection. *Journal of Finance*, 1952, **7**, 77–91.

Markowitz, H., *Portfolio selection: efficient diversification of investments*, 1959 (Wiley: New York).

Marsaglia, G. and Olkin, I., Generating correlation matrices. *SIAM Journal on Scientific and Statistical Computing*, 1984, **5**, 470–475.

Martin, R., A CDS option miscellany. Working paper, available at `http://arxiv.org/abs/1201.0111`, 2012.

Matoušek, J. and Nešetřil, J., *Diskrete Mathematik*, 2nd. ed. , 2007 (Springer: Berlin).

Meissner, G. (Ed.) *The definitive guide to CDOs*, 2008, Risk books.

Merton, R., On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 1974, **29**, 449–470.

Metwally, S. and Atiya, A., Using Brownian bridge for fast simulation of jump diffusion processes and barrier options. *Journal of Derivatives*, 2002, **10**, 43–54.

Musiela, M. and Rutkowski, M., *Martingale methods in financial modeling*, 2nd edition , 2009 (Springer: Berlin).

Nelsen, R.B., *An Introduction to Copulas*, 2nd ed. , 2006 (Springer: New York).

Neuberg, R. and Glasserman, P., Estimating a covariance matrix for market risk management and the case of credit default swaps. *Quantitative Finance*, 2019, **19**, 77–92.

Newman, M., The mathematics of networks. In *The New Palgrave Encyclopedia of Economics* (2nd. edn), edited by S. Durlauf and L. Blume, 2008 (Palgrave Macmillan: Basingstoke).

Oh, D.H. and Patton, A.J., Time-varying systemic risk: evidence from a dynamic copula model of CDS spreads. *Journal of Business & Economic Statistics*, 2018, **36**, 181–195.

Onnela, J.P., Chakraborti, A., Kaski, K., Kertész, J. and Kanto, A., Dynamics of market correlations: taxonomy and portfolio analysis. *Physical Review E*, 2003, **68**.

Pebesma, E., The meuse data set: a brief tutorial for the `gstat` R package. First accessed Oct. 2015, latest version Feb. 2019, 2019.

Peralta, G. and Zareei, A., A network approach to portfolio selection. *Journal of Empirical Finance*, 2016, **38**, 157–180.

Perreault, S., Duchesne, T. and Nešlehová, J., Detection of block-exchangeable structure in large-scale correlation matrices. *Journal of Multivariate Analysis*, 2019, **169**, 400–422.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N., Guhr, T. and Stanley, H.E., Random matrix approach to cross correlations in financial data. *Physical Review E*, 2002, **65**.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L.A.N. and Stanley, H.E., Universal and nonuniversal properties of cross correlations in financial time series. *Physical Review Letters*, 1999, **83**, 1471–1474.

Pötzelberger, K. and Wang, L., Boundary crossing probability for Brownian motion. *Journal of Applied Probability*, 2001, **38**, 152–164.

Pozzi, F., Di Matteo, T. and Aste, T., Spread of risk across financial markets: better to invest in the peripheries. *Scientific reports*, 2013, **3**.

Rasmussen, C. and Williams, C., *Gaussian Processes for Machine Learning*, 2006, MIT Press.

RiskMetrics Group, I., CreditMetrics^{TM}technical document. Technical report, RiskMetrics Group, Inc., 2007.

Rockafellar, R.T. and Uryasev, S., Optimization of conditional Value-at-Risk. *Journal of Risk*, 2000, **2**, 21–42.

Roll, R., A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics*, 1977, **4**, 129–176.

Ruf, J. and Scherer, M., Pricing corporate bonds in an arbitrary jump-diffusion model based on an improved Brownian-bridge algorithm. *Journal of Computational Finance*, 2011, **14**, 127–145.

Sato, K., *Lévy processes and infinitely divisible distributions*, 2007, Cambridge University Press.

Schoenberg, I., Metric spaces and completely monotone functions. *Annals of Mathematics*, 1938, **39**, 811–841.

Schönbucher, P., *Credit derivatives pricing models: models, pricing and implementation*, 2003 (Wiley: Chichester).

Schönbucher, P., A measure of survival. *Risk*, 2004, **17**, 79–85.

Steerneman, A. and van Perlo-ten Kleij, F., Spherical distributions: Schoenberg (1938) revisited. *Expositiones Mathematicae*, 2005, **23**, 281–287.

Stewart, G., The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM Journal on Numerical Analysis*, 1980, **17**, 403–409.

Tarazaga, P., Raydan, M. and Hurman, A., Perron–Frobenius theorem for matrices with some negative entries. *Linear Algebra and its Applications*, 2001, **328**, 57–68.

Tobler, W., A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 1970, **46**, 234–240.

Tofallis, C., A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 2015, **66**, 1352–1362.

Tsiveriotis, K. and Fernandes, C., Valuing convertible bonds with credit risk. *Journal of Fixed Income*, 1998, **8**, 95–102.

Vandewalle, N., Brisbois, F. and Tordoir, X., Non-random topology of stock markets. *Quantitative Finance*, 2001, **1**, 372–374.

Wang, L. and Pötzelberger, K., Boundary crossing probability for Brownian motion and general boundaries. *Journal of Applied Probability*, 1997, **34**, 54–65.

Wang, L. and Pötzelberger, K., Crossing probabilities for diffusion processes with piecewise continuous boundaries. *Methodology and Computing in Applied Probability*, 2007, **9**, 21–40.

Weinberger, K. and Saul, L., Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 2009, **10**, 207–244.

Yaglom, A., *Correlation theory of stationary and related random functions I, Basic Results*, 1987, Springer.

Zhou, C., The term structure of credit spreads with jump risk. *Journal of Banking and Finance*, 2001, **25**, 2015–2040.

List of Tables

List of Figures