universität wien

DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Robust Portfolio Optimization and Dimensional Analysis in Finance“

verfasst von / submitted by

Mathias Pohl

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 704 370 136
Dissertationsgebiet lt. Studienblatt / field of study as it appears on the student record sheet:	Statistik und Operations Research
Betreut von / Supervisor:	o. Univ.-Prof. Mag. Dr. Georg Pflug o. Univ.-Prof. Mag. Dr. Walter Schachermayer

Dedicated to the one and only Schneeberg …

ACKNOWLEDGEMENTS

Most of the doctoral theses I came across only had a very short section of acknowledgments. This thesis does not. The following acknowledgements are rather long. The reason is that I started to write the acknowledgments long before I started to write my thesis. In fact, I extended them whenever I met someone who motivated me to work on this thesis or my academic career in general. Moreover, I think that those who are interested in the content of my thesis can skip the next three pages. The others can read the acknowledgments and will still be exhausted afterwards.

Throughout the biggest part of my employment at the University of Vienna (which started in Dezember 2014) I was employed in an interdisciplinary WWTF project. First and foremost I want to convey my gratitude to all involved protagonists: Georg Pflug for selecting me as his PhD-student and being the perfect "Doktorvater" ever since, Walter Schachermayer for becoming my mentor and Nikolaus Hautsch for treating me more like a colleague than a student from the first day onwards. I would like to especially express my deep appreciation to Georg Pflug and Walter Schachermayer, who as my supervisors made this thesis possible by their brilliant guidance over the years in the first place and especially Walter Schachermayer for providing continuous funding throughout my work. To the two PostDocs, Alexander Ristig and Ludovic Tangpi, who shared an office with me for three years and were employed in the same project: It is incredible how much I learned from the two of you over the years. Thank you very much. Let me quote our motto a final time: "Discussion in english, jokes in german".

I was so lucky to get to know the academic community as a collection of people who want to help and support each other in order to serve science - and thereby society - in the best possible way. My aim to become a part of this community was probability the biggest motivation to work on this thesis. Obviously, I cannot thank everyone here, but I will still try. Thanks to: Ruduo Wang for being the first outside of the University of Vienna to give me valuable feedback. Wolfgang Trutschnig for his interest and feedback and for checking

with Fabrizio Durante if someone already worked on the relation between measures of concordance and the Wasserstein distance. Giovanni Puccetti for a long discussion in Almeria which helped me a lot. Daniel Kuhn for a great discussion and an excellent lecture series. Dojre Brody for offering me a PhD position at Brunel University which in fact led to the idea that I could do a PhD. Andrei Kirilenko for a brilliant discussion about what I called the clock-time periodicity of stock markets and Frank Hatheway for telling me a year later that this phenomenon is known as "the plus". Mathieu Rosenbaum for answering my questions and giving me new ones. Anna Obizhaeva and Pete Kyle for inspiration and desperation. Michael Kupper and Stephan Eckstein for what remains the best week during my PhD studies, and Daniel Bartl for cracking some of the best jokes. Friedrich Hubalek for showing me that $1/(2\pi)$ does not equal $1/6$ (which means $\pi \neq 3$). Daniel Lacker for an applied optimal transport seminar on a Friday afternoon at Merkenstein. Rama Cont for his magnificent metaphor involving the Danube and a swimmer. Johannes Muhle-Karbe for his useful advise after a presentations of mine. Sebastian Fuchs for helping me to fall in love with the copula community. Those who where not yet mentioned but made the following conferences to especially wonderful events: the winter school on energy systems and markets in March 2015, the conference on dependence modeling in finance, insurance and environmental science in May 2016, the Vienna congress on mathematical finance in September 2016, the workshop of high frequency trading in September 2016, the conference on copulas and their applications in July 2017, and the conference on model uncertainty and robust finance in March 2018.

Before coming to the academic staff at the University of Vienna, I want to thank the non-academic staff members from the Institute for Statistic and Operations Research as well as the Faculty for Mathematics - especially Dominique Sundt, not only for baking many cakes, and Astrid Kollros for many things but mostly for holding the financial mathematics group together. Now, let me come to my fellow PhD Students: Martin Glanzer helped me a lot by reading my first paper within a day and correcting a lot of typos and formal mistakes. Corina Birghila, my dear office partner, was always there to discuss and help me with a lot of small steps. So did Sandor Guzmics and Daniela Escobar. Together with Caroline Geiersbach, the latter mentioned folks were my „companions in misfortune".

To Georg Brandstätter, Nina Senitschnig, Lukas Steinberger and Ivana Milovic: be aware that I appreciate your help a lot. Thanks Gökhan Cebiroglu for entrusting me with an idea which we thought would change the world and in the end did not. Moreover, I want to thank Hannes Leeb in particular for his positiv feedback on my thesis proposal and Irene Klein as well as Immanuel Bomze for some crucial advises and support. Kory Johnson helped me with quite a few things that where easy for him but hard for me (not only the slab pitches

on the Stadelwand). Coming to the group of financial mathematics, due to the brown bag lunch it often felt like the weekend would start on Friday noon. Out of all the amazing people who joined to group over the years, I especially want to thank Rémi Peyre for finishing the tiramisu, Julio Daniel Backhoff Veraguas for putting a smile on my face whenever I see him, Anastasiia Zalashko for not only being extremely kind but also for being my optimal transport lexicon, Christa Cuchiero for helping me with kind words when I nearly turned crazy as well as Jiří Černý and Mathias Beiglböck for caring about all of us.

To keep up the motivation, the Tuesday Boulderbar connection became very important to me. As all of the participants have a connection to mathematics they really deserve to be mentioned here: Birgit, who simply is awesome. Mathias, who truly deserves to be mentioned twice for numerous reasons and certainly does not deserve to be described by a single adjective. Fritz, who became one of my best friends with the same speed in which he is climbing. Annemarie, who calms herself down by brushing holds in the crux sections. Martin, who has an Ape index from another species. Katharina, who knew that the grades do not matter long before I did. Dominik, who is the only one who ever reached something that I could not.

I gratefully acknowledge the financial support from the European Research Council (ERC) under grand FA506041, the Austrian Science Fund (FWF) under grand P25815 and P28661 and the Vienna Science and Technologie Fund (WWTF) through project MA14-008.

The basis for successfully completing the doctoral studies is to stay mentally and physically healthy - which is easier said than done. Hence, I want to thank all my dear friends (without naming them all) and stress that they helped me to keep a good work-life balance. Let me mention my buddy Adam Ondra nevertheless, who kept motivating me by showing that the impossible is possible. Moreover, I want to acknowledge the support of my wonderful family, i.e., my 92 year old grandfather, my awesome sister Michaela and my loving parents.

Finally, my deep gratitude goes to my girlfriend Leni, who always cares about me and loves me unconditionally. This dissertation would not have been possible without her love, faith and patience.

TABLE OF CONTENTS

PREFACE

Even for risk averse investors, it can be optimal to invest in only a single asset. This rather surprising result is proved in the first part of the thesis. It only requires that investors take into account uncertainty in the model for the dependence structure between assets. More generally, the first chapter of this thesis is concerned with robust portfolio optimization. The second and the third chapter deal with dimensional analysis in the context of financial markets. This topic is related to the first only by the obvious fact that both are part of the vast research area of finance.

Each of the chapters was developed and submitted as a separate paper. In this preface, I provide a detailed overview of the individual parts and relate them whenever possible. Before doing so, let me shortly describe the main idea behind the two topics.

In the classical portfolio selection problem, one selects the assets in a portfolio in order to maximize the expected portfolio return while minimizing the risk of the portfolio. The return distribution of the portfolio's assets is needed to evaluate both criteria. In practice however, the distribution of future returns is not known, but is estimated from data. Robust portfolio optimization protects against the effects of model ambiguity in this estimated reference distribution.

Chapter 1 reviews a recently established approach in this context. We take as given both a joint distribution of the returns and a quantification of the investor's trust in this reference distribution. The portfolio is optimized under the worst case distribution which is "close" to the reference distribution. Our measure of "closeness" is strongly related to the amount of trust the investor has in the reference model. It has been shown that as the investor's trust decreases, the optimal portfolio becomes more and more diversified. In the limit, when no distributional information on the future returns is available, the equally weighted portfolio investment strategy is optimal.

A conceptually different, yet mathematically similar approach is portfolio optimization under *dependence uncertainty*. The problem set-up is the same as above, but we additionally

assume that the marginal distributions of asset returns are know. Hence, the model uncertainty lies solely in the dependence structure between the assets. In many cases of practical interest, this so-called dependence uncertainty is a reasonable assumption. One might be able to determine the marginal distributions of a portfolio's individual assets but fail to estimate their dependence structure accurately. Focusing the analysis solely on dependence uncertainty yields novel results. In particular, it turns out that under high dependence uncertainty it is optimal to concentrate a portfolio into a single asset. Hence, we contrast these two approaches and their contradictory implications: diversification and concentration.

Chapters 2 and 3 explain the use of dimensional analysis in the context of financial markets. Dimensional analysis is a well-known line of argument in physics. It builds on the intuition that the validity of a relation between given quantities should not depend on the units which are used to measure the related dimensions. In physics, relevant dimensions are for instance length, time and mass, whereas their units of measurement are, e.g., meters, seconds, and grams. When applying this argument in a financial context, the relevant dimensions are typically time, shares, and money. Therefore, the validness of relations in finance should not be affected when changing the units of measurement from seconds to minutes, single shares to packages of hundred share and Euros to Euro-cents, respectively.

For the considered applications, the principle of dimensional analysis can be explained as follows: Suppose a variable of interest U depends on n explanatory variables $W_1, \dots, W_n$. Moreover, assume that there are the same number of relevant dimensions as there are explanatory variables, namely n. Then, under certain conditions, dimensional analysis allows us to determine the unique functional relation between U and the variables $W_1, \dots, W_n$, which respects the considered dimensions.

In Chapter 2, the variable of interest is the market impact of a meta-order,[1] which is defined as the size of the price movement caused by the execution of this meta-order. By identifying a certain set of explanatory variables and imposing an additional assumption called "leverage neutrality", we derive in particular a square-root dependence of the market impact on the size of the meta-order. This agrees with several other theoretical and empirical findings. Hence, we complement the existing literature with a rigorous derivation of this so-called square-root law for market impact based on a set of well defined assumptions.

We continue to foster our understanding of dimensional analysis when applied in the area of finance in Chapter 3. There, the variable of interest is the trading activity, which we define as the number of trades in a given interval. By considering various sets of explanatory

[1] A meta-order, also referred to as *bet*, is a collection of trades originating from the same trading decision of a single investor.

variables we derive relationships that have already been proposed by other authors. Whether the derived relations describe reality is however an empirical question. Indeed our results hinge on the assumption that the chosen explanatory variables indeed *fully* explain the variables of interest - in our case the number of trades. Hence, we perform an empirical analysis to determine the fit of the considered relations to data from the stock market.

Overview

Chapter 1: A review on ambiguity in stochastic portfolio optimization

In mean-risk portfolio optimization, it is typically assumed that the returns of the assets follow a known distribution P_0, which is estimated from observed data. We aim at an investment strategy which is robust against possible misspecification of P_0. Hence, the portfolio selection problem is considered with respect to the worst-case distribution within a Wasserstein-neighborhood of P_0. The investor's trust in the *reference distribution* P_0 is captured by the size of this Wasserstein-neighborhood. After motivating this approach in the introductory Section 1.1, Section 1.2 offers a general perspective on stochastic optimization under model ambiguity. In Section 1.3, we come back to the portfolio selection problem under model ambiguity mentioned above and review tractable formulations thereof by Pflug and Wozabal (2007), Pflug et al. (2012) as well as by Esfahani and Kuhn (2015). For instance, the latter two references established that high model ambiguity leads to equally-weighted portfolio diversification.

In Section 1.4, we introduce the portfolio selection problem under dependence uncertainty. This novel viewpoint is motivated by a situation in which the marginal return distributions of the assets can be estimated with high accuracy, whereas the dependence structure between the assets remains ambiguous. We therefore fix the marginal distributions and assume the model uncertainty only concerns the dependence structure. It is thus natural to define the Wasserstein-neighborhood, in which the worst case distribution has to lie, directly on the level of the dependence structure. We show that if the investment decision is based on the worst case among all kinds of dependencies, which corresponds to the case where the investor does not trust the reference dependence structure at all, then concentration of the portfolio into a single asset is optimal. Moreover, we develop an algorithm which allows us to approximately solve the portfolio selection problem for a fixed level of dependence uncertainty.

The empirical study in Section 1.5 compares the two approaches and their opposing implications.

Chapter 2: The amazing power of dimensional analysis: Quantifying market impact

We start to illustrate the amazing power of dimensional analysis by considering the classical example of the period of a pendulum. Inspired by the work of Kyle and Obizhaeva (2017a), we then apply the concept of dimensional analysis to determine the price impact of a meta-order and derive the following remarkable fact. If the market impact G only depends on the volatility σ, the asset price P, the traded volume V and the size of the meta-order Q while the assumption of leverage neutrality is also satisfied, then there is only one possible form of this dependence: $G \sim \sigma\sqrt{Q/V}$. This relation is known as the square-root law for market impact and its simple derivation based on dimensional analysis is the main contribution of this chapter. For this purpose the assumption of leverage neutrality, which was first considered in this context by Kyle and Obizhaeva (2017a), is crucial: Building on the findings of Modigliani and Miller (1958), we establish how the given quantities behave when the stock issueing company changes its capital structure by paying dividends or by raising new capital. We show that conceptually this restriction can be understood as an artificial "dimension" in our analysis.

The square root law can be regarded as a special case of a more general result of Kyle and Obizhaeva (2017a). Their result is based on an additional fifth explanatory variable. Hence, the previous unique relation is replaced by a richer variety of possible functional relations. Finally, we study another extension of the original set of our explanatory variables by including the length T of the execution interval of the considered meta-order.

Chapter 3: Theoretical and empirical analysis of trading activity

Chapter 3 builds on the ideas introduced in Chapter 2, i.e., dimensional analysis and leverage neutrality, but applies them in a different context, namely to study trading activity. We aim to understand how the trading activity, defined here as the number of trades N, relates to other financial quantities such as the volatility σ, the traded volume V, the assets price P or the bid-ask spread S as well as the trading cost C. In this regard, different proportionality relations have been proposed: the very basic connection that the number of trades is proportional to the squared volatility, i.e., $N \sim \sigma^2$ (see Jones et al., 1994), the so-called 3/2-law $N^{3/2} \sim \sigma PV/C$ (see Benzaquen et al., 2016) and the relation $N \sim (\sigma P/S)^2$ (see Wyart et al., 2008). In Section 3.2, we prove that these "laws" are the only possible relations for the considered sets of explanatory variables by means of dimensional analysis. Here, the latter two relations $N^{3/2} \sim \sigma PV/C$ and $N \sim (\sigma P/S)^2$ are estabished under the additional assumption of leverage neutrality. However, if we consider the product σP rather than the explanatory variables σ

and P separately, the complexity of the problem is reduced and the assumption of leverage neutrality is no longer required.

Section 3.3 is devoted to the question which of the discussed relations is indeed supported by data. It is important to complement our theoretical analysis with an empirical study since dimensional analysis does not provide any guarantees as to which relation actually represents reality. By its very nature, dimensional analysis only establishes which relations are consistent with the considered scaling invariances, given that a fixed set of variables *fully* explains the variable of interest. Hence, the art of the game is to choose the right set of explanatory variables. In the context of trading activity, our empirical analysis suggests that information concerning the spread, reflected either by the bid-ask spread S or by the trading cost C, is crucial to explain the number of trades N. Based on data from the NASDAQ stock exchange, Section 3.3 shows that the two relations $N^{3/2} \sim \sigma PV/C$ and $N \sim (\sigma P/S)^2$ hold with a certain degree of universality.

In Section 3.4, we finally discuss the time scaling of the volatility σ, which turns out to be more subtle than expected.

Individual contribution of co-authors

The text in Chapter 1 is taken from a joint paper with Georg Pflug which was published in the Journal of *Set-Valued and Variational Analysis*, see Pflug and Pohl (2017). Georg Pflug suggested the portfolio selection problem under dependence uncertainty as the topic for my PhD studies. Hence, he supervised my work throughout my studies while I sign responsible for the main workload. Concerning the writing of the paper, I prepared the initial draft of all sections, except the introductory Sections 1.1 and 1.2.

Chapter 2 coincides with the paper "The amazing power of dimensional analysis: Quantifying market impact" which is accepted for publication in the Journal of *Market Microstructure and Liquidity* and was written by Alexander Ristig, Walter Schachermayer, Ludovic Tangpi and myself, see Pohl et al. (2017). The initial idea was provided by Walter Schachermayer. I contributed by establishing the link to the Pi-Theorem, which is well known in physics, but was unknown to us. The paper then emerged from numerous discussions between the four of us. The writing of the paper was equally split among all co-authors: Walter Schachermayer drafted the introduction, i.e., Section 2.1, Alexander Ristig wrote the literature review in Section 2.2, Ludovic Tangpi and I provided the first draft of Section 2.3 and Section 2.4. It should be stressed that due to extensive discussions and subsequent changes, we all share equal reasonability not only for the whole paper, but also for each individual part.

Together with the same group of coauthors I also wrote the paper “Theoretical and empirical analysis of trading activity”. This paper is presented in Chapter 3 and has been submitted, see Pohl et al. (2018). Note that the text in Chapter 3 does not correspond exactly to this paper since I referenced the results in Chapter 2 directly rather than collecting them in an additional appendix like we did in the paper. The paper developed out of many discussions following our first above mentioned joint publication. During these discussions it became clear that this paper should form the final part of my thesis. Hence, I am responsible for the first written draft. My initial draft was then revised and improved by the three co-authors. Alexander Ristig helped a lot with the empirics and generated all the plots.

CHAPTER 1

A REVIEW ON AMBIGUITY IN STOCHASTIC PORTFOLIO OPTIMIZATION

JOINT WORK WITH G. CH. PFLUG

ABSTRACT.

In mean-risk portfolio optimization, it is typically assumed that the assets follow a known distribution P_0, which is estimated from observed data. Aiming at an investment strategy which is robust against possible misspecification of P_0, the portfolio selection problem is solved with respect to the worst-case distribution within a Wasserstein-neighborhood of P_0. We review tractable formulations of the portfolio selection problem under model ambiguity, as it is called in the literature. For instance, it is known that high model ambiguity leads to equally-weighted portfolio diversification. However, it often happens that the marginal distributions of the assets can be estimated with high accuracy, whereas the dependence structure between the assets remains ambiguous. This leads to the problem of portfolio selection under dependence uncertainty. We show that in this case portfolio concentration becomes optimal as the uncertainty with respect to the estimated dependence structure increases. Hence, distributionally robust portfolio optimization can have two very distinct implications: Diversification on the one hand and concentration on the other hand.

1.1 Introduction

The history of optimization is characterized by weakening the assumptions about the available information. Back in the old days, the paradigm was that both the objective function and the constraints are well known, deterministic functions. This strict assumption has been weakened even in the early work of Dantzig (1955) by allowing some parameters to be unknown, but assuming that they are realizations of known distributions and thus random. Other pioneering works in the area of stochastic programming include Charnes and Cooper (1959), Tintner (1941) and Wets (1966). More recently, the assumption of known distributions has been replaced by the weaker assumption that only a set of distributions is known which includes the unknown *true* distribution. We call this set of distributions *ambiguity set*. Optimizing with respect to the *worst case* distribution, i.e., the distribution contained in the ambiguity set which yields the worst value of the objective function, leads to distributionally robust decisions.

In this context, the portfolio selection problem is arguably one of the most prominent problems. Introduced by Markowitz (1952) in the early 1950s, more than 25000[1] citations of his celebrated work speak for themselves. We refer to Brandt (2009) and Steinbach (2001) for surveys listing some of the most important contributions in the vast field of portfolio selection.

Adapting the ideas mentioned above, this paper discusses portfolio selection strategies which are robust with respect to misspecification of the assets' distribution. In other words, we account for model uncertainty in the portfolio selection problem. Hence, our approach falls into the field of robust portfolio optimization. Two perceptive reviews of the developments in this area were written by Fabozzi et al. (2010) and Kim et al. (2014). The present paper differs from most of the existing literature in that we rely on a purely non-parametric approach which builds on the Wasserstein distance to construct the ambiguity set. Within this setting, we contrast two fundamentally different ways to address the problem: Firstly, we consider the case of model uncertainty with respect to the model for the joint distribution of the portfolio's assets, which we will refer to as *model ambiguity*. Secondly, we assume that the marginal distribution of the assets are known and the model uncertainty lies solely on the level of the dependence structure. The latter case is known as *dependence uncertainty*. The gist of our analysis is that when increasing the level of model uncertainty, portfolio diversification becomes optimal, while in the latter case of dependence uncertainty portfolio concentration becomes optimal.

[1] according to Google Scholar.

The remainder of this paper is structured as follows. Section 1.2 introduces ambiguity in stochastic optimization in a general context. The portfolio selection problem under model ambiguity is studied in section 1.3, whereas section 1.4 looks at dependence uncertainty in the context of portfolio selection. The empirical study in section 1.5 compares the two presented approaches. Section 1.6 concludes.

1.2 Ambiguity sets

The concept that probability models are given by ambiguity sets is quite popular in statistics as there is a plethora of parametric and semi-parametric models, which are based on confidence regions of parameters. The basic construction for parametric models, i.e., probability models of the form $(P_\theta, \theta \in \Theta)$, allows us to define an ambiguity set as a subset $(P_\theta, \theta \in \hat{\Theta})$, which is determined by a parametric estimate $\hat{\theta}$ of θ and some confidence region $\hat{\Theta}$ around it. Typical examples include:

- The scenario distribution stems from a normal distribution $N(\mu, \sigma^2)$, where μ and σ are only known to lie within given bounds, i.e., $\mu_1 \leq \mu \leq \mu_2$ and $\sigma_1^2 \leq \sigma^2 \leq \sigma_2^2$.
- The multivariate distributions possess first and second moments, where the mean vector μ satisfies $\mu_1 \leq \mu \leq \mu_2$ (componentwise) and the covariance matrix Σ satisfies $\Sigma_1 \preceq \Sigma \preceq \Sigma_2$ (in the sense of positive semidefinite ordering).

When it comes to semi-parametric and fully nonparametric ambiguity models, we have seen different approaches emerge in recent years. For instance, ambiguity set containing all distributions with given constraints on the moments are studied e.g. in Delage and Ye (2010); Edirisinghe (2011); Liu et al. (2017). This approach can be extended to ambiguity sets which are characterized, additionally to the moment bounds, by constraints on structural properties auch as symmetry, unimodality or independence patterns, see Hanasusanto et al. (2015). Another field of research deals with sample average approximations, where ambiguity sets appear as confidence regions (see e.g. Bertsimas et al., 2014; Meskarian et al., 2012; Xu and Zhang, 2012). An overview of these and related topics can be found in the book by Consigli et al. (2017).

In contrast to the cited concepts above, this paper relies on a very intuitive approach, where ambiguity sets are given by some baseline model P_0 and a ball around P_0 with respect to some (pseudo-) distance for probability measures. Examples for such distances are:

- the bounded Lipschitz distance

$$d_{BL}(P,P_0) = \sup\left\{\left|\int h(x)\,dP(x) - \int h(x)\,dP_0(x)\right| : |h(x)-h(y)| \le d(x,y); |h(x)| \le 1\right\};$$

- the Wasserstein distance (here of order 1)[2]

$$d_W(P,P_0) = \sup\left\{\left|\int h(x)\,dP(x) - \int h(x)\,dP_0(x)\right| : |h(x)-h(y)| \le d(x,y)\right\}; \tag{1.1}$$

- the Fortet-Mourier distance

$$d_{FM}(P,P_0) = \sup\left\{\left|\int h(x)\,dP(x) - \int h(x)\,dP_0(x)\right| : |h(x)-h(y)| \le L(x,y)\cdot d(x,y)\right\},$$

  where $L(x,y) = \max(\|x\|^r, \|y\|^r, 1)$;

- the relative entropy (Kullback-Leibler pseudo-distance)

$$d_{KL}(P,P_0) = \int \log\left(\frac{dP}{dP_0}\right)(x)\,dP_0(x);$$

- the Vapnik-Cervonenkis distance

$$d_{VC}(P,P_0) = \sup\{P(C) - P_0(C) :\ C \in \mathscr{K}\},$$

  where $\mathscr{K}$ is a Vapnik-Cervonenkis class, see Shorack and Wellner (2009).

More details on the ideas behind the above definitions can be found in Pflug and Pichler (2014). The following facts are well known: The bounded Lipschitz distance metricizes the weak topology of probability measures and therefore the empirical measure converges to the true one a.s. in this distance. If the true measure has a finite first moment, then a.s. convergence also holds with respect to the Wasserstein distance. Existence of higher moments is required for the convergence in Fortet-Mourier distance. It was shown in Pflug

[2]Wasserstein distance of order 1 is sometimes referred to as Kantorovich distance.

and Pichler (2011) that the Fortet-Mourier distance is a special case of the Wasserstein distance with a different metric on the real line. In contrast, the relative entropy and the VC distance generate stronger topologies implying that the empirical measure does not converge to the true one in these topologies. For this reason, these distances are excluded from our considerations. If a first moment exists, then the a.s. convergence of the empirical measure to the true measure in Wasserstein distance is an easy consequence of the law of large numbers and the dominated convergence theorem. However, rates of convergence have only been established recently, see Fournier and Guilin (2015).

While the relative entropy as a concept for closedness of models was introduced by Hansen and Sargent (2001), Wasserstein neighborhoods were considered for the first time by Pflug and Wozabal (2007) and recently followed by Esfahani and Kuhn (2015) as well as Gao and Kleywegt (2016) and Ji and Lejeune (2015). In the remainder of this paper we restrict our considerations to ambiguity sets defined by Wasserstein neighborhoods. In particular, we set the distance $d(x,y) = ||x-y||_1 := \sum_{i=1}^m |x_i - y_i|$ in (1.1). This choice is motivated by the fact that the expected return as well as the Average-Value-at-Risk functional, defined below, are Lipschitz continuous with respect to this distance. This property will prove to be useful when we discuss the portfolio selection problem below.

In all the approaches cited above, the model uncertainty aspect in decision making is taken into account by the following concept: While for a single model P one considers the standard stochastic program

$$\min_x \{\mathbb{E}_P[Q(x,\xi)] : x \in \mathbb{X}\},$$

where Q is a cost function depending on the decision x and the random part ξ, in case of an ambiguity set $\mathscr{P}$ one solves the minimax stochastic program

$$\min_x \max_{P\in\mathscr{P}} \{\mathbb{E}_P[Q(x,\xi)] : x \in \mathbb{X}\}. \tag{1.2}$$

Such a problem formulation is called risk-neutral. As an extension, we might also look at the risk-averse formulation

$$\min_x \max_{P\in\mathscr{P}} \{\mathscr{R}_P[Q(x,\xi)] : x \in \mathbb{X}\}. \tag{1.3}$$

where $\mathscr{R}$ is some risk-functional.

The existence of a solution of (1.3) is related to the existence of a saddle point of $f(x,P) : (x,P) \mapsto \mathscr{R}_P[Q(x,\cdot)]$. Recall that the pair (x^*,P^*) is called a saddle point for f, if for any $x \in \mathbb{X}$ and $P \in \mathscr{P}$

$$f(x^*,P) \leq f(x^*,P^*) \leq f(x,P^*).$$

If (x^*,P^*) is a saddle point, then x^* is the minimizer of $x \mapsto \max_P f(x,P)$ and of $P^* \in \operatorname{argmax}_P f(x^*,P)$. The converse is not true: by searching for all saddle points, one in general does not obtain all possible solutions of the minimax problem. In order to characterize all solutions of the minimax problem as saddle points, one has to impose stronger global conditions, such as quasi-convexity and quasi-concavity: A function $(x,P) \mapsto f(x,P)$ is quasi convex-concave if $\{x : f(x,P) \leq c\}$ is convex or empty for all $P \in \mathscr{P}$ and all $c \in \mathbb{R}$ and at the same time $\{P : f(x,P) \geq c\}$ is convex for all $x \in \mathbb{X}$ and all $c \in \mathbb{R}$. For such functions, Sion's famous saddle point Theorem holds, which is given in appendix A.2.

1.3 Portfolio selection under model ambiguity

We start this section by reviewing the classical mean-risk framework, which goes back to Markowitz (1952) and his mean-variance approach. In the literature, alternatives to Markowitz' choice of the variance as a risk functional have been proposed. For example, the semi-variance (Markowitz, 1959), the mean absolute deviation (Konno and Yamazaki, 1991), the Value-at-Risk (Basak and Shapiro, 2001), the Average-Value-at-Risk (Rockafellar and Uryasev, 2000) or, more generally, convex risk measures (Föllmer and Schied, 2002) as well as coherent risk measures (Artzner et al., 1999) were proposed.

In the remainder of this paper, let the random vector $\xi = (\xi_1,\ldots,\xi_m)^\top$describe the random return of the m risky assets in the capital market under consideration. Let the decision variable $x = (x_1,\ldots,x_m)^\top \in \mathbb{X} := \left\{x \in \mathbb{R}^m_+ : \sum_{j=1}^m x_j = 1\right\}$ denote the proportion of the portfolio to be invested in the m risky assets. Notice that we do not allow for short-selling. Moreover, assume that $\lambda \in \mathbb{R}_+$ quantifies the investor's risk-aversion and $\mathscr{R}$ is her preferred risk functional. The portfolio selection problem can then be formulated as the following single-stage stochastic program

$$\max_{x\in\mathbb{X}} \mathbb{E}\left(x^\top \xi^P\right) - \lambda \mathscr{R}\left(-x^\top \xi^P\right), \tag{1.4}$$

where P refers to the known probability distribution of the random vector ξ representing the asset returns. To put problem (1.4) into words, we aim to maximize the risk adjusted expected return of the portfolio, where the risk adjustment is controlled by the risk aversion coefficient λ as well as the risk functional $\mathscr{R}$. Together λ and $\mathscr{R}$ are assumed to capture the investor's risk preferences. Notice that for a fixed probability space (Ω,σ,μ) the risk functional $\mathscr{R} : (\Omega,\sigma,\mu) \to \mathbb{R}$ assigns a real value to the random variable X, which represent

the random future losses. To be more precise, $\mathscr{R}(X)$ quantifies the riskiness of X which means high values of $\mathscr{R}$ indicate more risk and hence a less desirable situation.

We choose the Average-Value-at-Risk as a risk functional throughout most of our analysis:

$$\mathrm{AV@R}_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 F_X^{-1}(t)\,dt,$$

where F_X^{-1} denotes the inverse distribution function of the random variable X. Notice that this definition implies that values of α are typically chosen close to 1 as we consider the upper tail of the loss distribution. As shown in Rockafellar and Uryasev (2000), the AV@R allows to reformulate problem (1.4) as a linear program in case P is a discrete probability distribution.

In practice, the distribution P of the future returns is typically not known. However, statistical methods allow us to come up with an - hopefully accurate - estimate P_0 of the true, unknown probability distribution P. Therefore, it seems reasonable to assume that the decision maker takes the available information in form of P_0 into account, while incorporating the ambiguity with respect to P_0 into her decision. We model this situation by defining the ambiguity set as the set of distributions, whose Wasserstein distance to the so-called *reference distribution* P_0 does not exceed a certain threshold κ. That is, we define the ambiguity set as

$$\mathscr{B}_\kappa(P_0) = \{Q \in \mathscr{P}(\mathbb{R}^m) : d_W(P_0, Q) \leq \kappa\},$$

where $\mathscr{P}(\mathbb{R}^m)$ denotes the space of all Borel probability measures on $\mathbb{R}^m$. We can now formulate the distributionally robust counterpart of the stochastic program (1.4) as the following maximin problem

$$\max_{x\in\mathbb{X}} \min_{Q\in\mathscr{B}_\kappa(P_0)} \mathbb{E}\left(x^\top \xi^Q\right) - \lambda\,\mathrm{AV@R}_\alpha\left(-x^\top \xi^Q\right). \tag{1.5}$$

As stated in Wozabal (2012), problem (1.5) is algorithmically intractable for the following three reasons: Firstly, it contains infinitely many constraints. Secondly, the solutions of the maximin problem (1.5) are elements of an infinite dimensional space. Lastly, solving for a saddle point of the objective function is usually harder than simply finding a maximum or a minimum. Nevertheless, three distinct approaches to solve problem (1.5) overcome the mentioned difficulties: Pflug and Wozabal (2007), who first considered the problem, restricted their analysis to all the discrete measures with atoms identical to those of P_0, which is assumed to be the empirical measure $\hat{P}_n$ constructed from n observations of the random vector ξ. They then proposed a successive convex programming solution method to solve the

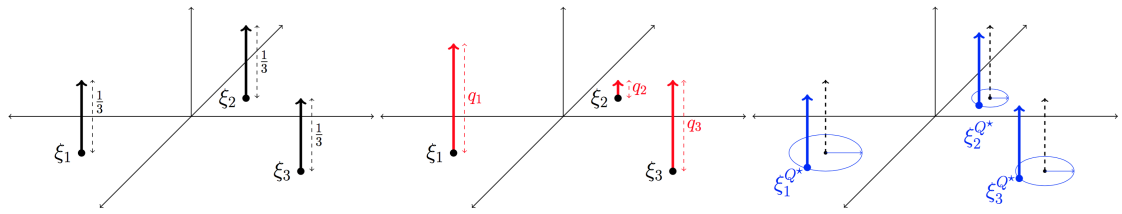


Fig. 1.1 The left panel illustrates an empirical measure $\hat{P}_3$ in $\mathbb{R}^2$ with support points (ξ_1, ξ_2, ξ_3). The middle panel illustrates how Pflug and Wozabal (2007) find the worst case probability measure, namely by changing the probability weights (q_1, q_2, q_3) of the support points (ξ_1, ξ_2, ξ_3). The right panel illustrates the approach in Pflug et al. (2012) as well as Esfahani and Kuhn (2015): they compute the worst case distribution by changing the support points (ξ_1, ξ_2, ξ_3) to $\left(\xi_1^{Q^\star}, \xi_2^{Q^\star}, \xi_3^{Q^\star}\right)$ while leaving their probability weights unchanged.

resulting semi-infinite program. As an alternative, Pflug et al. (2012) directly characterized the worst case distributions $Q^\star$, which satisfies $d_W(P_0, Q^\star) = \kappa$. Hence, problem (1.5) can be solved by a linear program. Recently, Esfahani and Kuhn (2015) derived a tractable reformulation of (1.5) as an LP by relying on state-of-the-art global optimization techniques. Loosely speaking, the authors change the location of the atoms of the empirical measure $\hat{P}_n$ rather than their probability weights, as done in Pflug and Wozabal (2007). Figure 1.1 illustrates this difference. In the remaining part of this section, we review and compare these three approaches to solve problem (1.5).

Proceeding in chronological order, we start our literature review with the paper by Pflug and Wozabal (2007). The authors proposed the following successive convex programming algorithm:

1. Set $i = 0$ and $\mathscr{P}_0 = \{\hat{P}_n\}$.

2. Solve the outer problem

$$\begin{aligned} &\max_{x \in \mathbb{X}, t \in \mathbb{R}} t \\ &\text{s.t.: } \mathbb{E}\left(x^\top \xi^Q\right) - \lambda \, \mathrm{AV@R}_\alpha\left(-x^\top \xi^Q\right) \geq t \qquad \forall Q \in \mathscr{P}_n, \end{aligned}$$

   and call the optimizer (x_i, t_i).

3. Set $Y_x = (x_i)^\top \xi$ and solve the inner problem

$$\min_{Q \in \mathscr{B}_\kappa(\hat{P}_n)} \mathbb{E}\left(Y_x^Q\right) - \lambda \, \mathrm{AV@R}\left(-Y_x^Q\right),$$

call the optimizer P_i and let $\mathscr{P}_{i+1} = \mathscr{P}_i \cup \{P_i\}$.

4. If either $\mathscr{P}_{i+1} = \mathscr{P}_i$ or the optimal value of the inner problem equals t_i, then a saddle point of problem (1.5) is found and the algorithm stops. Otherwise, set $i = i+1$ and go to step 2.

As $\mathscr{P}_n$ only contains finitely many objects, the outer problem can be solved as an LP. The inner problem however has a more complicated structure. As indicated above, Pflug and Wozabal (2007) restrict their analysis to probability measures Q which live on the same n points as the empirical measure $\hat{P}_n$. Hence, in the inner problem one optimizes over the probability weights $(q_i)_{i=1,\dots,n}$ of the n sample points, which have equal weight under the empirical measure $\hat{P}_n$. Still, the inner problem is non-convex due to bilinear terms defining the Wasserstein distance and the AV@R. We refer to section 4.2 of the follow-up paper by Wozabal (2012), which proposes a method to approximately solve this program rewriting it in terms of differences of convex (D.C.) functions.

Let us now come to the paper by Pflug et al. (2012), who explicitly derived the worst case distribution of $Q^\star \in \mathscr{B}_\kappa(P_0)$ for fixed portfolio weight x. In order to adapt the more general results in Pflug et al. (2012), notice that the maximin problem (1.5) is equivalent to the following minimax problem

$$\min_{x\in\mathbb{X}} \max_{Q\in\mathscr{B}_\kappa(P_0)} (\lambda+1)\mathscr{A}_\alpha(-x^\top \xi^Q) \tag{1.6}$$

where $\mathscr{A}_\alpha$ is a mixture of AV@R's with different α's, i.e., it has the Choquet representation:

$$\mathscr{A}_\alpha(X) = \frac{1}{\lambda+1}\mathrm{AV@R}_0(X) + \frac{\lambda}{\lambda+1}\mathrm{AV@R}_\alpha(X).$$

We denote by $Q^\star$ the worst case distribution in the Wasserstein-ball $\mathscr{B}_\kappa(P_0)$ for a fixed x, i.e., $Q^\star = \mathrm{argmax}_{Q\in\mathscr{B}_\kappa(P_0)}\{\mathscr{A}_\alpha(-x^\top \xi^Q)\}$. Applying the result in Pflug et al. (2012) to $\mathscr{A}_\alpha$, we find that the random vector of the asset returns $\xi^{Q^\star} = \left(\xi_1^{Q^\star},\dots,\xi_m^{Q^\star}\right)^\top$ under the worst case distribution $Q^\star$ is given by

$$\xi_j^{Q^\star} = \begin{cases} \xi_j^{P_0} + \mathrm{sign}(x_j)\mathrm{sign}(Z)\frac{\kappa}{k(1+\lambda)}\left(\frac{\lambda}{1-\alpha}+1\right) & \text{if } |x_j| = ||x||_\infty \\ \xi_j^{P_0} & \text{otherwise} \end{cases},$$

for all $j = 1,\dots,m$, where Z is in the set $\partial\mathscr{A}_\alpha\left(-x^\top \xi^{P_0}\right)$ of subgradients of $\mathscr{A}_\alpha$ at $-x^\top \xi^{P_0}$ and $k = \left|\left\{j \in \{1,\dots,m\} : |x_j| = ||x||_\infty\right\}\right|$. Setting the reference distribution P_0 to the empirical distribution $\hat{P}_n$, problem (1.6) can be written as the following linear program

$$\begin{aligned}
&\min_{a,W,x,z} \lambda a + \frac{\lambda}{1-\alpha}\frac{1}{n}\sum_{i=1}^{n} z_i - \frac{1}{n}\sum_{i=1}^{n} x^\top \xi_i + W && (1.7)\\
&\text{s.t.:} \sum_{j=1}^{m} x_j = 1 \\
&\quad z_i \geq -x^\top \xi_i + W - a && i = 1,\ldots,n \\
&\quad x_j \frac{\kappa}{\lambda+1}\left(\frac{\lambda}{1-\alpha}+1\right) \leq W && j = 1,\ldots,m \\
&\quad x_j \geq 0, z_i \geq 0 && i = 1,\ldots,n; j = 1,\ldots,m.
\end{aligned}$$

Notice that a very important implication of the above is that under high model ambiguity (i.e., $\kappa \to \infty$) portfolio diversification becomes optimal. We refer to Pflug et al. (2012) for a detailed and more general proof of this result. It should be mentioned that this result is confirmed by Esfahani and Kuhn (2015), which we review next.

In the above discussion of Pflug and Wozabal (2007), we have seen that quite evolved techniques are needed to solve problem (1.5), when restricting the feasible set to discrete measures with the same number of atoms as P_0, fixing the location of these atoms and optimizing over their probability. But what happens when one is more ambitious and instead fixes the probability and optimizes over the location of the atoms? Esfahani and Kuhn (2015) show that this leads to a tractable reformulation of the problem as an LP. We only restate their result here and refer to the original paper Esfahani and Kuhn (2015) for their derivation of more general results than needed in our context. First, we provide yet another equivalent reformulation of the original problem (1.5):

$$\min_{x\in\mathbb{X},\tau\in\mathbb{R}} \max_{Q\in\mathscr{B}_\kappa(P_0)} \mathbb{E}\left[\max_{k\leq 2} a_k\left(x^\top \xi^Q\right) + b_k\tau\right],$$

where $a_1 = -1, a_2 = (1-\alpha+\lambda)/(\alpha-1), b_1 = \lambda$ and $b_2 = \lambda\alpha/(\alpha-1)$. Assuming that $P_0 = \hat{P}_n$ and that the support of the *true* distribution P of ξ is contained in the set $\Xi := \{\xi \in \mathbb{R}^m : C\xi \leq d\}$, where C and d are of appropriate size, Esfahani and Kuhn (2015) show

that the above minimax problem can be written as the following linear program

$$\begin{aligned}
&\min_{x,\tau,\mu,z,\gamma_{ik}} \mu\kappa + \frac{1}{n}\sum_{i=1}^{n} z_i \qquad (1.8)\\
&\text{s.t.:} \sum_{j=1}^{m} x_j = 1\\
&\quad b_k\tau + a_k\left(x^\top \xi_i\right) + \gamma_{ik}^\top (d - C\xi_i) \le z_i \qquad k=1,2; i=1,\dots,n\\
&\quad |C^\top \gamma_{ik} - a_k x_j| \le \mu \qquad k=1,2; i=1,\dots,n, j=1,\dots,m\\
&\quad \gamma_{ik} \ge 0, x \ge 0 \qquad k=1,2; i=1,\dots,n, j=1,\dots,m.
\end{aligned}$$

Comparing the three discussed approaches, it is clear that solving the LP (1.7) takes the least computational effort as the worst case distribution is known a priori. The LP (1.8) given yields the same solution as (1.7), while being computationally considerably more expensive.[3] Still the results in Esfahani and Kuhn (2015) allow us to solve the problem surprisingly fast given that the worst case distribution is not computed a priori. Lastly, we should mention that the results in Pflug and Wozabal (2007) cannot keep up. The SCP algorithm is computationally very intense and a large sample size n is required if the algorithm should produce results similar to those of the other two approaches. Here the follow-up paper by Wozabal (2012) should be mentioned, which sightly improves the requirement of a large sample size n but does so at the cost of even more computational complexity.

We refer to section 1.5, where we present an empirical study based on problem (1.7) or equivalently (1.8) and consider the optimal portfolio decomposition when the level κ of model ambiguity increases.

1.4 Portfolio selection under dependence uncertainty

In many cases of practical interest, one might be able to determine the distribution of a portfolio's individual assets but fail to estimate their dependence structure accurately. In such a situation only the dependence structure of the asset returns is ambiguous and not their joint distribution. This situation is known as *dependence uncertainty*. The main motivation to study this topic comes from risk management, as it might lie in the nature of an application at hand, that the dependence structure of different risks cannot be estimated with high accuracy. Still the dependence structure is crucial when it comes to aggregating the individual risks. We refer to Aas and Puccetti (2014) for an illustrative example. Since risk measurement

[3] Notice that the LP in (1.8) has $2+m+3n$ variables and $2n+4mn$ inequality constraints compared with $2+m+n$ variables and $m+n$ inequality constraints in the LP (1.7).

is the corner stone of any portfolio selection strategy, the methods and results concerning dependence uncertainty are particularly relevant in our context. As it turns out, computing bounds for aggregated risks with given marginal distributions is a challenging task on its own. Therefore, we provide a short literature review and some examples in the following.

1.4.1 Background

First of all, let us introduce the notion of *copulas*. The intriguing yet simple idea underlying copulas is that they allow us to split the modeling of multivariate random variables into two parts: Firstly, the modeling of the marginal distributions and secondly, the modeling of the dependence between the univariate random variables. The latter is done via copula functions. Thus, in the presence of dependence uncertainty the first task is straight forward whereas the second task is complicated due to the lack of information. To put it in a nutshell: dependence uncertainty means marginals known and copula unknown.

Formally, let H be an m-dimensional cumulative distribution function with margins $F_1, F_2, \dots, F_m$. Then there exists a function C, which is called *m-copula* such that

$$H(t_1, t_2, \dots, t_m) = C\left(F_1(t_1), F_2(t_2), \dots, F_m(t_m)\right), \qquad \forall t_1, \dots, t_m \in \mathbb{R}.$$

This basic connection is known as Sklar's Theorem. For a precise definition of copulas and an introduction to the topic, the reader is referred to Nelsen (2007).

Let us consider a simple example to understand the essence of dependence uncertainty and its connection to the concept of copulas: What is the Average-Value-at-Risk $\mathrm{AV@R}_\alpha(U+V)$ of the sum of two standard uniforms U and V, i.e., $U, V \sim \mathrm{Uniform}[0,1]$? The answer depends on the copula linking the uniforms U and V:

1. Assume U and V are counter-monotonic, i.e., perfectly negative dependent. Their copula is given by the lower Fréchet Hoeffding bound $W(u,v) = \max(u+v-1, 0)$ for all $u, v \in [0,1]$. In this case $\mathrm{AV@R}_\alpha^W(U+V) = 1$ for all α.[4]

2. Assume U and V are comonotonic, i.e., perfectly positive dependent. Their copula is given by the upper Fréchet Hoeffding bound $M(u,v) = \min(u,v)$ for all $u, v \in [0,1]$. In this case $\mathrm{AV@R}_\alpha^M(U+V) = 1+\alpha$.

3. Assume U and V are independent and hence coupled with the product copula $\Pi(u,v) = uv$ for all $u, v \in [0,1]$. A tedious calculation shows that if $\alpha > 0.5$, $\mathrm{AV@R}_\alpha^\Pi(U+V) = 2 - 2\sqrt{2(1-\alpha)}/3$.

[4] Notice that only in this example we indicate the copula linking U and V in the superscript of AV@R.

Given that we have no information regarding the copula linking U and V, the best we can do is compute bounds for $\mathrm{AV@R}_\alpha(U+V)$:

$$1 = \mathrm{AV@R}_\alpha^W(U+V) \leq \mathrm{AV@R}_\alpha^C(U+V) \leq \mathrm{AV@R}_\alpha^M(U+V) = 1+\alpha,$$

for any bivariate copula C. Note that these bounds are very far apart. This issue is addressed below. Let us first discuss the upper bound. A nice property of the AV@R is that the copula which yields the upper bound turns out to be the comonotonic copula M. This property holds true in arbitrary dimensions due to the fact that the AV@R is subadditive (see Pflug, 2000). Intuitively, it is clear that perfect positive dependence implies no diversification benefit when trying to minimize the overall risk of your portfolio. Hence, this dependence should be the least favored by any investor. The AV@R reflects this, whereas other risk functionals, like the Value-at-Risk (V@R), do not. In fact, finding upper bounds for the V@R is much harder. Makarov (1982) and Rüschendorf (1982) solved this problem in two dimensions, showing that the following copula is the one maximizing the V@R with confidence level α:

$$C_\alpha(u,v) = \begin{cases} \max(W(u,v),\alpha) & (u,v) \in [\alpha,1]\times[\alpha,1] \\ M(u,v) & \text{otherwise} \end{cases}. \tag{1.9}$$

As can been seen, counter-monotonicity in the upper tail is needed for the construction. This is precisely the reason why the result cannot be generalized to higher dimensions and leads us directly to the discussion of the difficulties that arise when computing lower bounds in dimensions strictly larger than 2. As a matter of fact, the concept of counter-monotonicity does not naturally extend to higher dimensions. To give some intuition, perfect negative dependence between two random variables implies that whenever one random variable moves into one direction, the other one moves into the opposite direction. Arguably, considering perfect negative dependence between three random variables is not as simple since it is not clear when three direction stand in maximal contrast. More formally, this is reflected by the fact that the multivariate extension of the lower Fréchet Hoeffding bound $W(u_1,\ldots,u_n) = \max\left(\sum_{i=1}^n u_i - n, 0\right)$ ceases to be a copula (see Nelsen, 2007). We refer to Wang and Wang (2015) for a discussion of possible ways to define perfect negative dependence in arbitrary dimensions.

In the context of dependence uncertainty, the best and worst case value of the V@R as well as the worst case value of the AV@R are in general unknown. Nevertheless, these bounds have been extensively studied in recent years and partial solutions have been obtained (see Puccetti and Wang, 2015) for an overview. Most importantly, Puccetti and Rüschendorf (2012) introduced the so-called *rearrangement algorithm* which is a fast procedure to numerically

compute the bounds of interest. Under quite restrictive assumptions, analytical bounds can be computed based on the notion of *complete mixability*, which was introduced by Wang and Wang (2011).

In practical applications, these lower and upper bounds cease to be useful as they are too far apart. Thus, the case where additional dependence information is available, received some attention in recent years. The literature often refers to this topic as *partial dependence uncertainty* in contrast to *complete dependence uncertainty*, which we discussed above. Bounds for aggregated risks have been computed in the case where (a) the copula is given on a subset of the unit cube, (b) bounds for the copula itself are given or (c) bounds for the variance of the aggregated risks are given. We refer to the survey by Rüschendorf (2016) and references therein. Most interestingly, we want to point out the work by Lux and Papapantoleon (2016), who among other intriguing results derived V@R-bounds for the case where the copula is known to be close in the sense of a statistical distance to some reference copula.

This short detour to the existing literature on dependence uncertainty should serve as a motivation to model the situation where the marginal distribution of the individual assets is assumed to be known, whereas their dependence structure (i.e., the copula linking them) is only partially known in the context of portfolio selection. The fact, that the combination of portfolio optimization with dependence uncertainty is of practical and theoretical interest, has of course been noted by others before. For instance, Kakouris and Rustem (2014) introduce an investment strategy which is robust against possible misspecification of the chosen parametric copula family. On the contrary, the more involved framework by Doan et al. (2015) does not need any parametric assumptions. The authors derive AV@R bounds for the Fréchet class of discrete, multivariate and overlapping marginals with finite support. Optimizing these bounds leads to an investment strategy based on a minimum of distributional assumptions.

We, however, take a different and novel approach to the portfolio selection problem under dependence uncertainty, which is very much in the spirit of section 1.3:

$$\max_{x\in\mathbb{X}} \min_{C\in\mathscr{C}_\varepsilon(C_0)} \mathbb{E}(x^\top \xi) - \lambda \mathscr{R}(-x^\top \xi^C), \tag{1.10}$$

where $\mathscr{C}$ denotes the set of all m-dimensional copulas and the ambiguity sets $\mathscr{C}_\varepsilon(C_0) := \{C \in \mathscr{C} : d_W(C_0, C) \leq \varepsilon\}$ refers to the set of m-copulas which are close to some reference copula C_0 with respect to the Wasserstein distance. Since the copula has no impact on the expected value of $x^\top \xi$, we omit the letter C in the superscript of the first ξ in (1.10). We use the Wasserstein distance as a distance measure between copulas in order to guarantee that problem (1.10) is defined similarly to the portfolio selection problem under model ambiguity

introduced in equation (1.4). In fact, the only difference is that in (1.10) we assume that the marginal distributions of the asset returns are known whereas in (1.4) this assumption is not made. Still the question may arise why the Wasserstein distance is a meaningful distance measure for copulas. The reason lies in the following fact: Suppose that the pairs of random variables (X_1, X_2) resp. (Y_1, Y_2) have the same marginals F_1 and F_2 but differ in their copulas C_X and C_Y. If h is a Lipschitz function with two arguments and Lipschitz constant L and both quantile functions F_1^{-1} and F_2^{-1} are Lipschitz with constant K, then

$$\mathbb{E}\left[|h(X_1, X_2) - h(Y_2, Y_2)|\right] \leq LKd_W(C_X, C_Y).$$

More generally, take a distribution P on $\mathbb{R}^m$ with non-Lipschitz quantile functions, which are assumed not to grow faster then $-x^\gamma$ near zero and $(1-x)^\gamma$ near 1, then one may define β-Lipschitz functions h, which satisfy

$$|h(x_1, \ldots, x_m) - h(y_1, \ldots, y_m)| \leq L \sum_{i=1}^{m} \left| \operatorname{sgn}(x_i) \cdot |x_i|^\beta - \operatorname{sgn}(y_i) \cdot |y_i|^\beta \right|,$$

for $\beta < 1$. If $\beta < 1/\gamma$, then $P \mapsto \int h \, dP$ is, for fixed marginals, Lipschitz in the Wasserstein distance of the copula.

Concerning the choice of the risk functional $\mathscr{R}$, we want to point out that the AV@R is *aggregation-robust*, see Embrechts et al. (2015), which implies that the AV@R is less sensitive to model uncertainty at the level of the dependence structure than, for instance, the Value-at-Risk. Hence, the AV@R will be of special interest in our subsequent analysis.

1.4.2 Complete dependence uncertainty

Before discussing how to approach problem (1.10) computationally, we first want to consider the case of *complete dependence uncertainty*: Assume that the degree of dependence uncertainty, given by the radius ε of the ambiguity set in (1.10), converges to infinity. Then the ambiguity set $\mathscr{C}_\varepsilon(C_0)$ extends to the set of all copulas $\mathscr{C}$. Thus, we obtain the problem

$$\max_{x \in \mathbb{X}} \min_{C \in \mathscr{C}} \mathbb{E}(x^\top \xi) - \lambda \mathscr{R}(-x^\top \xi^C). \tag{1.11}$$

In the following we show that the above problem can be simplified significantly when we restrict our considerations to (a) subadditive, (b) comonotone additive and (c) positively homogeneous risk functionals $\mathscr{R}$. A risk functional $\mathscr{R}$ is said to be

(a) subadditive, if for all random variables X and Y, $\mathscr{R}(X+Y) \leq \mathscr{R}(X) + \mathscr{R}(Y)$.

(b) comonotone additive, if for any comonotonic[5] random variables X and Y it holds that $\mathscr{R}(X+Y) = \mathscr{R}(X) + \mathscr{R}(Y)$.

(c) positively homogeneous, if for all random variables X and all constants $c \geq 0$ it holds that $\mathscr{R}(cX) = c\mathscr{R}(X)$.

We refer to Pflug and Römisch (2007) for a detailed discussion of these and other properties of risk functionals. In the following we use the notation ξ_j to denote the j-th component of the $m-$dimensional (random) vector ξ.

Proposition 1.1. *If the risk functional $\mathscr{R}$ is subadditive, comonotone additive and positively homogeneous, then*

$$\max_{x\in\mathbb{X}} \min_{C\in\mathscr{C}} \mathbb{E}(x^\top \xi) - \lambda \mathscr{R}(-x^\top \xi^C) = \mathbb{E}\left(\xi_{j^\star}\right) - \lambda \mathscr{R}(-\xi_{j^\star}),$$

for any $j^\star \in \operatorname{argmax}_j \mathbb{E}\left(\xi_j\right) - \lambda \mathscr{R}(-\xi_j)$.

Proof. First notice that

$$\mathscr{R}\left(-x^\top \xi^C\right) \overset{\text{(a)}}{\leq} \sum_{j=1}^m \mathscr{R}\left(-x_j \xi_j\right) \overset{\text{(b)}}{=} \mathscr{R}\left(-x^\top \xi^M\right),$$

for all copulas C. Hence, we have that $\max_{C\in\mathscr{C}} \mathscr{R}\left(-x^\top \xi^C\right) = \mathscr{R}\left(-x^\top \xi^M\right)$ and we obtain

$$\begin{aligned}
&\max_{x\in\mathbb{X}} \min_{C\in\mathscr{C}} \mathbb{E}(x^\top \xi) - \lambda \mathscr{R}(-x^\top \xi^C) \\
&= \max_{x\in\mathbb{X}} \mathbb{E}(x^\top \xi) - \lambda \mathscr{R}(-x^\top \xi^M) \\
&\overset{\text{(b)}}{=} \max_{x\in\mathbb{X}} \sum_{j=1}^m x_j \mathbb{E}(\xi_j) - \lambda \sum_{j=1}^m \mathscr{R}(-x_j \xi_j) \\
&\overset{\text{(c)}}{=} \max_{x\in\mathbb{X}} \sum_{j=1}^m x_j \left(\mathbb{E}(\xi_j) - \lambda \mathscr{R}(-\xi_j)\right) \\
&= \mathbb{E}\left(\xi_{j^\star}\right) - \lambda \mathscr{R}(-\xi_{j^\star}),
\end{aligned}$$

for any $j^\star \in \operatorname{argmax}_{j\in\{1,\ldots,m\}} \mathbb{E}\left(\xi_j\right) - \lambda \mathscr{R}(-\xi_j)$. In the last equality we use the fact that $\mathbb{X}$ is the m-dimensional unit simplex. □

[5] Two random variables X and Y are said to be comonotonic if there exists a real-valued random variable Z and non-decreasing functions f_1 and f_2 such that $X = f_1(Z)$ and $Y = f_2(Z)$ or equivalently if the copula linking them is given by the comonotonic copula M.

Note that due to the high precision of modern data processors the argmax set will in practice only contain one element. Thus, we find that complete dependence uncertainty implies that concentration of the portfolio into one single asset is optimal. At first, this insight might be somewhat counterintuitive since putting all the capital into a single asset does not seem to be a robust decision. In fact, the decision when facing high model uncertainty on the level of the joint distribution could not differ more from the decision which is optimal under complete dependence uncertainty: the first implies portfolio diversification, the latter portfolio concentration. However, looking at it from the technical side, on the one hand subadditivity rewards portfolio diversification but on the other hand comonotone additivity does not attribute any diversification benefits in case of comonotone losses. Hence, without any knowledge of the dependence between the asset returns, the worst case is that all asset prices rise and fall simultaneously. Basing our decision on this worst case, portfolio diversification cannot be beneficial. Rather, investing into a single asset, namely one of those which performs best, is optimal.

Alternatively, we can approach the result from an information theoretical point of view: a decision is better the more informed it is. The equal weighted portfolio investment strategy requires no information at all. Thus, equally-weighted portfolio diversification is optimal in case of high model ambiguity and suboptimal under complete dependence uncertainty. Portfolio concentration, however, requires full information on the marginal distributions to determine which asset is the best. Therefore, it is not surprising that portfolio concentration is optimal when the model uncertainty lies solely on the level of the dependence structure.

We conclude that subadditivity, comonotone additivity and positively homogeneity are indeed reasonable properties for risk functionals in the portfolio selection problem under dependence uncertainty. Let us therefore take a closer look at risk functionals with these three properties. Spectral risk measures, introduced by Acerbi (2002), and distortion risk measures, which developed from research on premium principles by Wang (1996), are well-known families of risk functionals with these properties as they are coherent (in the sense of Artzner et al. (1999)) and comonotone additive and law invariant (see Föllmer and Schied, 2011; Rüschendorf, 2013). Arguably, the most relevant representative of these families is the AV@R, as it can be seen as the basic building block for spectral risk measures (see Kusuoka, 2001).

Nevertheless, well known risk functionals violate at least one of the mentioned properties: the variance (which has neither of these properties), the standard deviation (which is not comonotone additive), the Value-at-Risk (which is not subadditive) and expectiles (which are in general not comonotone additive), see Emmer et al. (2015). The questions arises what happens if one of the latter risk functionals is chosen in the portfolio selection problem (1.11)

under complete dependence uncertainty. The following example addresses this question.

Example 2.1 Let some portfolio be composed of two assets. We represent the two assets by random variables U and W, where $W = 2V^2$ and U and V are both Uniform[0,1]. Thus, the second asset promises a higher return at a higher risk. For $0 \leq x \leq 1$, the portfolio with weights $1-x$ resp. x has return

$$Y_x^C = (1-x)U + x2V^2,$$

where C denotes the copula linking U and V. Assuming the copula C is unknown leads to the following portfolio selection problem:

$$\max_{x\in[0,1]} \min_{C\in\mathscr{C}} \mathbb{E}(Y_x^C) - \lambda \mathscr{R}(Y_x^C). \tag{1.12}$$

We analyze four different risk measures $\mathscr{R}$: i) the variance (Var), ii) the standard deviation (Std), iii) the Average-Value-at-Risk (AV@R) and iv) the Value-at-Risk (V@R). In case i) and ii), $\mathscr{R}(Y_x^C)$ is maximal if $\mathrm{Cov}(U,V^2)$ is maximal. According to Theorem 5.25 in McNeil et al. (2015), this means that U and V^2 are comonotone, i.e., the worst case copula is given by the comonotonic copula M and we can set $V = U$. The same statement holds in case iii) where $\max_{C\in\mathscr{C}} \mathrm{AV@R}(Y_x^C) = \mathrm{AV@R}(Y_x^M)$. In case iv), it was mentioned above that $\max_{C\in\mathscr{C}} \mathrm{V@R}(Y_x^C) = \mathrm{V@R}(Y_x^{C_\alpha})$ where the worst case copula C_α is given in equation (1.9). Note that C_α depends on the confidence level α but not on the portfolio weight x. Problem (1.12) can now be solved explicitly for the four different choices of $\mathscr{R}$:

i) $\max_{x\in[0,1]} \mathbb{E}(Y_x) - \lambda \mathrm{Var}(Y_x^M) = \max_{x\in[0,1]} \frac{1}{2} + \frac{x}{6} - \lambda\left(\frac{1}{12} + \frac{x}{6} + \frac{19x^2}{180}\right)$.

ii) $\max_{x\in[0,1]} \mathbb{E}(Y_x) - \lambda \mathrm{Std}(Y_x^M) = \max_{x\in[0,1]} \frac{1}{2} + \frac{x}{6} - \lambda\sqrt{\frac{1}{12} + \frac{x}{6} + \frac{19x^2}{180}}$.

iii) $\max_{x\in[0,1]} \mathbb{E}(Y_x) - \lambda \mathrm{AV@R}_\alpha(Y_x^M) = \max_{x\in[0,1]} \frac{1}{2} + \frac{x}{6} - \lambda\left(\frac{4\alpha^2 x+\alpha x+x+3\alpha+3}{6}\right)$.

iv) $\max_{x\in[0,1]} \mathbb{E}(Y_x) - \lambda \mathrm{V@R}_\alpha(Y_x^{C_\alpha})$
$= \max_{x\in[0,1]} \frac{1}{2} + \frac{x}{6} - \lambda \min\left(2x\alpha^2 + (1-x), \alpha(1-x) + 2x\right)$ for $\alpha \geq 0.5$. [6]

The optimal portfolio decomposition for the four different risk functionals is illustrated in figure 1.2 as a function of λ: For low risk aversion λ the portfolio is concentrated in second asset which promises a higher return, whereas in case of high risk aversion, the portfolio is concentrated in the less risky first asset. As indicated by Proposition 1.1, when using the

[6] For the computation of $\mathrm{V@R}_\alpha(Y_x^{C_\alpha})$ Proposition 1 in Denuit et al. (1999) is very useful.

AV@R as a risk measure, the optimal portfolio is concentrated for all λ (see the second panel from the bottom in figure 1.2). The panel at the bottom of figure 1.2 shows that the same also holds true when using the V@R. In case i) and ii) we observe that the portfolio is composed for certain values of risk aversion. For instance, in case ii) $\lambda = \sqrt{695}/49 \approx 0.54$ implies that portfolio diversification into half/half is optimal. □

This example shows us that independently of the chosen risk functionals, optimal portfolios under complete dependence uncertainty are typically concentrated. Only for specific values of the risk aversion coefficient λ it might happen that the portfolio is composed. For subadditive, comonotone additive and positively homogeneous risk functionals, composed portfolios can even be ruled out a priori, which constitutes another advantage of this class of risk measures. In fact, we can go even further and conjecture that for any risk functional the optimal portfolio under complete dependence uncertainty is never composed of more than two assets.

Summing up, portfolio concentration is a robust decision when facing complete dependence uncertainty and the choice of subadditive, comonotone additive and positively homogeneous risk functionals is particularly convenient in this context.

1.4.3 Partial dependence uncertainty

Let us now discuss problem (1.10) for $\varepsilon > 0$. In particular, we assume that a reference copula C_0 is given and we consider an ambiguity set of copulas which are close to this reference copula. In the spirit of section 1.3, we avoid all kinds of parametric assumptions and choose the empirical copula $\hat{C}_n$ to be the reference copula C_0. The empirical m-copula $\hat{C}_n$ generated by n points is defined as

$$\hat{C}_n(u_1,\dots,u_m) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left(U_{i,1} \le u_1,\dots,U_{i,m} \le u_m\right), \tag{1.13}$$

with $u_1,\dots,u_m \in [0,1]$ and $U_{i,j} = n\hat{F}_j(\xi_{i,j})/(n+1)$ for all $i = 1,\dots,n$ and $j = 1,\dots,m$, where $\hat{F}_j$ denotes the empirical marginal distribution function of asset j, $(\xi_{i,j})_{i=1,\dots,n;j=1,\dots m}$ denotes the matrix which contains n observations of the m asset returns and the scaling factor $n/(n+1)$ is only introduced to avoid potential problems at the boundary of $[0,1]^m$ (see Genest et al., 2009). Various conditions under which $\hat{C}_n$ is a consistent estimator of the *true* underlying copula are given in Fermanian et al. (2004). In the formulation of problem (1.10) the ambiguity set $\mathscr{C}_\varepsilon(C_0)$ contains infinitely many objects for all $\varepsilon > 0$. This reminds us of the more general, yet similar problem (1.5) and the ambiguity set $\mathscr{B}_\kappa(P_0)$. As shown in Esfahani

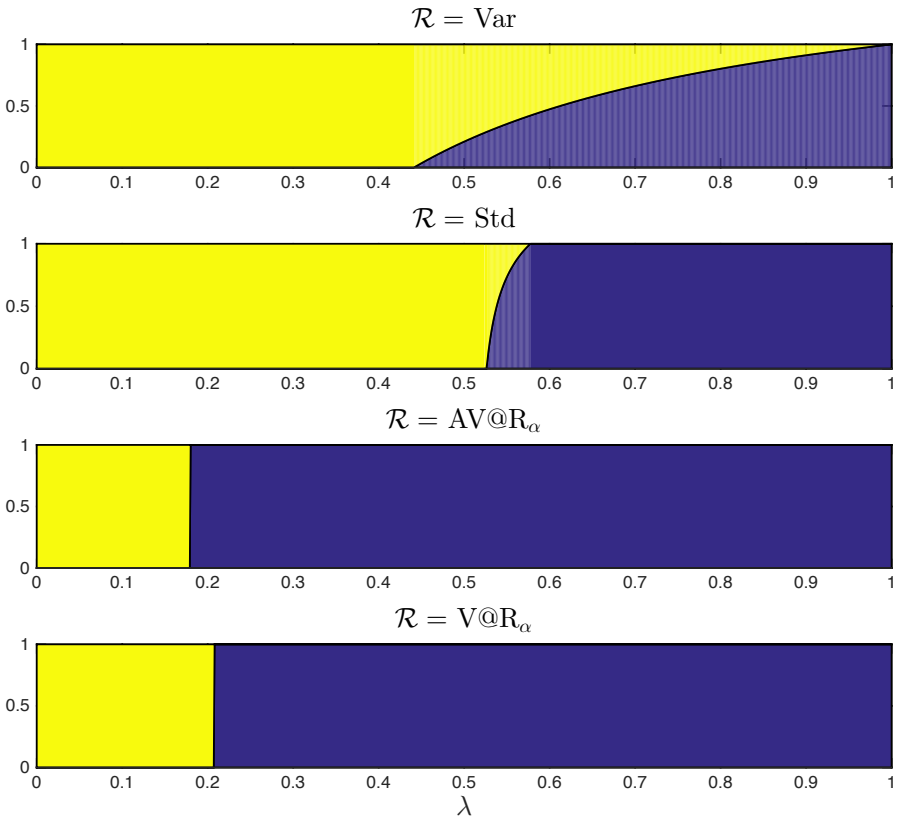


Fig. 1.2 The optimal portfolio decomposition of the portfolio selection problem under complete dependence uncertainty, which is given in equation (1.12), is shown for four different risk functionals $\mathscr{R}$ as a function of the risk aversion $\lambda \in [0,1]$ and for fixed $\alpha = 0.95$. For a fixed λ, the bright area corresponds to the fraction x which should be invested in the second asset whereas the dark area gives the optimal weight $1-x$ of the first asset.

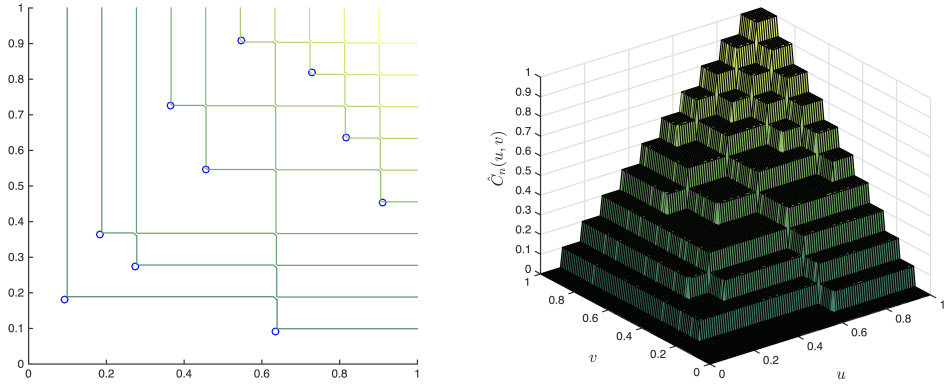


Fig. 1.3 The left panel shows a scatter plot of points $(U_{i,j})_{i=1,\ldots,10,j=1,2}$ generating the empirical copula $\hat{C}_n$ with $n = 10$ plotted in the right panel.

and Kuhn (2015); Pflug et al. (2012) and discussed in section 1.3, when setting $P_0 = \hat{P}_n$, the worst case distribution also lives on n points. Thus, we conjecture that when setting $C_0 = \hat{C}_n$, the worst case copula also lives on n points. We therefore restrict our considerations to the smaller set $\mathscr{C}_{\varepsilon,n}(\hat{C}_n) := \left\{C \in \mathscr{C}_n : d_W(C,\hat{C}_n) \leq \varepsilon\right\}$ of all empirical m-copulas resulting from n observations which are close to $\hat{C}_n$. This restriction as well as choosing the AV@R as a risk functional, leads to the following problem:

$$\max_{x\in\mathbb{X}} \min_{C\in\mathscr{C}_{\varepsilon,n}(\hat{C}_n)} \mathbb{E}\left(x^\top \xi\right) - \lambda \mathrm{AV@R}_\alpha\left(-x^\top \xi^C\right). \tag{1.14}$$

In order to solve this problem, we have to understand the objects in the set $\mathscr{C}_{\varepsilon,n}(\hat{C}_n)$. From definition (1.13) we know that any empirical copula $C \in \mathscr{C}_n$ is determined by its support points $U_{i,j} = n\hat{F}_j(\xi_{i,j})/(n+1)$, where $i = 1,\ldots,n$ and $j = 1,\ldots,m$. The construction in (1.13) guarantees that the margins of the distribution induced by $(U_{i,j})$ are uniformly distributed on $[0,1]$. Figure 1.3 shows an example of points $\left(U_{i,j}\right)_{i=1,\ldots,n;j=1,2}$ generating an empirical bivariate copula. This graph illustrates the so-called *chess-tower property*, which is generally satisfied by all points $\left(U_{i,j}\right)_{i=1,\ldots,n;j=1,\ldots,m}$ generating an empirical copula $C \in \mathscr{C}_n$: For fixed $j \in \{1,\ldots,m\}$ and fixed $k \in \{1,\ldots,n\}$ there exists a unique $i \in \{1,\ldots,n\}$ such that $U_{i,j} = k/(n+1)$, since $U_{i,j} = n\hat{F}_j(\xi_{i,j})/(n+1) = R_{i,j}/(n+1)$ where $(R_{i,j})$ denotes the ranks corresponding the the observations $(\xi_{i,j})$. This implies that there are $(m-1)!n!$ elements in the set $\mathscr{C}_n$.

Hence, solving problem (1.14) for fixed portfolio weights x amounts to a combinatorial problem. Solving it by enumeration requires to compute the Wasserstein distance $d_W(C,\hat{C}_n)$ for all $(m-1)!n!$ empirical copulas C in $\mathscr{C}_n$. Even for fixed portfolio weights x, this task

is computationally infeasible for reasonable values of m and n. However, we can propose a heuristic, which allows us to rapidly compute a lower bound $C_L \in \mathscr{C}_{\varepsilon,n}(\hat{C}_n)$ and an upper bound $C_U \in \mathscr{C}_n \setminus \mathscr{C}_{\varepsilon,n}(\hat{C}_n)$ such that

$$\mathrm{AV@R}_\alpha\left(-x^\top \xi^{C_L}\right) \leq \max_{C \in \mathscr{C}_{\varepsilon,n}(\hat{C}_n)} \mathrm{AV@R}_\alpha\left(-x^\top \xi^{C}\right) \leq \mathrm{AV@R}_\alpha\left(-x^\top \xi^{C_U}\right), \qquad (1.15)$$

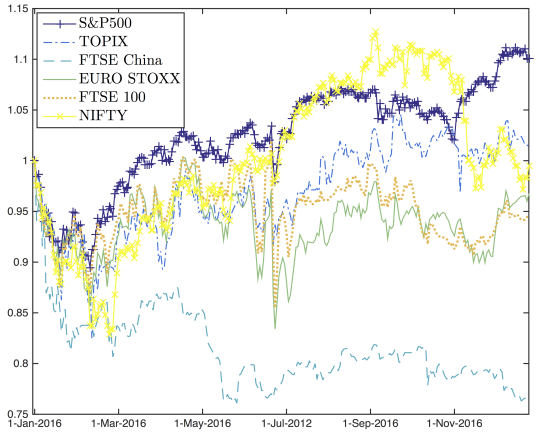
for all $x \in \mathbb{X}$. An illustration of the "true" worst case copula and the corresponding bounds can be seen in figure A.2 in the appendix. Let us sketch the main idea behind the proposed heuristic: First, note that the discrete counterpart M_n of the comonotonic copula M is defined by the points $U_{i,j} = i/(n+1)$ for all i and j, i.e., M_n lives on the main diagonal of the unit cube. Therefore, our heuristic iteratively transports points to the main diagonal of the unit cube starting with the smallest points in the lower tail of the copula until the bound ε of the maximal feasible Wasserstein distance is exceeded. The last feasible discrete copula is set to be C_L and the first infeasible copula defines C_U. Appendix A.1 describes this heuristic for the case $m = 2$. The procedure extends naturally to higher dimensions m. We omit a further discussion here.

1.5 Empirical study: concentration versus diversification

In the example studied in this section, the asset universe consists of the following six indices: S&P 500, TOPIX, FTSE China B35, EURO STOXX 50, FTSE 100 and NIFTY 500. The markets corresponding to these indices are USA, Japan, China, Eurozone, UK and India. We look at daily returns from the year 2016 (i.e., January 1, 2016 to December 31, 2016) computed from the prices in USD[7] of the indices which we normalized to one at the initial time. Figure 1.4 shows these normalized prices and summarizes the expected value and the AV@R of the corresponding returns. The data $\left(\xi_{i,j}\right)_{i=1,\dots,n;j=1,\dots,m}$, consisting of $n = 260$ observations of daily returns of $m = 6$ assets, is retrieved from the Thomson Reuters datastream licensed at the University of Vienna.

The aim of this empirical study is to graphically compare the two distinct approaches in section 1.3 and 1.4: Firstly, we solve the portfolio selection problem under model ambiguity as given in (1.5) relying on the methods of Pflug et al. (2012) and Esfahani and Kuhn (2015). Thus, we solve the LP (1.7) and (1.8) given ξ, α and λ for different values of κ. It should be mentioned once more that the two LPs yield indeed the same solution. Studying the optimal portfolio decomposition given by the portfolio weights x as the level of model ambiguity κ increases, we expect to find that portfolio diversification becomes optimal.

[7] Note that the FX-risk is neglected in this study.



	$\mathbb{E}(\cdot)$ (10^{-4})	$\mathrm{AV@R}_{0.95}(\cdot)$ (10^{-2})
S&P500	3.67	1.94
TOPIX	0.48	3.40
FTSE China	-9.06	3.41
EURO STOXX	-0.86	3.19
FTSE 100	-1.57	3.44
NIFTY 500	0.30	2.75

Fig. 1.4 The graph on the left shows six equity indices in 2016, which have been normalized to 1 on January 1, 2016. The table on the right shows the expected value as well as the 95%-AV@R of the corresponding daily losses.

Indeed, this is confirmed by the right plot in Figure 1.5, in which we set $\alpha = 0.95, \lambda = 10$ and ξ as discussed in the previous paragraph and plot the optimal portfolio decomposition x for different values of κ. For small values of κ, which means that the investor trusts in the empirical distribution function computed from the observed returns, the right plot in Figure 1.5 indicates that investing more than half of the capital into the S&P 500 is optimal. This fraction decreases to $1/6$ as the level κ of model ambiguity increases and stands in contrast to the left plot of Figure 1.5, which illustrates the optimal portfolio decomposition under dependence uncertainty. Here, we assume that the empirical marginal distribution functions indeed provide an accurate model for the marginal returns, whereas the dependence structure is ambiguous and only known to be within an $\varepsilon-$neighborhood of the empirical copula computed from the observations. In other words, we consider the maximin problem (1.14). If ε is close to zero, we are in the same situation as discussed before when $\kappa \approx 0$. For sufficiently large ε however, the worst case dependence structure is comonotonicity and portfolio concentration becomes optimal, as proven in section 1.4.2. In our example we find that, as ε increases, more and more capital should be invested into the S&P 500, which is in our example the best performing asset. It should be mentioned that we have used the heuristic algorithm presented in section 1.4.3 and appendix A.1 to generate the left graph in figure 1.5. In fact, this graph displays the portfolio weights x which are optimal with respect to the feasible lower bound C_L in equation (1.15). The lower bound C_L corresponds to the upper bound shown in the left graph of figure 1.6, which plots the value of the objective function in equation (1.14) as a function of ε. The absolute difference between the upper and

the lower bound of the objective function's value is 0.0051 ± 0.0037 when averaged over 20 different values of $\varepsilon \in (10^{-4}, 10^{-1})$. Hence, we conclude that the results of the proposed heuristic algorithm need to treated carefully when a fixed degree ε of dependence uncertainty is considered. Nevertheless, the heuristic allows us to determine the structure of the optimal portfolio decomposition as the degree of dependence uncertainty increases.

1.6 Conclusion

Figure 1.5 sums up the discussion of model ambiguity and dependence uncertainty in this paper. In the center of the Figure, we can see the optimal portfolio decomposition when we assume that the returns follow a known distribution P_0. Aiming at an investment strategy which is robust with respect to this distributional assumption, we can go into two different directions. On the one hand, we can assume that the true distribution lies in a $\kappa-$neighborhood of the distribution P_0 and optimize the portfolio with respect to the worst case distribution within the $\kappa-$neighborhood. As shown in the right plot of Figure 1.5, the higher the level of model ambiguity κ gets, the more diversified the portfolio becomes. On the other hand, we might only consider ambiguity with respect to the dependence structure of the distribution P_0 while fixing the marginal distributions. This leads to portfolio optimization under dependence uncertainty and is shown in the left plot of Figure 1.5: The higher the degree of dependence uncertainty gets, the more concentrated the optimal portfolio becomes. Although the set-up of these two approaches is very similar as all parametric assumptions are avoided and the definition of the ambiguity set relies in both cases on the Wasserstein distance, the implication could not be more dissimilar.

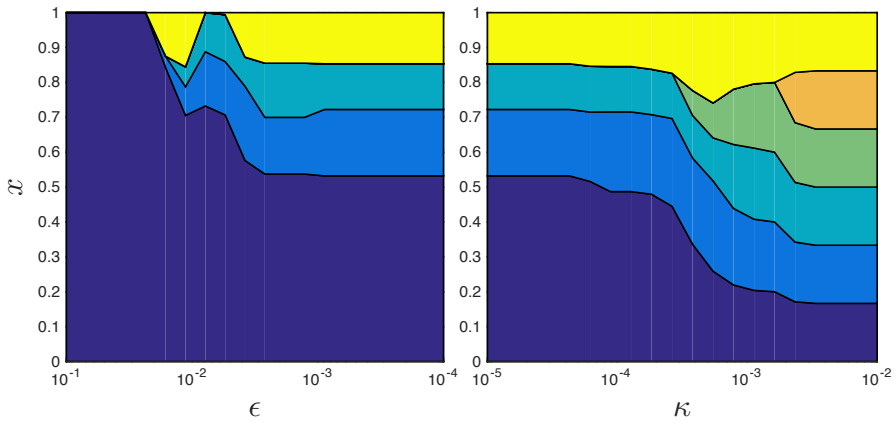


Fig. 1.5 In the graph on the right, the optimal portfolio decomposition is shown as a function of the level κ of model ambiguity. The graph on the left shows an approximation to the optimal portfolio decomposition as the degree ε of dependence uncertainty increases from right to left.

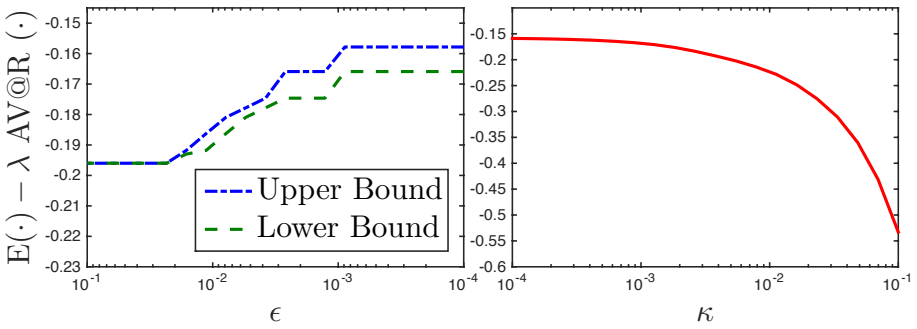


Fig. 1.6 The value of the objective function corresponding to the portfolio decomposition shown in figure 1.5 are plotted as a function of the level κ of model ambiguity as well as of the degree ε of dependence uncertainty.

CHAPTER 2

THE AMAZING POWER OF DIMENSIONAL ANALYSIS: QUANTIFYING MARKET IMPACT

JOINT WORK WITH A. RISTIG, W. SCHACHERMAYER AND L. TANGPI

ABSTRACT.

This note complements the inspiring work on dimensional analysis and market microstructure by Kyle and Obizhaeva. Following closely these authors, our main result shows by a similar argument as usually applied in physics the following remarkable fact. If the market impact of a meta-order only depends on four well-defined and financially meaningful variables, and some obvious scaling relations as well as the assumption of leverage neutrality are satisfied, then there is only one possible form of this dependence. In particular, the market impact is proportional to the square-root of the size of the meta-order.

This theorem can be regarded as a special case of a more general result of Kyle and Obizhaeva. These authors consider five variables which might have an influence on the size of the market impact. In this case one finds a richer variety of possible functional relations which we precisely characterize. We also discuss the analogies to classical arguments from physics, such as the period of a pendulum.

2.1 Introduction

Dimensional analysis is a well known line of argument in physics. The idea is best explained by considering a classical example: The period of a pendulum.

The *basic assumption* is that the period depends *only* on the following quantities:

- the length l of the pendulum, measured in meters,
- the mass m of the pendulum, measured in grams,
- the acceleration g caused by gravity, measured in meters per second squared.

The basic assumption amounts to the formula,

$$\text{period} = f(l, m, g), \tag{2.1}$$

where the period is measured in seconds and f is an – a priori – arbitrary function.

Of course, relation (2.1) should not depend on whether we measure length by meters or inches, time by seconds or minutes, and mass by grams or pounds. Combining these three requirements with the *ansatz*

$$f(l, m, g) = \text{const} \cdot l^{y_1} m^{y_2} g^{y_3}, \tag{2.2}$$

these requirements translate into three linear equations in the variables y_1, y_2, y_3. The unique solution yields the well-known relation (see Huntley (1967) as well as Appendix B.1 below for the details)

$$\text{period} = \text{const} \cdot \sqrt{\frac{l}{g}}. \tag{2.3}$$

This result goes back as far as Galileo. The elementary linear algebra used in the above argument has been formalized in proper generality in the nineteenth century and is known under the name of "Pi-Theorem" (see Section 2.3 below). It is worth mentioning that in the present case, the *ansatz* (2.2) does not restrict the generality of the solution (2.3) (see Appendix B.1 below).

Kyle and Obizhaeva (2017a) have applied this line of argument to analyze the *market impact* of a *meta-order* (note that similar scaling invariance arguments were introduced by Kyle and Obizhaeva (2016)): think of an investor who wants to buy (or sell) a sizeable amount of an underlying stock within a limited time (e.g. two days). Of course, when placing this *meta-order* she will split it into smaller pieces, the actual orders, in some (hopefully)

clever way. Nevertheless, we expect the quoted prices to move to the disadvantage of the agent. We call the expected size of this price movement, measured in percentage of the price, the *market impact*, see Eisler et al. (2012).

We start by identifying the variables (and their dimensions $[\cdot]$) which we expect to have an influence on the size of the market impact:

- Q the size of the meta-order, measured in units of shares $[Q] = \mathbb{S}$,
- P the price of the stock, measured in units of money per share $[P] = \mathbb{U}/\mathbb{S}$,
- V the traded volume of the stock, measured in units of shares per time $[V] = \mathbb{S}/\mathbb{T}$,
- σ^2 the squared volatility of the stock, measured in percentage of the stock price per unit of time $[\sigma^2] = \mathbb{T}^{-1}$.

These 4 variables are measured in the units of the 3 fundamental dimensions shares $\mathbb{S}$, time $\mathbb{T}$ and money $\mathbb{U}$. Now we formulate the following basic assumption.

Assumption 2.1. *The market impact G depends* only *on the above* 4 *variables, i.e.,*

$$G = g(Q, P, V, \sigma^2), \tag{2.4}$$

where the function $g : \mathbb{R}_+^4 \to \mathbb{R}_+$ *as well as the quantity G are invariant under changes of the units chosen to measure the "dimensions"* $\mathbb{S}$, $\mathbb{T}$ *and* $\mathbb{U}$.

We note that G is a percentage of the quoted price of the stock; hence G is referred to as "dimensionless", i.e., it is invariant under a change of the units in which $\mathbb{S}$, $\mathbb{T}$ and $\mathbb{U}$ are measured. We thus encounter an analogous situation as in the pendulum example. There is, however, a serious difference to the pleasant situation encountered above: We now have 4 variables, namely Q, P, V and σ^2, but only 3 equations resulting from the scaling invariance for the fundamental dimensions $\mathbb{S}$, $\mathbb{T}$, $\mathbb{U}$. We need one more equation to obtain such a crisp result as in (2.3) above. Kyle and Obizhaeva (2017a) found a remedy; an additional "no arbitrage" type argument which can be deduced from transferring the Modigliani-Miller invariance principle to market microstructure. To fix ideas, consider a stock which is a share of a company. Suppose that the company changes its capital structure by paying dividends or, passing to the opposite sign, by raising new capital. The Modigliani-Miller theorem precisely tells us which quantities *remain unchanged* when varying the leverage in terms of the relation between debt and equity of the company. This insight should furnish one more equation to be satisfied by (2.4). For the details we refer to Section 2.4 below. The subsequent assumption hints at this additional restriction which Kyle and Obizhaeva (2017a) call "leverage neutrality"

and is quoted from Proposition 1 in the seminal paper by Modigliani and Miller (1958) (see Assumption 2.5 below for a more formal definition).

Assumption 2.2 (Leverage neutrality)**.** *The market value of any firm is independent of its capital structure.*

It turns out that this invariance indeed provides one more linear equation analogous to the equations obtained by the scaling arguments above. We therefore find ourselves in a perfectly analogous situation as with the pendulum and have the same number of equations as unknowns, namely four.

Theorem 2.1. *Under Assumptions 2.1 and 2.2, the market impact is of the form*

$$G = \text{const} \cdot \sigma \sqrt{\frac{Q}{V}}, \tag{2.5}$$

for some constant const > 0.

In particular, we find the square-root dependence of the market impact on the order size Q in accordance with several theoretical as well as empirical findings (see the review of the literature below).

In fact, the above line of argument does not correspond exactly to what Kyle and Obizhaeva (2017a) have done (compare also Kyle and Obizhaeva, 2016, 2017b). They have considered one more variable which may have influence on the market impact. These authors suppose that the agent faces a cost C when preparing the placement of a meta-order, which the authors refer to as "bet cost".[1] This "bet cost" C may vary independently of the order size Q as well as of the quantities P, V and σ^2 discussed above. Hence, they consider an additional fifth quantity which might influence the market impact:

- C the "bet cost", measured in units of money $[C] = \mathbb{U}$.

In other words, Kyle and Obizhaeva (2017a) only use the subsequent hypothesis which is weaker than Assumption 1 above.

Assumption 2.3. *The market impact G depends only on the above 5 variables, i.e.,*

$$G = g(Q, P, V, \sigma^2, C), \tag{2.6}$$

where the function $g : \mathbb{R}_+^5 \to \mathbb{R}_+$ *as well as the quantity G are invariant under changes of the units chosen to measure the "dimensions"* $\mathbb{S}, \mathbb{T}$ *and* $\mathbb{U}$.

[1] In the version by Kyle and Obizhaeva (2017a), released in July 2017, C is defined as the unconditional expected dollar costs of executing a bet, i.e., meta-order.

Starting from this weaker assumption Kyle and Obizhaeva (2017a) apply a similar reasoning as above, in particular the argument of leverage neutrality. This leads to a system of four linear equations in five unknowns. The solution is not unique anymore, but leaves us with a degree of freedom which is expressed by the function f below.

Theorem 2.2 (Kyle and Obizhaeva)**.** *Under Assumptions 2.2 and 2.3, the market impact is of the form*

$$G = \frac{1}{L} f(Z),$$

where $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a function and the quantities L and Z are given by

$$L = \left(\frac{PV}{\sigma^2 C}\right)^{1/3} \quad \text{and} \quad Z = \left(\frac{Q^3 P^2 \sigma^2}{V C^2}\right)^{1/3}. \tag{2.7}$$

A priori, the generality of the function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is not restricted by Assumption 2.3. Specializing further as in the *ansatz* (2.2), one may assume f to be of the form $f(z) = \text{const} \cdot z^p$, for some $p \geq 0$. This implies that $G = \text{const} \cdot Z^p / L$. In particular, the choice $p = 1/2$ leads precisely to the relation (2.5) obtained in Theorem 2.1 above. Other choices of p lead to different relations, some of them already considered in the literature. Moreover, we would like to emphasize that the quantities L and Z have a financially meaningful interpretation in terms of measuring liquidity and the size of meta-orders, see Kyle and Obizhaeva (2017a).

The roadmap of this note is as follows. In Section 2.2, we provide a brief review of the existing literature. Section 2.3 introduces some notation as well as the so-called Pi-Theorem from dimensional analysis, which is the key to rigorously prove Theorem 2.1 and Theorem 2.2 in Section 2.4. Moreover, we discuss in Section 2.4 two additional extentions of Theorem 2.1: we expand the set of explanatory variables given in Assumption 2.1 by including the "spread cost" $\mathscr{C}$ and the length T of the execution interval, respectively. Section 2.5 concludes. Appendix B.1 discusses the example mentioned in the introduction, namely the period of a pendulum, in somewhat more detail, while some proofs are moved to Appendix B.2.

2.2 Literature review

As pointed out in recent reviews by Bouchaud et al. (2009) as well as Foucault et al. (2013), market impact can arise from different sources. For instance, Kyle (1985) in his seminal paper derives from an agent-based model that market impact should be linear in the order size and permanent in time. The majority of studies, however, does not support this conclusion

of Kyle's model. Instead, a body of literature finds market impact being non-linear in the order size and fading in time, e.g. Bouchaud et al. (2009). In particular, the market impact is frequently found to be concave in the size of the meta-order and especially close to the square-root function, which causes the name *square-root law* for market impact (see Bacry et al., 2015; Bershova and Rakhlin, 2013; Brokmann et al., 2015; Engle et al., 2012; Gomes and Waelbroeck, 2015; Mastromatteo et al., 2014; Moro et al., 2009; Tóth et al., 2011). Among other results, a market microstructure foundation in favor of the square-root law is provided by Gabaix et al. (2006). The broad evidence for the square-root law relies on studies having data from different venues, maturities, historical periods and geographical areas and thus provides the square-root law with universality. On the other hand, it deserves to be mentioned that some studies reveal empirically deviations from the square-root law, (e.g. Almgren et al., 2005; Zarinelli et al., 2015).

Let us try to elaborate on the relation between dimensional analysis and a general theory by alluding once more to the analogy with the period of the pendulum. Complementary to the introductory example, relation (2.3) from physics can, of course, also be derived from solving differential equations. Analogously, the square-root law for market impact can also be derived via solving partial differential equations, (see Donier et al., 2015). These authors formulate the dynamics of the average buy and sell volume density of the *latent order book* in terms of partial differential equations under minimal model requirements. While the latent order book is a theoretical concept which records the trading intentions of market participants, traders typically do not display their true supply and demand, so that the fictitious, non-public latent order book differs from the observed limit order book. As we derive the square-root law for market impact via dimensional analysis, Theorem 2.1 complements the existing literature.

2.3 Some linear algebra

To review the basic results of dimensional analysis we follow Chapter 1 of the book by Bluman and Kumei (2013). Additionally, the interested reader is referred to Pobedrya and Georgievskii (2006) for a historical perspective and to Curtis et al. (1982) for a purely mathematical treatment of dimensional analysis. We formalize the assumptions behind dimensional analysis in proper generality. However, for the purpose of the present paper we shall only need the degree of generality covered by Corollaries 2.4 and 2.5 below.

Assumption 2.4 (Dimensional analysis)**.**

(i) Let the quantity of interest $U \in \mathbb{R}_+$ depend on n quantities $W_1, \ldots, W_n \in \mathbb{R}_+$, i.e.,

$$U = H(W_1, W_2, \ldots, W_n), \tag{2.8}$$

for some function $H : \mathbb{R}_+^n \to \mathbb{R}_+$.

(ii) The quantities $U, W_1, \ldots, W_n$ are measured in terms of m fundamental dimensions labelled $L_1, \ldots, L_m$, where $m \leq n$. For any positive quantity X, its dimension $[X]$ satisfies $[X] = L_1^{x_1} \cdots L_m^{x_m}$ for some $x_1, \ldots, x_m \in \mathbb{R}$. If $[X] = 1$, the quantity X is called dimensionless.

The dimensions of the quantities $U, W_1, W_2, \ldots, W_n$ are known and given in the form of vectors a and $b^{(i)} \in \mathbb{R}^m$, $i = 1, \ldots, n$, satisfying $[U] = L_1^{a_1} \cdots L_m^{a_m}$ and $[W_i] = L_1^{b_{1i}} \cdots L_m^{b_{mi}}$, $i = 1, \ldots, n$. Denote by $B = (b^{(1)}, b^{(2)}, \ldots, b^{(n)})$ the $m \times n$ matrix with column vectors $b^{(i)} = (b_{1i}, \ldots, b_{mi})^\top$, $i = 1, \ldots, n$.

(iii) For the given set of fundamental dimensions $L_1, \ldots, L_m$, a system of units *is chosen in order to measure the value of a quantity. A change from one system of units to another amounts to rescaling all considered quantities. In particular, dimensionless quantities remain unchanged and formula* (2.8) *is invariant under arbitrary scaling of the fundamental dimensions.*

We can now state the main result from dimensional analysis (see Bluman and Kumei, 2013).

Theorem 2.3 (Pi-Theorem)**.** *Under Assumption 2.4, let $x^{(i)} := (x_{1i}, \ldots, x_{ni})^\top$, $i = 1, \ldots, k := n - \operatorname{rank}(B)$ be a basis of the solutions to the homogeneous system $Bx = 0$ and $y := (y_1, \ldots, y_n)^\top$ a solution to the inhomogeneous system $By = a$ respectively. Then, there is a function $F : \mathbb{R}_+^k \to \mathbb{R}_+$ such that*

$$U \cdot W_1^{-y_1} \cdots W_n^{-y_n} = F(\pi_1, \ldots, \pi_k),$$

where $\pi_i := W_1^{x_{1i}} \cdots W_n^{x_{ni}}$ are dimensionless quantities, for $i = 1, \ldots, k$.

We shall only need the special cases $k = 0$ and $k = 1$, which are spelled out in the two subsequent corollaries.

Corollary 2.4. *Under Assumption 2.4, suppose that $\operatorname{rank}(B) = n$ and let $y := (y_1, \ldots, y_n)^\top$ be the unique solution to the linear system $By = a$. Then there is a constant $\text{const} > 0$ such that*

$$U = \text{const} \cdot W_1^{y_1} \cdots W_n^{y_n}.$$

Corollary 2.5. *Under Assumption 2.4, suppose that* $\operatorname{rank}(B) = n-1$ *and define by* $x := (x_1,\dots,x_n)^\top$ *and* $y := (y_1,\dots,y_n)^\top$ *the non-trivial solutions to the homogeneous and inhomogeneous systems* $Bx = 0$ *and* $By = a$ *respectively. Then there is a function* $F : \mathbb{R}_+ \to \mathbb{R}_+$ *such that*

$$U = F(W_1^{x_1}\cdots W_n^{x_n})W_1^{y_1}\cdots W_n^{y_n}.$$

2.4 Market impact

The aim of this section is to formalize and prove Theorems 2.1 and 2.2 stated in the introduction by applying Corollaries 2.4 and 2.5. In order to derive the market impact function from these corollaries, we need to formalize Assumption 2.2 in the framework of Section 2.3. Therefore, we take a closer look at this assumption and hence, the behavior of the quantities G, Q, P, V, σ^2 and C in case of changing the firm's leverage defined below. From a conceptual point of view, the assumption of leverage neutrality gives an additional constraint on their behavior. This constraint can be understood as an additional though synthetic dimension in our analysis, which we refer to as the Modigliani-Miller "dimension" $\mathbb{M}$. The Modigliani-Miller dimension $\mathbb{M}$ of a share of a company can be measured in terms of the leverage $\mathscr{L}$, i.e., the quantity

$$\mathscr{L} = \frac{\text{total assets}}{\text{equity}}.$$

Multiplying $\mathscr{L}$ by a factor $A > 1$ is equivalent to paying out $(1 - A^{-1})$ of the equity as cash-dividends. On the other hand, multiplying $\mathscr{L}$ by a factor $0 < A < 1$ corresponds to raising new capital in order to increase the firm's own capital by $(A^{-1} - 1)$ times its equity. Following Kyle and Obizhaeva (2017a), Assumption 2.2 can be reformulated in the following way:

Assumption 2.5 (Leverage neutrality)**.** *Scaling the Modigliani-Miller "dimension"* $\mathbb{M}$ *by a factor* $A \in \mathbb{R}_+$ *implies that*

- *Q, V and C remain constant,*
- *P changes by a factor* A^{-1}*,*
- σ^2 *changes by a factor* A^2*,*
- *G changes by a factor A.*

To recapitulate in prose: Setting $A = 2$ corresponds to paying out half of the equity as dividends in the sense that each share yields a dividend of $(1 - A^{-1})P = P/2$. The stock price, thus, is multiplied by $A^{-1} = 1/2$ while the volatility σ and the percentage market impact G are multiplied by $A = 2$. The remaining quantities are not affected by changing the leverage, in accordance with the insight of Modigliani and Miller (1958) and the recent work by Kyle and Obizhaeva (2017a). As indicated in the introduction, this assumption is referred to as *leverage neutrality*.

We now reformulate Theorem 2.1 by replacing the informally stated Assumption 2.2 by the more formal Assumption 2.5 and provide a proof.

Theorem 2.6. *Under Assumptions 2.1 and 2.5, the market impact is of the form*

$$G = \text{const} \cdot \sigma \sqrt{\frac{Q}{V}}, \tag{2.9}$$

for some constant const > 0.

Proof. Combining Assumptions 2.1 and 2.5 with the dimensions of Q, P, V and σ^2 introduced in Section 2.1, we obtain that the matrix B and the vector a are given by

$$B = \begin{pmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & -1 & 0 & 2 \end{pmatrix} \quad \text{and} \quad a = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{2.10}$$

Table 2.1 summarizes how B and a can be derived and should be read as follows: Assume the measurement of dimension $\mathbb{S}$ referring to the unit of shares is rescaled by a factor S, then Q changes by a factor S, P changes by a factor S^{-1}, V changes by a factor S, while σ^2 does not change. Likewise, the last row labelled by $\mathbb{M}$ indicates that if the leverage $\mathscr{L}$ of the firm changes by a factor A, Q does not change, P changes by a factor A^{-1}, and so on.

As the matrix B has full rank, i.e., $\operatorname{rank}(B) = 4 = n$, and Assumption 2.4 is satisfied, applying Corollary 2.4 yields

$$G = \text{const} \cdot Q^{y_1} P^{y_2} V^{y_3} \sigma^{2y_4},$$

for some constant const> 0, where $y = (y_1, y_2, y_3, y_4)^\top$ is the unique solution of the linear system $By = a$ which is given by $y = (1/2, 0, -1/2, 1/2)^\top$. □

	Q	P	V	σ^2	C	G
$\mathbb{S}$	1	-1	1	0	0	0
$\mathbb{U}$	0	1	0	0	1	0
$\mathbb{T}$	0	0	-1	-1	0	0
$\mathbb{M}$	0	-1	0	2	0	1

Table 2.1 A labelled overview of the matrix B related to the dimensions of the quantities (Q,P,V,σ^2) and the matrix K related to the dimensions of the quantities (Q,P,V,σ^2,C) respectively, as well as the vector a related to the dimensions of G.

Theorem 2.6 implies the well known *square-root law* for market impact. We would like to highlight that the present derivation of the square-root law does not rely on economic, empirical or theoretical assumptions except the dependence of G on Q,P,V and σ^2 only, as well as leverage neutrality. Donier et al. (2015) present an alternative derivation of the square-root law relying on partial differential equations.

As discussed above, Kyle and Obizhaeva (2017a) consider yet another variable to influence the market impact, namely the "bet cost" C leading to the weaker Assumption 2.3. An economic motivation to include C in the analysis is provided also by Kyle and Obizhaeva (2017b). Based on Assumption 2.3, Kyle and Obizhaeva (2017a) derive a more general result summarized in Theorem 2.2. The methodology of Section 2.3 can be employed to prove a similar result stated below.

First of all, the matrix B used in the proof of Theorem 2.6 is extended by one column, which corresponds to C, to obtain the matrix

$$K = \begin{pmatrix} 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & -1 & -1 & 0 \\ 0 & -1 & 0 & 2 & 0 \end{pmatrix}. \tag{2.11}$$

The vector $a = (0,0,0,1)^\top$ defined in (2.10) remains unchanged. Table 2.1 illustrates how the additional variable C is related to the considered "dimensions".

Let us compute the solution space $\mathscr{H}$ of the homogeneous system $Kx = 0$, which is given by the kernel of the linear map induced by the matrix K, as well as the solution space $\mathscr{I}$ of

the inhomogeneous linear system $Ky = a$:

$$\mathscr{H} = \left\{ \lambda \begin{pmatrix} 3 \\ 2 \\ -1 \\ 1 \\ -2 \end{pmatrix}, \lambda \in \mathbb{R} \right\} \quad \text{and} \quad \mathscr{I} = \left\{ \begin{pmatrix} -1 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 3 \\ 2 \\ -1 \\ 1 \\ -2 \end{pmatrix}, \lambda \in \mathbb{R} \right\}.$$

Theorem 2.7. *Suppose Assumptions 2.3 and 2.5 hold. Fix* $x = (x_1, \dots, x_5)^\top \in \mathscr{H}$ *and* $y = (y_1, \dots, y_5)^\top \in \mathscr{I}$. *There is a function* $f : \mathbb{R}_+ \to \mathbb{R}_+$ *such that*

$$G = Q^{y_1} P^{y_2} V^{y_3} \sigma^{2y_4} C^{y_5} f\left(Q^{x_1} P^{x_2} V^{x_3} \sigma^{2x_4} C^{x_5}\right). \tag{2.12}$$

Proof. Combining Assumptions 2.3 and 2.5 with the dimensions of Q, P, V, σ^2 and C introduced in Section 2.1, we recover the matrix K and the vector a given in (2.11) and (2.10) respectively. Since Assumption 2.4 is satisfied and $\operatorname{rank}(K) = 4$, applying Corollary 2.5 completes the proof. □

For example, by setting

$$x = \left(1, \frac{2}{3}, -\frac{1}{3}, \frac{1}{3}, -\frac{2}{3}\right)^\top \in \mathscr{H} \qquad \text{and} \qquad y = \left(0, -\frac{1}{3}, -\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^\top \in \mathscr{I},$$

Theorem 2.7 yields precisely the formula of Theorem 2.2 given in the introduction, i.e.,

$$\begin{aligned} G &= \left(\frac{\sigma^2 C}{PV}\right)^{1/3} f\left(\left(\frac{Q^3 P^2 \sigma^2}{VC^2}\right)^{1/3}\right) \\ &= \frac{1}{L} f(Z), \end{aligned} \tag{2.13}$$

where L and Z are defined in (2.7). One may also consider other choices for $x \in \mathscr{H}$ and $y \in \mathscr{I}$, for example:

$$x = (3, 2, -1, 1, -2)^\top \in \mathscr{H} \qquad \text{and} \qquad y = \left(\frac{1}{2}, 0, -\frac{1}{2}, \frac{1}{2}, 0\right)^\top \in \mathscr{I}.$$

Formula (2.12) then takes the form

$$G = \sigma \sqrt{\frac{Q}{V}}\, h\left(Z^3\right) = \sigma \sqrt{\frac{Q}{V}}\, h\left(\frac{Q^3 P^2 \sigma^2}{VC^2}\right), \tag{2.14}$$

for some function $h : \mathbb{R}_+ \to \mathbb{R}_+$. If the function h in (2.14) is not a constant, this formula describes nicely the deviation from the square-root law (2.9) in a multiplicative way.

Remark 2.1. *It is important to note that* (2.13) *as well as* (2.14) *are both* the general solution *of the functional relation described by Theorem 2.7 and therefore coincide. The difference is that the (arbitrary) functions f and h are not identical, but rather in a one-to-one relation when passing from* (2.13) *to* (2.14).

As pointed out by Kyle and Obizhaeva (2017a), different choices of f in equation (2.13) (resp. h in (2.14)) lead to some particularly relevant market impact models studied in the literature.

(a) The proportional market impact: $f \equiv \text{const}$ (resp. $h(x) = \text{const} \cdot x^{-1/6}$) leads to

$$G = \text{const} \cdot \left(\frac{\sigma^2 C}{PV} \right)^{1/3}.$$

(b) The square-root impact: $f(z) = \text{const} \cdot z^{1/2}$ (resp. $h \equiv \text{const}$) leads to

$$G = \text{const} \cdot \sigma \sqrt{\frac{Q}{V}},$$

the unique solution which does not depend on C.

(c) The linear market impact: $f(z) = \text{const} \cdot z$ (resp. $h(x) = \text{const} \cdot x^{1/6}$) leads to

$$G = \text{const} \cdot Q \left(\frac{\sigma^4 P}{CV^2} \right)^{1/3}.$$

Remark 2.2. *It is also important to note that only two properties of the variable C enter the above dimensional analysis: the "dimension" of C is money, i.e., $[C] = \mathbb{U}$, and C remains unchanged by scaling the Modigliani-Miller "dimension" $\mathbb{M}$ by a factor $A \in \mathbb{R}_+$. The above result, therefore, does not rely on the interpretation of the quantity C as "bet cost" as considered in Kyle and Obizhaeva (2017a), but applies to any other quantity with the two aforementioned properties just as well.*

For example, an interesting alternative to C enjoying these properties can be found in the work on the intraday trading invariance hypothesis by Benzaquen et al. (2016). Rather than C, these authors consider the spread cost $\mathscr{C}$, which can be interpreted as the transaction cost incurred by trading Q shares. More formally, denote by S the bid-ask spread measured in units

of money per share $[S] = \mathbb{U}/\mathbb{S}$.[2] The spread cost $\mathscr{C}$ of a meta-order with size Q is then defined by $\mathscr{C} := QS$ and hence measured in units of money $[\mathscr{C}] = \mathbb{U}$. Thus, the mathematical analysis above remains *totally unchanged* when C is replaced by $\mathscr{C}$. In particular, the analogue of formula (2.14) then reads as

$$G = \sqrt{\frac{Q}{V}}\, h\left(\frac{Q^3 P^2 \sigma^2}{V \mathscr{C}^2}\right). \tag{2.15}$$

To finish this section, we shall consider one more possible set of 5 explanatory variables which will lead us into a somewhat different direction. Instead of C (or any appropriate alternative such as $\mathscr{C}$), we consider a variable with a different dimension, namely the length of the time interval $[0,T]$ over which the meta-order is executed. Clearly, the length T is an obvious candidate to influence the market impact:

- T the length of the execution interval, measured in units of time $[T] = \mathbb{T}$.

In practice, the interval length T can vary from a fraction of hours up to several days or even weeks. In an analogous manner as the "bet cost" C enters Assumption 2.3, the subsequent assumption incorporates the length of the execution interval.

Assumption 2.6. *The market impact G depends only on the variables Q, P, V, σ^2 and T, i.e.,*

$$G = g(Q,P,V,\sigma^2,T), \tag{2.16}$$

where the function $g : \mathbb{R}_+^5 \to \mathbb{R}_+$ as well as the quantity G are invariant under changes of the units chosen to measure the "dimensions" $\mathbb{S}, \mathbb{T}$ and $\mathbb{U}$.

Playing a similar game as above, the matrix B used in the proof of Theorem 2.6 is extended by one column corresponding to T and now given by

$$K = \begin{pmatrix} 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 1 \\ 0 & -1 & 0 & 2 & 0 \end{pmatrix}.$$

[2] It should be noticed that the spread S remains unchanged when scaling the Modigliani-Miller dimension $\mathbb{M}$ by a factor $A \in \mathbb{R}_+$. For instance, this can be inferred from an argument of (Kyle and Obizhaeva, 2017a, Section 3), whose empirical analysis uses that G and S/P have the same dimensional properties. Since the concept of leverage neutrality tells us precisely how G and P change when $\mathbb{M}$ is scaled by a factor $A \in \mathbb{R}_+$, the spread S has to remain unchanged.

The vector a given in (2.10) remains unchanged. The solution spaces $\mathscr{H}$ and $\mathscr{I}$ of the homogeneous system $Kx = 0$ and the inhomogeneous system $Ky = a$ respectively, are given by

$$\mathscr{H} = \left\{ \lambda \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \lambda \in \mathbb{R} \right\} \quad \text{and} \quad \mathscr{I} = \left\{ \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ -1 \end{pmatrix}, \lambda \in \mathbb{R} \right\}.$$

Under the assumptions of leverage neutrality and the exclusive dependence of the market impact G on Q, P, V, σ^2 and T, dimensional analysis leads to the following result.

Theorem 2.8. *Suppose Assumptions 2.5 and 2.6 hold. Fix $x = (x_1, \dots, x_5)^\top \in \mathscr{H}$ and $y = (y_1, \dots, y_5)^\top \in \mathscr{I}$. There is a function $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that*

$$G = Q^{y_1} P^{y_2} V^{y_3} \sigma^{2y_4} T^{y_5} f\left(Q^{x_1} P^{x_2} V^{x_3} \sigma^{2x_4} T^{x_5}\right).$$

Proof. Since Assumption 2.4 is satisfied and $\operatorname{rank}(K) = 4$, the result follows by Corollary 2.5. □

For instance, setting

$$x = (1, 0, -1, 0, -1)^\top \in \mathscr{H} \qquad \text{and} \qquad y = \left(\frac{1}{2}, 0, -\frac{1}{2}, \frac{1}{2}, 0\right)^\top \in \mathscr{I},$$

we obtain

$$G = \sigma \sqrt{\frac{Q}{V}}\, h\left(\frac{Q}{VT}\right), \tag{2.17}$$

for some function $h : \mathbb{R}_+ \to \mathbb{R}_+$. Similar to (2.14), where the "bet cost" C appears as a quantity influencing the deviation of (2.14) from (2.9), the function h in (2.17) characterizes the deviation from the square-root law (2.9) in dependence of the length of the execution interval T.

Donier et al. (2015) derive the square-root law based on a model taking the execution horizon T into account. Thus, the question arises under which conditions on T we recover the square-root law (2.9). The answer is simple: If the length of the execution interval T depends exclusively on either or all of the quantities Q, V, P and σ^2, Assumption 2.6 can be replaced by Assumption 2.1 and we are back in the setting of Theorem 2.1, where the market impact obeys the square-root law. In practice, the condition that T depends on Q, P, V and

σ^2 can be satisfied in case the investor determines the execution horizon T according to the latter quantities.

2.5 Conclusion

The main contribution of this paper is a derivation of the *square-root law* for market impact. The strong empirical support in favor of this *law* provides it with a universal character. Inspiring for our work, Kyle and Obizhaeva (2017a) derive a general form for the market impact function relying on dimensional analysis as well as the concepts of leverage neutrality and market microstructure invariance, where the *square-root law* turns out to be a special case of their result. Complementary to their approach, we present a direct and simple derivation of the *square-root law* by requiring only two assumptions: Firstly, the market impact of a given meta-order *only* depends on its size, the corresponding stock price, the traded volume in the stock as well as its volatility. Secondly, we employ the concept of leverage neutrality as done by Kyle and Obizhaeva (2017a). This idea is in line with the Modigliani-Miller invariance principle (see Modigliani and Miller, 1958) and explains how the considered quantities behave when changing the leverage of a firm. Relying on these plausible assumptions, we apply dimensional analysis in a rigorous way to show that the market impact of a meta-order is proportional to the volatility as well as to the square-root of this order's size and inversely proportional to the square-root of the traded volume.

We also discuss several extensions of this result by including the following quantities as additional explanatory variables: the "bet cost" C (2.14) like Kyle and Obizhaeva (2017a), the "spread cost" $\mathscr{C}$ (2.15) like Benzaquen et al. (2016), or the length T of the execution interval (2.17).

CHAPTER 3

THEORETICAL AND EMPIRICAL ANALYSIS OF TRADING ACTIVITY

JOINT WORK WITH A. RISTIG, W. SCHACHERMAYER AND L. TANGPI

ABSTRACT.

Understanding the structure of financial markets deals with suitably determining the functional relation between financial variables. In this respect, important variables are the trading activity, defined here as the number of trades N, the traded volume V, the asset price P, the squared volatility σ^2, the bid-ask spread S and the cost of trading C. Different reasonings result in simple proportionality relations ("scaling laws") between these variables. A basic proportionality is established between the trading activity and the squared volatility, i.e., $N \sim \sigma^2$. More sophisticated relations are the so called 3/2-law $N^{3/2} \sim \sigma PV/C$ and the intriguing scaling $N \sim (\sigma P/S)^2$. We prove that these "scaling laws" are the only possible relations for considered sets of variables by means of a well-known argument from physics: dimensional analysis. Moreover, we provide empirical evidence based on data from the NASDAQ stock exchange showing that the sophisticated relations hold with a certain degree of universality. Finally, we discuss the time scaling of the volatility σ, which turns out to be more subtle than one might naively expect.

3.1 Introduction

Understanding the structure of financial markets is of obvious relevance for traders, investors and regulators. Among others, the relation between trading activity and price variability received a lot of attention in the financial literature over the last five decades. The pioneers of this field, e.g Clark (1973), Epps and Epps (1976) and Tauchen and Pitts (1983), defined trading activity via trading volume and derived a proportionality relation between the trading volume and the price variability. The rationale behind this definition and the implied relation is the widely-cited aphorism, "it takes volume to move prices". We refer to Karpoff (1987) for a survey of these early works on the *price-volume relation*.

Due to minor empirical evidence for the hypotheses developed in these early approaches, the volume-based definition of trading activity has been replaced by the number of trades. This definition is caused by the substantial link between the observed price variability and the number of trades (see Ané and Geman, 2000; Dufour and Engle, 2000; Jones et al., 1994). For example, Jones et al. (1994) find no predictive power in the volume for the price variability but that the number of trades scales proportionally to the squared volatility. This scaling relation will be the starting point of our discussion. Building on the aforementioned ideas numerous other studies followed, e.g. Andersen (1996) and Liesenfeld (2001). In particular, let us point out the contribution by Wyart et al. (2008), who argue that the price volatility per trade, i.e., (price) $\times$ (volatility) $\times$ (number of trades)$^{-1/2}$, is proportional to the bid-ask-spread. This connection can be seen as a somewhat refined version of the relation proposed by Jones et al. (1994).

More recently, general relations between financial quantities have been derived based on the invariance of markets' microstructure, see Kyle and Obizhaeva (2016). In particular, the authors postulate a *trading invariance principle* which (in contrast to the above relations) is formulated on the latent level of *meta-orders*.[1] Andersen et al. (2016) and Benzaquen et al. (2016) confirm empirically that an analogue of this invariance principle holds true for intradaily observable quantities. The fundamental relation may then be formulated as follows: the nominal value of the exchanged risk during a period of time, defined as the product (volatility) $\times$ (traded volume) $\times$ (price), is proportional to the number of trades to the power $3/2$. This so called *intraday trading invariance principle* and its connection to the relations proposed by Jones et al. (1994) and Wyart et al. (2008) is the focus of the present paper.

[1] A meta-order, also referred to as *bet*, is a collection of trades originating from the same trading decision of a single investor.

Our aim is to critically analyze these three relations as well as variants thereof by applying a method well known from physics: dimensional analysis. It is a tool which allows for the *falsification* of a proposed relation, e.g. of the above mentioned formulas for the number of trades, but not for its *verification*. This principle is similar in spirit to K. Popper's approach to epistemology which in turn is inspired by the classical theory of statistics: There one can possibly reject a null hypothesis, but never prove it. Similarly, dimensional analysis can only isolate those functional relations between variables involving certain "dimensions" which do not violate the obvious scaling invariance of these dimensions. Hence, it a priori rules out those functional relations which are in conflict with these scaling requirements. But this does *not* imply that the identified functional relations, which are in accordance with the scaling requirements, describe the reality in a reasonable way. This has to be confirmed by other methods. In the present setting the ultimate challenge is, of course, to fit to empirical data. To complete the picture, we perform an empirical analysis of the relations described above and show that the *intraday trading invariance principle* provides an appropriate fit to empirical data, but fails to be a "universal law".

In dimensional analysis one uses the rather obvious argument that a meaningful relation between quantities involving some "dimensions" should not be affected by the units in which these "dimensions" are measured. In the present context the relevant "dimensions" are time, shares, and money, denoted as $\mathbb{T}, \mathbb{S}$ and $\mathbb{U}$, respectively. We shall also use an additional argument, namely "leverage neutrality" as introduced by Kyle and Obizhaeva (2017a). We emphasize that these authors were the first to combine the concepts of "leverage neutrality" and dimensional analysis. The assumption of leverage neutrality is based on the Modigliani-Miller theorem (see Modigliani and Miller, 1958) and leads to a scaling invariance principle which, mathematically speaking, is perfectly analogous to the dimensional scaling requirements mentioned above.

The remainder of the paper is structured as follows. In Section 3.2, we first deduce the proportionality between the number of trades and the price variability as proposed by Jones et al. (1994) from dimensional arguments. Next, we derive the more involved scaling relations proposed by Benzaquen et al. (2016) as well as Wyart et al. (2008), again using dimensional analysis, and discuss the assumption of leverage neutrality in this context. Having a theoretical foundation for the discussed relations, we then turn to the empirical analysis in Section 3.3: Based on data from the NASDAQ stock market, we show that the relation proposed by Benzaquen et al. (2016) fits the data rather well. In Section 3.4, we take a closer look at volatility and analyze implications of different time scalings thereof. We conclude with some empirical results in this respect. A reminder on the Pi-theorem

from dimensional analysis as well as proofs for all considered relations can be found in the appendix.

3.2 The trading invariance principle

We are interested in explaining the arrival rate of trades in a given stock measured as

- $N = N_t^{t+T}$ the number of trades within a fixed time interval $[t, t+T]$ so that N is measured per units of time. Following the notation from Pohl et al. (2017), this link between the variable N and its dimensional unit is therefore given by

$$[N] = \mathbb{T}^{-1}.$$

Let us identify the variables (and their dimensions $[\cdot]$) which are likely to influence the number of trades N in a given interval $[t, t+T]$. Three obvious candidates are:

- $V = V_t^{t+T}$ the traded volume of the stock during the time interval $[t, t+T]$, measured in units of shares per time

$$[V] = \mathbb{S}/\mathbb{T}.$$

- $P = P_t^{t+T}$ the average price of the stock in the interval $[t, t+T]$, measured in units of money per share

$$[P] = \mathbb{U}/\mathbb{S}.$$

- $\sigma^2 = (\sigma^2)_t^{t+T} = \mathbb{V}\mathrm{ar}\left(\log(P_{t+T}) - \log(P_t)\right)$ the variance of the log-price over the time interval $[t, t+T]$. We assume

$$[\sigma^2] = \mathbb{T}^{-1}.$$

If the price process $(P_t)_{t\geq 0}$ follows, e.g. the Black-Scholes model, see (3.24), we clearly find the above scaling $[\sigma^2] = \mathbb{T}^{-1}$ and shall retain this assumption in most of the paper. However, the scaling of σ^2 turns out to be more subtle than it seems at first glance. In Section 3.4 below, we shall investigate the implications of a scaling relation $[\sigma^2] = \mathbb{T}^{-2H}$, where $H \in (0,1)$ may be different from $1/2$. For instance, such a scaling may result from price processes based on a fractional Brownian motion $(B_t^H)_{t\geq 0}$ with Hurst parameter $H \in (0,1)$, see Mandelbrot and Ness (1968).

Based on these identified dimensions, let us turn to the basic idea of dimensional analysis: the validity of a considered relation should not depend on whether we measure time $\mathbb{T}$ in seconds or in minutes, shares $\mathbb{S}$ in single shares or in packages of hundred shares, and money $\mathbb{U}$ in Euros or in Euro-cents.

Defintion 1 (Dimenisonal invariance)**.** *A function $h : \mathbb{R}_+^n \to \mathbb{R}_+$ relating the quantity of interest U to the explanatory variables $W_1, \ldots, W_n$, i.e,*

$$U = h(W_1, \ldots, W_n),$$

is called dimensionally invariant *if it is invariant under rescaling the involved dimensions (in our case $\mathbb{S}, \mathbb{T}$ and $\mathbb{U}$).*

As a first - and rather naive - approach we analyze the assumption that the three variables σ^2, P and V *fully* explain the number of trades N.

Proposition 3.1. *Assume that the number of trades N depends* only *on the three quantities σ^2, P and V, i.e.,*

$$N = g(\sigma^2, P, V), \tag{3.1}$$

where the function $g : \mathbb{R}_+^3 \to \mathbb{R}_+$ is dimensionally invariant*. Then, there is a constant $c > 0$ such that the number of trades N obeys the relation*

$$N = c \cdot \sigma^2. \tag{3.2}$$

The proof relies on elementary linear algebra and is given in Appendix B.3 below (compare also the proof of Theorem 1 below which is similar). Recall that relation (3.2) goes back to Jones et al. (1994).

As mentioned in the introduction, one should read the present "dimensional" argument in favor of relation (3.2) as a pure "if…then…" assertion: **if** N really is fully explained by σ^2, P and V **and** the obvious scaling invariances of $\mathbb{S}$, $\mathbb{T}$ and $\mathbb{U}$ are satisfied, **then** (3.2) is the only possible relation. As we shall see below, the empirical data does not reconfirm the validity of (3.2). In other words, we have to turn the above statement upside down: as (3.2) is not reconfirmed by empirical data, the variables σ^2, P and V cannot fully explain the quantity N. It is therefore natural to introduce more/other quantities in order to explain the number of trades N.

Regarding the uniqueness of the function g in (3.1), the mathematical reason for the unique choice of g given by (3.2) is that we have three scaling relations (pertaining to the

invariance of the "dimensions" $\mathbb{S}, \mathbb{U}$ and $\mathbb{T}$) as well as the three explanatory variables σ^2, P and V. This leads to three linear equations in three unknowns, yielding a unique solution.

Let us now try to go beyond the scope of relation (3.1) by considering further explanatory variables. Motivated by Wyart et al. (2008), we consider the following quantity as relevant for the number of trades N in a given interval $[t, t+T]$, additionally to σ^2, P and V:

- $S = S_t^{t+T}$ the average bid-ask spread in the interval $[t, t+T]$, measured in units of money per share
$$[S] = \mathbb{U}/\mathbb{S}.$$

Following Benzaquen et al. (2016), it is also convenient to alternatively consider the quantity

- $C = C_t^{t+T}$ the average cost per trade in the interval $[t, t+T]$, measured in units of money
$$[C] = \mathbb{U}.$$

To visualize things, suppose that for some stock we observe in average during the time interval $[t, t+T]$ an ask price of EUR12.30 and a bid price of EUR12.20 so that the bid-ask spread S equals 10 cents. If the average trade size in the interval $[t, t+T]$, denoted by $Q = Q_t^{t+T}$, is 500 shares, we obtain that the average cost per trade $C = QS$ is EUR50. A discussion of the difference between using S rather than C as an explanatory variable can be found at the end of this section. For now, let us follow Benzaquen et al. (2016) for our derivation of the *intraday trading invariance principle* and pass to the set σ^2, P, V and C of explanatory variables, i.e.,

$$N = g(\sigma^2, P, V, C), \tag{3.3}$$

for some function $g : \mathbb{R}_+^4 \to \mathbb{R}_+$. As we now have four explanatory variables, the three equations yielded by the scale invariance of the dimensions $\mathbb{S}, \mathbb{U}$ and $\mathbb{T}$ are not sufficient anymore to imply an (essentially) unique solution for g. In fact, the four explanatory variables above combined with the three invariance relations pertaining to $\mathbb{S}$, $\mathbb{T}$ and $\mathbb{U}$ only yield a general solution of (3.3) of the form

$$N = \sigma^2 f\left(\frac{PV}{\sigma^2 C}\right), \tag{3.4}$$

where $f : \mathbb{R}_+ \to \mathbb{R}_+$ is an arbitrary function whose generality cannot be restricted by only relying on arguments pertaining to dimensional analysis with respect to the three dimensions

$\mathbb{S}$, $\mathbb{T}$ and $\mathbb{U}$ (see Appendix B.3).

Hence, in order to obtain such a crisp result as in (3.2), an additional "dimensional invariance" is required. Kyle and Obizhaeva (2017a) found a remedy: a no-arbitrage type argument, referred to as "leverage neutrality".[2] This concept is inspired by the findings of Modigliani and Miller (1958) (compare Pohl et al., 2017): Consider a stock of a company, and suppose that the company changes its capital structure by paying dividends or by raising new capital. The Modigliani-Miller theorem tells us precisely which features of the company are *not affected* by a change in the capital structure. This allows us to establish how certain quantities behave when varying the leverage in terms of the relation between debt and equity of a company.

From a conceptual point of view, the assumption of leverage neutrality gives a constraint on the behavior of the quantities N, σ^2, P, V, C (resp. S) in case of changing the firm's capital structure. This constraint can be understood as an additional though synthetic dimension in our analysis, which we refer to as the Modigliani-Miller "dimension" $\mathbb{M}$. The Modigliani-Miller "dimension" $\mathbb{M}$ of a share of a company is measured in terms of the leverage $\mathscr{L}$, i.e., the quantity

$$\mathscr{L} = \frac{\text{total assets}}{\text{equity}}.$$

Multiplying $\mathscr{L}$ by a factor $A > 1$ is equivalent to paying out $(1 - A^{-1})$ of the equity as cash-dividends. On the other hand, multiplying $\mathscr{L}$ by a factor $0 < A < 1$ corresponds to raising new capital in order to increase the firm's equity by a factor A^{-1}. Following Kyle and Obizhaeva (2017a) as well as Pohl et al. (2017), we are led to the following assumption:

Leverage Neutrality Assumption (Kyle and Obizhaeva (2017a); Pohl et al. (2017))**.** *Scaling the Modigliani-Miller "dimension" $\mathbb{M}$ by a factor $A \in \mathbb{R}_+$ implies that*

- *N, V and C (as well as S) remain constant,*
- *P changes by a factor A^{-1},*
- *σ^2 changes by a factor A^2.*

To recapitulate: Setting $A = 2$ corresponds to paying out half of the equity as dividends so that each share yields a dividend of $(1 - A^{-1})P = P/2$. The stock price is, thus, multiplied by $A^{-1} = 1/2$ while the volatility σ is multiplied by $A = 2$. The remaining quantities are not

[2] Note that Kyle and Obizhaeva (2017a) use the argument of leverage neutrality in the context of market impact. But, of course, the same idea applies in the present situation.

affected by changing the leverage, in accordance with the insight of Modigliani and Miller (1958) and the recent work by Kyle and Obizhaeva (2017a). The economic reason is that the value of the assets of the corresponding company and hence the associated risk does not change.

Defintion 2 (Leverage neutrality)**.** *A function $h : \mathbb{R}_+^n \to \mathbb{R}_+$ relating the quantity N to the explanatory variables σ^2, P, V, C and S, i.e,*

$$N = h(\sigma^2, P, V, C, S),$$

is called leverage neutral *if it is invariant when rescaling the Modigliani-Miller dimension $\mathbb{M}$ of the variables N, σ^2, P, V, C, S as defined in the assumption above.*

We can now derive the following relation, which is the focus of the present paper. It relies on the basic fact that under the "Leverage Neutrality Assumption" we now find four linear equations in order to determine four unknowns. Note that Benzaquen et al. (2016) coined this relation the "3/2-law".

Theorem 3.2 ($(3/2)$-law)**.** *Suppose the "Leverage Neutrality Assumption" holds and that the number of trades N depends* only *on the four quantities σ^2, P, V and C, i.e.,*

$$N = g(\sigma^2, P, V, C), \tag{3.5}$$

where the function $g : \mathbb{R}_+^4 \to \mathbb{R}_+$ is dimensionally invariant *and* leverage neutral*. Then, there is a constant $c > 0$ such that the number of trades N obeys the relation*

$$N^{3/2} = c \cdot \frac{\sigma P V}{C}. \tag{3.6}$$

The proof follows from the general Pi-theorem reviewed in Appendix B.3. For the convenience of the reader, we also present a direct proof of Theorem 3.2. Although slightly longish and repetitive, we hope that it helps the intuition.

Proof of Theorem 3.2. First, we make the following *ansatz* for the function g in (3.5):

$$g(\sigma^2, P, V, C) = c \cdot (\sigma^2)^{y_1} P^{y_2} V^{y_3} C^{y_4}, \tag{3.7}$$

where $c > 0$ is a constant and $y_1, \ldots, y_4$ are unknown real numbers. Looking at the first row of Table 3.1 yields the relation

$$-y_2 + y_3 = 0. \tag{3.8}$$

Indeed, when passing from counting shares in packages of 100 units rather than in single units, the number P is replaced by $100P$ while the number V is replaced by $V/100$. Since the function g in (3.7) is assumed to be dimensionally invariant, g should remain unchanged by this passage, i.e.,

$$c \cdot \left(\sigma^2\right)^{y_1} P^{y_2} V^{y_3} C^{y_4} = c \cdot \left(\sigma^2\right)^{y_1} (100P)^{y_2} \left(\frac{V}{100}\right)^{y_3} C^{y_4} \tag{3.9}$$

which is only possible if (3.8) holds true. Looking at the other rows of Table 3.1 we therefore get the system of linear equations

$$\begin{cases} -y_2 + y_3 = 0 \\ y_2 + y_4 = 0 \\ -y_1 - y_3 = -1 \\ 2y_1 - y_2 = 0 \end{cases}$$

whose unique solution is

$$y = \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}, -\frac{2}{3}\right)^\top, \tag{3.10}$$

which gives (3.6) as one possible solution of (3.5).

We still have to show the uniqueness of (3.6). To do so, it is convenient to pass to logarithmic coordinates: suppose that there is a function $G : \mathbb{R}^4 \to \mathbb{R}$ such that $\log(N) = G\left(\log(\sigma^2), \log(P), \log(V), \log(C)\right)$ or equivalently,

$$\log(N) - G(X_1, X_2, X_3, X_4) = 0, \tag{3.11}$$

where we write $\left(\log(\sigma^2), \log(P), \log(V), \log(C)\right)$ as (X_1, X_2, X_3, X_4). We have to show that G has the form

$$\log(N) = y_1 X_1 + y_2 X_2 + y_3 X_3 + y_4 X_4 + \text{const},$$

where y_1, y_2, y_3, y_4 are given by (3.10) and const is a real number. Denote by $r_1 := -e_2 + e_3$ the first row of Table 3.1, considered as a vector in $\mathbb{R}^4$, where $(e_i)_{i=1}^4$ is the canonical basis of $\mathbb{R}^4$. Similarly as in (3.9), the first row of Table 3.1 and dimensional invariance imply that

$$\begin{aligned} &G\left(\log(\sigma^2), \log(P), \log(V), \log(C)\right) \\ &\quad = G\left(\log(\sigma^2), \log(P) + \log(100), \log(V) - \log(100), \log(C)\right). \end{aligned}$$

	σ^2	P	V	C	N
$\mathbb{S}$	0	-1	1	0	0
$\mathbb{U}$	0	1	0	1	0
$\mathbb{T}$	-1	0	-1	0	-1
$\mathbb{M}$	2	-1	0	0	0

Table 3.1 A labelled overview of the dimensions of the quantities P, V, σ^2 and C.

Clearly we can replace $\log(100)$ by any real number. Speaking abstractly, this means that $G : \mathbb{R}^4 \to \mathbb{R}$ must be constant on any straight line parallel to the vector r_1. A similar argument applies to $r_2 = e_2 + e_4$ and $r_4 = 2e_1 - e_2$. As regard $r_3 = -e_1 - e_3$ the situation is slightly different, as the third row of Table 3.1 also involves a non-zero entry of N.

The third row of Table 3.1 and (3.11) imply that for any $\lambda \in \mathbb{R}$,

$$G(X_1 - \lambda, X_2, X_3 - \lambda, X_4) = G(X_1, X_2, X_3, X_4) - \lambda.$$

Setting const $:= G(0,0,0,0)$, we have

$$G(-\lambda, 0, -\lambda, 0) = -\lambda + \text{const} \quad \text{for all } \lambda \in \mathbb{R},$$

which uniquely determines G on the one-dimensional space spanned by $r_3 = -e_1 - e_3$ in $\mathbb{R}^4$. As we have seen that G also must be constant along each line in $\mathbb{R}^3$ parallel to r_1, r_2 and r_4, and as r_1, r_2, r_3, r_4 span the entire space $\mathbb{R}^4$, we conclude that there is only one choice for the function G, up to the constant const $= G(0,0,0,0)$. □

For an alternative derivation of relation (3.6), we pass from considering σ^2, the variability of the *relative* price changes, to considering σ_B^2, the variability of the *absolute* price changes. This will allow us to reduce the *two* explanatory variables σ^2 and P to *one* explanatory variable $\sigma_B^2 = \sigma^2 P^2$. We call σ_B the *Bachelier volatility* as it corresponds to Bachelier's original model from 1900 (see Bachelier, 1900). Recall that the dynamics of the price process $(P_t)_{t \geq 0}$ of the Black-Scholes versus the Bachelier model are

$$\begin{aligned} dP_t &= \sigma P_t dW_t, && \text{(Black-Schloes model)} \\ dP_t &= \sigma_B dW_t, && \text{(Bachelier model)} \end{aligned} \tag{3.12}$$

where W_t is a standard Brownian motion. Defining $\sigma_B = \sigma P$ the two models coincide remarkably well as long as P_t does not move too much (Schachermayer and Teichmann, 2008, compare e.g.). We therefore define

- $\sigma_B^2 = \sigma^2 P^2$ the Bachelier volatility in the interval $[t, t+T]$. Plugging in the dimensions $[\sigma^2] = \mathbb{T}^{-1}$ and $[P] = \mathbb{U}\mathbb{S}^{-1}$, we obtain

$$[\sigma_B^2] = \mathbb{U}^2\mathbb{S}^{-2}\mathbb{T}^{-1}.$$

A glance at Table 3.2 reveals that σ_B^2 has Modigliani-Miller dimension $\mathbb{M}$ equal to zero (just as the other variables V, C and N). This enables us to derive the assertion of Theorem 3.2 by using only the three obvious scaling invariances, but *without* imposing a priori the requirement of leverage neutrality.

Corollary 3.3. *Suppose the number of trades N depends* only *on the three quantities σ_B^2, V and C, i.e.,*

$$N = g(\sigma_B^2, V, C), \tag{3.13}$$

where the function $g : \mathbb{R}_+^3 \to \mathbb{R}_+$ is dimensionally invariant. *Then, there is a constant $c > 0$ such that the number of trades N obeys the relation*

$$N^{3/2} = c \cdot \frac{\sigma_B V}{C}. \tag{3.14}$$

The proof is analogous to (and even easier than) the above proof. Note that Proposition 3.1 and Corollary 3.3 both only rely on the very convincing invariance assumption with respect to $\mathbb{S}$, $\mathbb{T}$ and $\mathbb{U}$, but not on the "Leverage Neutrality Assumption".

Anticipating that relation (3.14) gives a superior fit to empirical data than relation (3.2) we can draw the following conclusion: the choice of σ_B^2, V, C as explanatory variables for the quantity N is superior to the choice σ^2, P, V made in Proposition 3.1 above.

Here is a "dimensional argument" why we should expect a better result from Corollary 3.3 as compared to Proposition 3.1. It follows from the very approach of dimensional analysis that everything hinges on the assumption that the chosen explanatory variables indeed "fully explain" the dependent variable. Of course, in reality such an assumption will – at best – only be approximately satisfied. The art of the game is to find a combination of explanatory variables which "best" explain the resulting variable. The choice of the variables σ_B^2, V, C as in Corollary 3.3 *automatically* implies that the "Leverage Neutrality Assumption" is satisfied as shown in Table 3.2. Indeed, the variables σ_B^2, V, C as well as N have a zero entry for the Modigliani-Miller dimension $\mathbb{M}$. Therefore, *any* function relating these variables is *automatically* leverage neutral. This is in contrast to the choice of variables σ^2, P, V in Proposition 3.1 as Table 3.1 reveals that P and σ^2 have a non-trivial dependence on $\mathbb{M}$. It follows that formula (3.2) does not satisfy the invariance relation dictated by the "Leverage

	σ_B^2	V	C	N
$\mathbb{S}$	-2	1	0	0
$\mathbb{U}$	2	0	1	0
$\mathbb{T}$	-1	-1	0	-1
$\mathbb{M}$	0	0	0	0

Table 3.2 A labelled overview of the dimensions of the quantities $V, \sigma_B^2 = \sigma^2 P^2$ and C.

Neutrality Assumption".

Finally, we examine the implications of substituting the cost per trade C by its more common counterpart, the bid-ask spread S, introduced above. In fact, in the present context it is equivalent to use either C or S as explanatory variables for the number of trades N - provided that the traded volume V is already one of the explanatory variables. Indeed, we have the relation $C = SQ = SV/N$ since the average trade size Q in the interval $[t, t+T]$ is given by the traded volume V divided by the number of trades N. Hence, if we know the functional relation between N and V, we also know the functional relation between N and Q and can therefore pass from S to $C = SQ$ and vice versa. Thus, we may restate Theorem 3.2 (and, equivalently, Corollary 3.3) in terms of the bid-ask spread S rather than the cost per trade C in the following corollary.

Corollary 3.4. *Suppose that the number of trades N depends* only *on the three quantities σ_B^2, V and S, i.e.,*

$$N = g(\sigma_B^2, V, S), \tag{3.15}$$

where the function $g : \mathbb{R}_+^3 \to \mathbb{R}_+$ dimensionally invariant *and* leverage neutral. *Then, there is a constant $c > 0$ such that the number of trades N obeys the relation*

$$N = c^2 \cdot \left(\frac{\sigma_B}{S}\right)^2. \tag{3.16}$$

We observe that the variables σ_B^2, V and S again have no Modigliani-Miller dimension $\mathbb{M}$, i.e., they are invariant under changes of the leverage. Therefore, formula (3.16) satisfies the invariance principle given by the "Leverage Neutrality Assumption". We note again that given the relations $C = SQ = SV/N$ as well as $\sigma_B^2 = \sigma^2 P^2$ the two equations (3.6) and (3.16) are indeed equivalent.

Relation (3.16) is precisely the one proposed by Wyart et al. (2008). By rearranging the terms, we find that

$$S^2 = c^2 \cdot \frac{\sigma_B^2}{N}. \tag{3.17}$$

The interpretation is that the squared Bachelier volatility per trade is proportional to the square of the spread. If we elaborate further on (3.17), we find that

$$\frac{S}{P} = c \cdot \frac{\sigma}{\sqrt{N}}. \tag{3.18}$$

Without loss of generality, we can determine the price P on the left hand side of (3.18) as midquote price, i.e., the average of the best ask- and bid price. Then, S/P refers to the so called proportional bid-ask spread which can be used to approximate a dealer's "round trip" transaction costs. Clearly, the approximate round-trip costs increase in the volatility of a relative price change and decrease in the trading activity.

Summing up this section, we have seen that the relation $N \sim \sigma^2$ proposed by Jones et al. (1994) follows from the restrictive assumption that the number of trades N *only* depends on the quantities σ^2, P and V as well as dimensional arguments (see Proposition 3.1). Going beyond the latter relation, it seems reasonable to include information concerning the bid-ask spread in our analysis. Depending on whether we choose the trading cost C or the bid-ask spread S directly, we are led to either the 3/2-law $N^{3/2} \sim \sigma PV/C$ proposed by Benzaquen et al. (2016) (see Theorem 3.2) or to the relation $S \sim \sigma_B/\sqrt{N}$ proposed by Wyart et al. (2008) (see Corollary 3.4). When proving the two latter relations we have seen that the assumption of leverage neutrality comes into play. Alternatively, we can also consider the product $\sigma^2 P^2$, rather than σ^2 and P separately. This consideration of the "Bachelier volatility" $\sigma_B = \sigma P$ reduces the complexity of the problem inasmuch as the assumption of leverage neutrality is not needed anymore. Again, the *actual* validity of any of the above scaling laws should be confirmed by exhaustive empirical analyses.

3.3 Empirical evidence

3.3.1 Degrees of universality and related literature

We now turn to the empirical analysis of relation (3.2) as well as of the $3/2$-law (3.6). When collecting data for the relevant quantities N, σ^2, V, P and C, one has to specify the considered

asset and the considered time period as well as the length T of the time interval over which the data is aggregated. We cannot expect that the constant c appearing in relations (3.2) resp. (3.6) is the same for each considered interval *and* each possible interval length *and* each considered asset in either one of the relations. We can only hope that a given relation holds *on average*. Based on the nomenclature introduced in Benzaquen et al. (2016), we therefore distinguish the following three degrees of universality attached to the validity of relations (3.2) and (3.6):

1. *No universality:* The relation holds on average for a fixed asset and a fixed interval length. However, the constant c varies significantly for different assets and different interval lengths.

2. *Weak universality:* The relation holds on average for some assets and some interval lengths with similar values from the constant c.

3. *Strong universality:* The relation holds on average for all assets and all interval lengths with similar values from the constant c.

Note that this distinction does not allow for the possibility that the validity attached to a given relation changes over time, simply because we consider only one specific time period.

Let us shortly discuss the relevant empirical evidence which can be found in the literature before turning to our own empirical analysis. Andersen et al. (2016) conducted an important empirical study in the present context. They test the relation

$$I = \frac{\sigma PV}{N^{3/2}}, \tag{3.19}$$

where I is independently and identically distributed across assets and time for E-mini S&P 500 futures contract. Neglecting the price P, they show that relation $N^{3/2} \sim V\sigma$ holds when averaging within and across trading days for this particular asset. In fact, their data fits the latter relation nearly perfectly compared to the relations $V \sim \sigma^2$ resp. $N \sim \sigma^2$ proposed by Tauchen and Pitts (1983) resp. Jones et al. (1994). Benzaquen et al. (2016) address the same question by examining eleven additional futures contracts as well as 300 US stocks. Aiming to confirm that $\beta = 3/2$ in the relation $N^{\beta} \sim \sigma PV$, they estimate β for each considered stock individually. They find that $\hat{\beta} = 1.54 \pm 0.11$, where the uncertainty here is the root mean square cross-sectional dispersion. Thus, these authors note that this provides evidence that the relation $N^{3/2} \sim \sigma PV$ holds also on the stock market and not only on the very liquid futures market. Moreover, they show that the distribution of I in (3.19) depends significantly on the studied asset and thus, conclude that relation (3.19) holds only with weak universality.

As an additional contribution, the authors reveal that the inclusion of the trading cost C is beneficial in the sense that their proposed invariant $\mathscr{I} = \sigma PVC^{-1}N^{-3/2}$ is almost constant for different assets.

Finally, let us mention the evidence in the earlier work by Wyart et al. (2008). These authors show that relation (3.17) describes the data very well when the right level of aggregation is chosen. When examining the France Telecom stock, S and $\sigma_B/\sqrt{N}$ are averaged over two trading days, while in case of NYSE stocks these quantities are averaged over an entire year. The constant c in relation (3.17) is found to lie between 1.2 and 1.6. Moreover, the authors note that the typical intraday pattern of the considered quantities is in line with (3.17): The U-shaped pattern of the volatility σ_B is explained by the decline of the bid-ask spread S and an increase of the number of trades N within the trading day.

3.3.2 Description of data

Our empirical analysis is based on limit order book data provided by the LOBSTER database (`https://lobsterdata.com`). The considered sampling period begins on January 2, 2015 and ends on August 31, 2015, leaving 167 trading days. Among all NASDAQ stocks, $d = 128$ sufficiently liquid stocks with high market capitalizations are chosen. Stocks are considered to be "sufficiently liquid" as long as the aggregated variables (defined below) can be reasonably treated as continuously distributed, i.e., the empirical distributions of the aggregated variables do not have points with obviously concentrated mass. Observations made during the thirty minutes after the opening of the exchange as well as trading halts are removed.

Let us fix an interval length $T \in \{30, 60, 120, 180, 360\}$ min for which a developed hypothesis is tested. For the sake of illustration, set the length of the considered time interval T to 60min. This interval length balances the tradeoff between sufficient aggregation of the data on the one hand and some intraday variability on the other hand. As a result, we are left with $n = 1002$ non-overlapping time intervals with equal length $T = 60$min. Let us concentrate on a specific asset $i \in \{1, \dots, d\}$ (omitting the index i for ease of notation in the remainder of Section 3.3.2) and let $j \in \{1, \dots, n\}$ refer to an arbitrary interval. Suppose the trades in the considered interval j arrive at irregularly spaced transaction times $t_1, t_2, \dots, t_{N_j}$. Then,

N_j denotes the number of trades in the interval j,

$Q_j = N_j^{-1} \sum_{k=1}^{N_j} Q_{t_k}$ denotes the average size of the trades in the interval j, where Q_{t_k} denotes the number of shares traded at time t_k,

V_j $= N_j \times Q_j$ is the traded volume in the interval j,

P_j $= N_j^{-1} \sum_{k=1}^{N_j} P_{t_k}$ denotes the average midquote price in the interval j, where $P_{t_k} = (A_{t_k} + B_{t_k})/2$ and A_{t_k} (resp. B_{t_k}) denotes the best ask (resp. bid) price after the transaction at time t_k,

$\hat{\sigma}_j^2$ denotes the estimated squared volatility in the interval j,

S_j $= N_j^{-1} \sum_{k=1}^{N_j} S_{t_k}$ denotes the average bid-ask spread in the interval j, where $S_{t_k} = A_{t_k} - B_{t_k}$ is the bid-ask spread after the transaction at time t_k, and

C_j $= Q_j \times S_j$ is the cost per trade in the interval j.

Note the following four details: Firstly, even though transaction times are recorded on a nano-second level, a time-stamp t_k is recorded L-times $(t_{k_1}, \ldots, t_{k_L})$ in the raw dataset when a market order is executed against L limit orders at time t_k. Such a multiple entry of the same time-stamp enters the number of trades N_j only once (not L-times). The size Q_{t_k} of the trade at time t_k is determined by summing the L-records in the dataset $Q_{t_{k_\ell}}$, $\ell = 1, \ldots, L$, i.e., $Q_{t_k} = \sum_{\ell=1}^{L} Q_{t_{k_\ell}}$. The midquote price P_{t_k} and the bid-ask spread S_{t_k} related to the merged market order of size Q_{t_k} are computed as volume-weighted averages

$$P_{t_k} = Q_{t_k}^{-1} \sum_{\ell=1}^{L} Q_{t_{k_\ell}} P_{t_{k_\ell}} \quad \text{and} \quad S_{t_k} = Q_{t_k}^{-1} \sum_{\ell=1}^{L} Q_{t_{k_\ell}} S_{t_{k_\ell}}.$$

Secondly, the aggregated variables, i.e., the average market order size Q_j, the average midquote price P_j and the average bid-ask spread S_j of interval j, are in fact not computed by the sample averages as state above. Since simple sample averages are sensitive with respect to outliers, e.g. huge market orders, Q_j, P_j and S_j are based on robust averages. In detail, we compute trimmed means of $Q_{t_1}, \ldots, Q_{t_{N_j}}$, $P_{t_1}, \ldots, P_{t_{N_j}}$ and $S_{t_1}, \ldots, S_{t_{N_j}}$ to obtain Q_j, P_j and S_j respectively. These trimmed means discard the upper 0.5% and the lower 0.5% of the corresponding ordered data and compute the average based on the remaining 99% of the data.

Thirdly, the estimated squared volatility σ_j^2 is computed as realized variance in interval j

$$\hat{\sigma}_j^2 = \sum_{k=2}^{N_j} \left(\log(P_{t_k}) - \log(P_{t_{k-1}}) \right)^2. \tag{3.20}$$

The properties of the estimator $\hat{\sigma}_j^2$ are well understood for a variety of models for the efficient price process $(P_t)_{t \geq 0}$. For example, if the dynamics of the efficient price process follows the stochastic model $dP_t = \sigma P_t dW_t$, with $\sigma > 0$, the estimator $\hat{\sigma}_j^2$ converges weakly in

probability to $\sigma^2 T$ (the quadratic variation of the increments of $(\log(P_t))_{t\geq 0}$) as the number of transactions within interval j becomes dense (as $N_j \to \infty$). The limit of $\hat{\sigma}_j^2$, however, does not coincide with the quadratic variation of the efficient price process, if the observed midquote price is contaminated by market microstructure noise. This noise, for instance, arises from market imperfections such as price discreteness or informational content in price changes (see Black, 1986). To check the robustness of our analysis with respect to the presence of market microstructure noise, several results below can likewise be confirmed by replacing the realized variance by the noise-robust estimator of the quadratic variation proposed by Hautsch and Podolskij (2013). It should be noticed that a distortion of the analysis by the bid-ask bounce is already avoided by considering midquote prices rather than transaction prices. The interested reader are referred to the gentle introduction explaining how noisy price observations erode the realized variance by Aït-Sahalia and Yu (2009).

Last but not least, note that Benzaquen et al. (2016) in fact define the cost per trade by $\widetilde{C}_j = N_j^{-1} \sum_{k=1}^{N_j} Q_{t_k} S_{t_k}$. This slight difference in the definitions becomes obviously negligible, if the bid-ask spread S_{t_k} is constant over the entire interval j. The results presented below are robust with respect to the employed version of the cost per trade as we shall see.

3.3.3 $N \sim \sigma^2$ versus $N^{3/2} \sim \sigma PV/C$

To check which of the relations $N \sim \sigma^2$ and $N^{3/2} \sim \sigma PV/C$ is superiorly supported by data, we consider for each stock ($i = 1, \ldots, d$) a multiplicative model of the form

$$N_{ij} = \exp(\alpha_i)(\hat{\sigma}_{ij}^2)^{\beta_i} \left(\frac{P_{ij} V_{ij}}{C_{ij}} \right)^{\gamma_i} \exp(\varepsilon_{ij}) \quad \text{with} \quad j = 1, \ldots, n, \tag{3.21}$$

where ε_{ij}, $j = 1, \ldots, n$, is an error term that satisfies standard regularity conditions and α_i, β_i and γ_i are unknown real valued parameters. A logarithmic transformation of (3.21) yields the linear model

$$\log(N_{ij}) = \alpha_i + \beta_i \log\left(\hat{\sigma}_{ij}^2\right) + \gamma_i \log\left(\frac{P_{ij} V_{ij}}{C_{ij}} \right) + \varepsilon_{ij}. \tag{3.22}$$

Since dimensional analysis imposes the restriction $\beta_i + \gamma_i = 1$ on the parameters β_i and γ_i, the value $\gamma_i = 0$ would imply the relation $N \sim \sigma^2$, whereas $\gamma_i = 2/3$ would imply the relation $N^{3/2} \sim \sigma PV/C$ from Theorem 3.2. The estimation of the coefficients β_i and γ_i subject to the restriction $\beta_i + \gamma_i = 1$ therefore allows to infer which of the two discussed relations is backed by stronger empirical evidence.

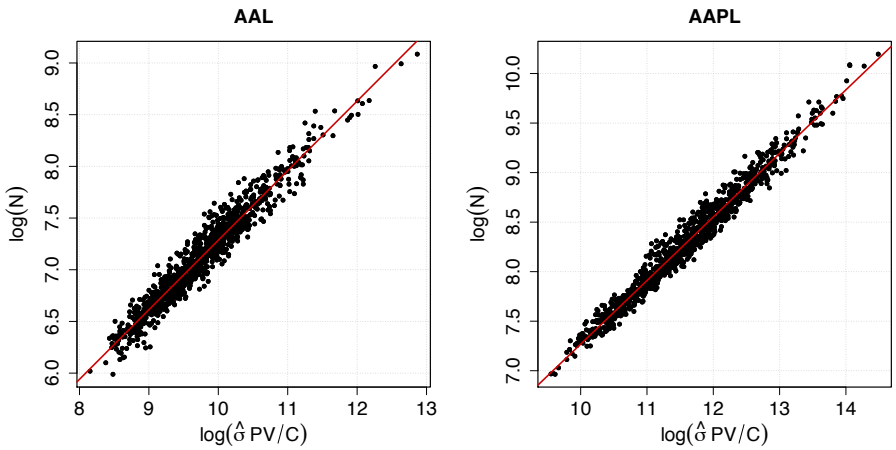


Fig. 3.1 The logarithmic dependent variable $\log(N)$ is plotted versus the logarithmic explanatory variable $\log(\hat{\sigma}PV/C)$ for the fixed interval length $T = 60$min and the two stocks AAL and AAPL. The lines indicate the estimated linear relations between the considered quantities.

Before turning to the constrained estimation of the parameters β_i and γ_i, it deserves to be emphasized that the functional relation between the logarithmic dependent variable $\log(N_j)$ and the logarithmic explanatory variable $\log(\hat{\sigma}_{ij}P_{ij}V_{ij}/C_{ij})$ can be reasonably assumed to be linear for all stocks $i = 1,\ldots,d$. To conclude this, we have visually inspected the bivariate point-clouds of dependent and explanatory variable. Figure 3.1 illustrates this relation for the stocks of the American Airline Group, Inc. (AAL) and Apple Inc. (AAPL). The remaining 126 stocks show similar patterns.

For each stock ($i = 1,\ldots,d$) and all interval lengths $T \in \{30,60,120,180,360\}$ min, we estimate the parameters β_i and γ_i in (3.22) by ordinary least squares subject to the constraint $\beta_i + \gamma_i = 1$. The corresponding estimate of interest is denoted by $\hat{\gamma}_i$. To present the results of these regressions in an informative and compact way, Figure 3.2 shows kernel density estimates of $\hat{\gamma}_i$ across i and for fixed T.

First, let us come to the main result of this section and concentrate on the solid graphs in Figure 3.2 referring to the standard setting based on the realized variance $\hat{\sigma}_{ij}^2$ defined in (3.20) and the cost per trade $C_{ij} = Q_{ij} \times S_{ij}$. If the parameter γ_i of the linear model (3.22) is equal to zero, then the underlying variables satisfy the simple relation $N \sim \sigma^2$. Similarly, if the parameter γ_i is equal to $2/3$, then we can conclude that the 3/2-law from Theorem 3.2 holds. As seen in Figure 3.2, the averages of the estimates $\hat{\gamma}_i$ (across i for different T) are clearly much closer to $2/3$ than to zero for all considered interval lengths T. This result

supports the claim made in Section 3.2 that there is stronger empirical support for the 3/2-law (or equivalently for the relation $N \sim (\sigma P/S)^2$) than for the relation $N \sim \sigma^2$.

Regarding the robustness of this insight, we have re-conducted the above regression analysis for two slightly different scenarios. One alternative setting considers replacing the realized variance in the linear model (3.22) by the market microstructure noise robust estimator of the quadratic variation of Hautsch and Podolskij (2013). The dashed graphs in Figure 3.2 are related to density estimates relying on corresponding parameter estimates $\hat{\gamma}_i$, $i = 1,\ldots,d$. The second modification of the initial setting replaces the cost per trade C_j in the linear model (3.22) by the variant $\widetilde{C}_j$ of Benzaquen et al. (2016). The dotted graphs in Figure 3.2 refer to corresponding density estimates. Despite some deviation in the estimates $\hat{\gamma}_i$ for these two alternative settings from the initial one, the solid, dashed and dotted graphs document a rather similar pattern among the estimates of the parameters γ_i for all interval lengths $T \in \{30, 60, 120, 180, 360\}$ min. These similarities lead to the conclusion that neither market microstructure noise nor the exact definition of the cost per trade erode the overall relation between the dependent and explanatory variables. In the remaining part of the manuscript, we take a closer look on the 3/2-law and try to find reasonable explanations for the systematic deviations of the estimates $\hat{\gamma}_i$ from $2/3$.

3.3.4 On the universality of the 3/2-law

In order to check the validity and universality of the $3/2$-law, $N^{3/2} = c \cdot \sigma PV/C$ (or equivalently of the relation $N = c^2 \cdot (\sigma P/S)^2$), we examine the variation of the constant c across assets and interval lengths. Hence, we do not rely on the estimators $\hat{\gamma}_i$ computed in Section 3.3.3. Instead, we compute for a fixed interval length T the quantity

$$\hat{c}_i = n^{-1} \sum_{j=1}^{n} \frac{C_{ij} N_{ij}^{3/2}}{\hat{\sigma}_{ij} P_{ij} V_{ij}} = n^{-1} \sum_{j=1}^{n} \frac{N_{ij}^{1/2}}{\hat{\sigma}_{ij}} \frac{S_{ij}}{P_{ij}}, \quad \text{for} \quad i = 1,\ldots,d,$$

where n is the number of non-overlapping time intervals with equal length T. The left panel of Figure 3.3 shows the estimates $\hat{c}_i$ for different values of T. Note that the rainbow-color-code refers to the ordered values of $\hat{c}_i$ for $T = 120$min. As we recover the same rainbow-pattern also for the other interval lengths $T \in \{30, 60, 180, 360\}$ min, we can conclude that there is little variation of the estimates $\hat{c}_i$ for a fixed stock i across different interval lengths T. This small variation of $\hat{c}_i$ for fixed i and varying $T \in \{30, 60, 120, 180, 360\}$ min endows the 3/2-law with a certain degree of universality. However, the present cross-sectional dispersion in $\hat{c}_i$ across different assets i, i.e., the fact that depending on the considered stock the estimates $\hat{c}_i$ range from two to five, does not allow awarding the 3/2-law with strong universality. Thus,

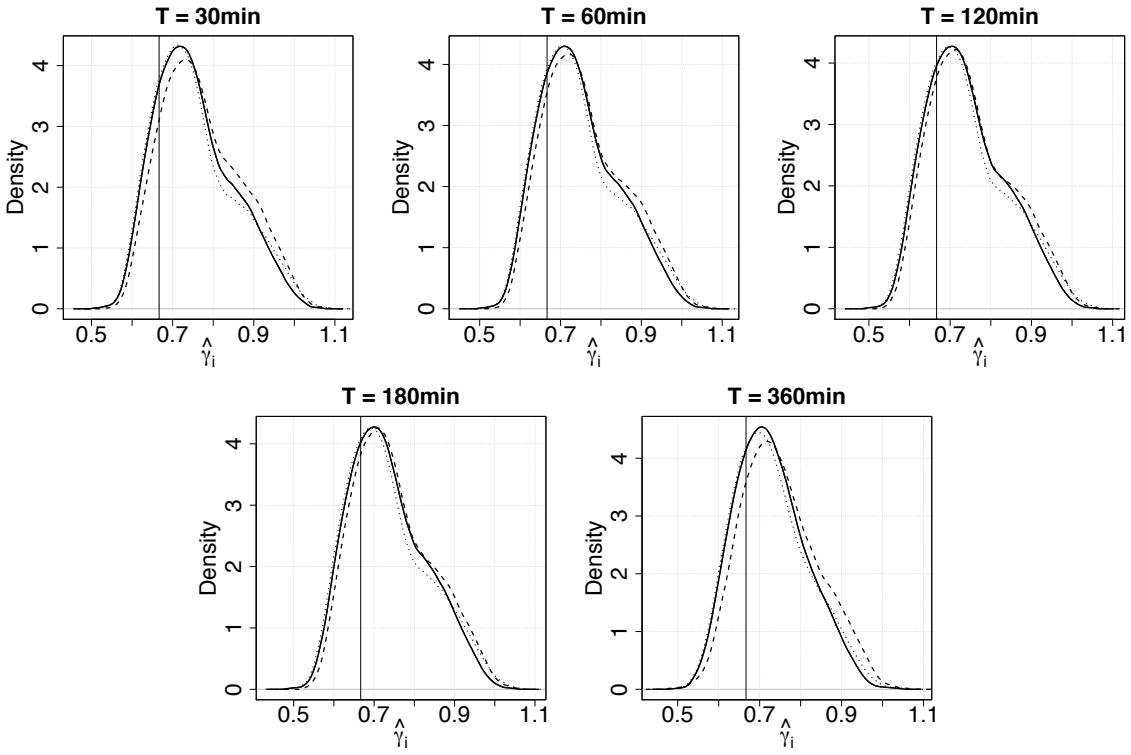


Fig. 3.2 The panels show kernel density estimates across the estimated parameters $\hat{\gamma}_i$ for different interval lengths $T \in \{30, 60, 120, 180, 360\}$ min.

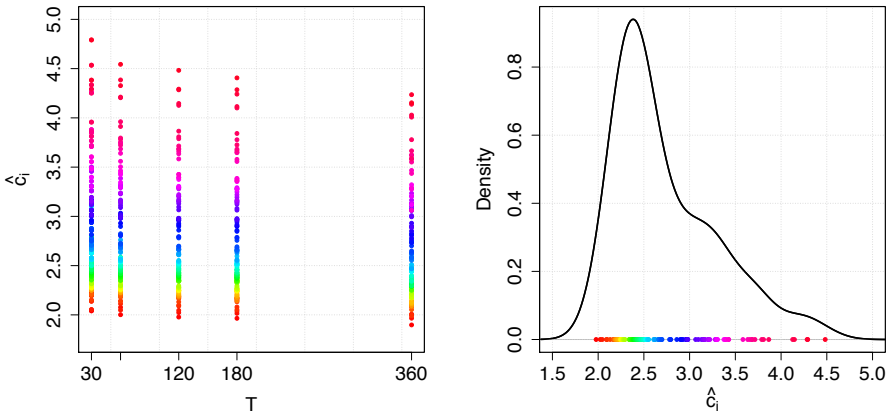


Fig. 3.3 The left panel shows the computed values for $\hat{c}_i$ in dependence of $T \in \{30, 60, 120, 180, 360\}$ min. The right panel shows a kernel density estimate across the estimates $\hat{c}_i$ for fixed $T = 120$min.

we draw the same conclusion as Benzaquen et al. (2016) that the 3/2-law holds with weak universality. For completeness, the kernel density estimate in the right panel of Figure 3.3 illustrates the distribution of the estimates $\hat{c}_i$, $i = 1,\dots,d$ for $T = 120$min.

3.4 A closer look on volatility

We have seen that the volatility σ plays a dominant role in explaining the trading activity N. The squared volatility σ^2 of a given stock during a fixed interval $[t,t+T]$ was defined as the variance of the change of the log-price

$$\sigma^2 := \mathbb{V}\mathrm{ar}\left(\log(P_{t+T}) - \log(P_t)\right). \tag{3.23}$$

When specifying the definition of σ^2 in this way we had in mind the Black-Scholes model,

$$dP_t = P_t\left(\sigma dW_t + \mu dt\right), \tag{3.24}$$

where, fixing the normalization $T = 1$, formula (3.23) indeed recovers the constant σ in (3.24). Going beyond Black-Scholes, consider a price process of the form

$$P_t = P_0 \exp\left(\int_0^t \sigma_u dW_u\right) \tag{3.25}$$

where $(\sigma_t)_{t\geq 0}$ is an arbitrary stochastic process (satisfying suitable regularity conditions). In this case, formula (3.23) should, of course, be interpreted conditionally on the sigma-algebra $\mathscr{F}_t$ and we obtain the "Wald identity"

$$\mathbb{V}\mathrm{ar}\left(\log(P_{t+T}) - \log(P_t)|\mathscr{F}_t\right) = \mathbb{E}\left(\int_t^{t+T} \sigma_u^2 du|\mathscr{F}_t\right). \tag{3.26}$$

This implies in particular that, as long as we are in the framework of processes of the form (3.25), the above chosen scaling

$$[\sigma^2] = \mathbb{T}^{-1},$$

is the only reasonable choice.

But let us have a closer look at what we are actually doing here. The above reasoning tacitly assumes that we are starting from a *stochastic model* of a price process. The present situation, however, dictates a different point of view: we start from empirical tick data observed during the interval $[t,t+T]$. Even when we make the heroic assumption that this data is accurately modeled, e.g. by the Black Scholes model (3.24), the number σ^2 which we

plug into the formula $N = g(\sigma^2, \dots)$ can only be an *estimator* of σ^2 obtained from the data at hand. This implies that, strictly speaking, we should write our formulas as $N = g(\hat{\sigma}^2, \dots)$ in dependence of the *estimated* squared volatility $\hat{\sigma}^2$. The gist of the argument is that for the purpose of dimensional analysis the scaling which is relevant is that of the *estimator* of the volatility rather than that of the *true* volatility (whatever this is). To be concrete, suppose that we are given price data $(P_{t_k})_{k=1,\dots,N}$ for a grid $t \le t_1 < \cdots < t_N \le t+T$ in the interval $[t, t+T]$. An obvious choice for the estimator of the squared volatility, which is also used in Section 3.3 above, is

$$\hat{\sigma}^2 := \sum_{k=2}^{N} \big(\log(P_{t_k}) - \log(P_{t_{k-1}})\big)^2. \tag{3.27}$$

Clearly, this estimator has the dimension $[\hat{\sigma}^2] = \mathbb{T}^{-1}$ if we suppose that the typical distance $\Delta t_k = t_{k+1} - t_k$ (in absolute terms) does not depend on whether we measure time in seconds or in minutes. Hence, for the estimator $\hat{\sigma}^2$, the hypothesis $[\hat{\sigma}^2] = \mathbb{T}^{-1}$ underlying the dimensional analysis in Section 3.2 is satisfied.

However, we can also think of other estimators. Fix $H \in (0,1)$ and define the estimator $\hat{\sigma}^2(H)$ by

$$\hat{\sigma}^2(H) := \left(\sum_{k=2}^{N} |\log(P_{t_k}) - \log(P_{t_{k-1}})|^{1/H} \right)^{2H}. \tag{3.28}$$

To motivate this estimator, consider the model

$$P_t = P_0 \exp(\sigma W_t^H), \qquad t \ge 0, \tag{3.29}$$

where $\sigma > 0$ is a fixed number and $(W_t^H)_{t \ge 0}$ is a *fractional* Brownian motion with Hurst parameter H, starting at $W_0^H = 0$. In this case, the estimator $\hat{\sigma}^2(H)$ in (3.28) is a consistent estimator for the parameter σ^2 in (3.29). But the estimator $\hat{\sigma}^2(H)$ now scales differently in time than the quadratic estimator $\hat{\sigma}^2$ (see Coutin, 2007; Pratelli, 2011), namely

$$[\hat{\sigma}^2(H)] = \mathbb{T}^{-2H}. \tag{3.30}$$

Models for the price process $(P_t)_{t \ge 0}$ involving fractional Brownian motion as in (3.29) have been proposed, notably by B. Mandelbrot, already more than 50 years ago (Mandelbrot, 1963; Mandelbrot and Ness, 1968) and there may be good reasons not to rule them out a priori.

Here is another example where a sub-diffusive behavior of the price process $(P_t)_{t\geq 0}$ occurs, due to a micro-structural effect: the discrete nature of the prices in the real world (compare Benzaquen et al., 2016, we thank Jean-Philippe Bouchaud for bringing this phenomenon to our attention). To present the idea in its simplest possible form, suppose that the price process $(\check{P}_t)_{t\geq 0}$ is given by

$$\log(\check{P}_t) = \mathrm{int}(W_t),$$

where $(W_t)_{t\geq 0}$ is a standard Brownian motion and $\mathrm{int}(x)$ denotes the integer closest to the real number x, i.e., $\mathrm{int}(x) = \sup\{n \in \mathbb{Z} : n \leq x + 0.5\}$. Fix again an interval $[t, t+T]$ and consider the quantity

$$\check{\sigma}^2 = (\check{\sigma}^2)_t^{t+T} = \mathbb{V}\mathrm{ar}\left(\log(\check{P}_{t+T}) - \log(\check{P}_t)\right).$$

For small $T > 0$, we show in Appendix B.4 that

$$(\check{\sigma}^2)_t^{t+T} \approx \mathrm{c}\,\sqrt{T},$$

for some constant $c > 0$. Hence, if the interval length T is sufficiently small, we recover that $[\check{\sigma}^2] = \mathbb{T}^{-1/2}$, rather than the usual scaling in the dimension time, i.e., $\mathbb{T}^{-1}$.

This observation indicates, that if the interval length T is small compared to the width of the price grid, i.e., the tick value, we observe a sub-diffusive behavior of the price process even if the "efficient", unobserved price process is assumed to be a diffusion. We refer to Robert and Rosenbaum (2010) for a detailed discussion of how to account for the discrete nature of prices. For now, this rough argument should only serve as motivation that there might be plenty of reasons why the scaling $[\sigma^2] = \mathbb{T}^{-1}$ is, in practical situations, not as clearly granted as it might seem at first glance.

For all these reasons we drop in this section the convenient dimensional assumption $[\sigma^2] = \mathbb{T}^{-1}$ and replace it by the subsequent more general assumption.

***H*-Assumption.** *There is $H \in (0,1)$ such that the squared volatility estimator $\hat{\sigma}^2(H)$ has dimension*

$$[\hat{\sigma}^2(H)] = \mathbb{T}^{-2H}.$$

Proposition 3.5 ($(1+H)$-law)**.** *Suppose that the "Leverage Neutrality Assumption" as well as the "H-Assumption" hold true and that the number of trades N depends* only *on the four quantities $\hat{\sigma}^2(H), P, V$ and C, i.e.,*

$$N = g(\hat{\sigma}^2(H), P, V, C),$$

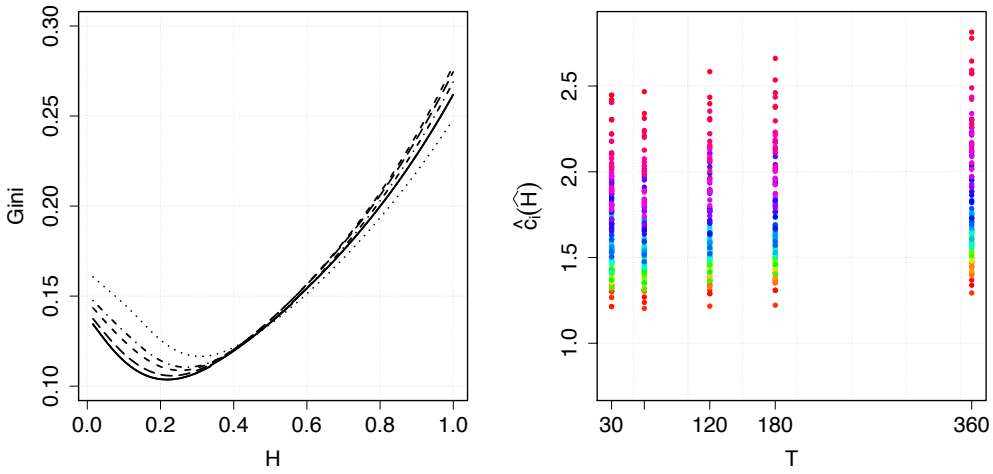


Fig. 3.4 The left panel illustrates the Gini-coefficient in dependence of H for $T = 30$min (solid), $T = 60$min (long-dashed), $T = 120$min (dashed), $T = 180$min (dashed-dotted) and $T = 360$min (dotted). The right panel shows the computed values for $\hat{c}_i(\widehat{H})$ such that $\widehat{H}$ minimizes the Gini-coefficient for fixed $T \in \{30, 60, 120, 180, 360\}$ min.

where the function $g : \mathbb{R}_+^4 \to \mathbb{R}_+$ *is* dimensionally invariant *and* leverage neutral. *Then, there is a constant* $c > 0$ *such that the number of trades* N *obeys the relation*

$$N^{1+H} = c \cdot \frac{\hat{\sigma}(H)PV}{C}. \tag{3.31}$$

The proof is analogous to the proof of Theorem 3.2 and is given in Appendix B.3.

The hypothesis of the above proposition assumes that $H \in (0,1)$ is known a priori. As H is typically unknown in practical applications, we can therefore ask the following question: For which H does relation (3.31) fit the empirical data best? We address this question in the following subsection.

3.4.1 Empirical evidence under the H-Assumption

According to arguments from dimensional analysis, the constant c and the parameter H from Equation (3.31) should at best be identical for all stocks and all interval lengths T. The empirical results above, however, have revealed cross-sectional dispersion which might be related to the restrictive assumption $[\hat{\sigma}^2] = \mathbb{T}^{-1}$. This restriction motivates the empirical exercise of this section: Can we determine an $H \in (0,1)$ in (3.31) that minimizes the cross-sectional dispersion across the estimates of c?

Following Proposition 3.5, we therefore compute the estimates $\hat{c}_i(H)$ for different H as

$$\hat{c}_i(H) = n^{-1} \sum_{j=1}^{n} \frac{N_{ij}^{1+H} C_{ij}}{\hat{\sigma}_{ij}(H) P_{ij} V_{ij}} = n^{-1} \sum_{j=1}^{n} \frac{N_{ij}^{H}}{\hat{\sigma}_{ij}(H)} \frac{S_{ij}}{P_{ij}}, \quad \text{for} \quad i = 1, \ldots, d,$$

where $\hat{\sigma}_{ij}^2(H)$ is defined in (3.28), $H \in (0,1)$. Both variables N_{ij}^H and $\hat{\sigma}_{ij}(H)$ increase as H increases, so that it is not obvious how $\hat{c}_i(H)$ behaves when H increases. We find empirically that overall the constant $\hat{c}_i(H)$ typically increases in H. Addressing the above question therefore requires a scale invariant measure for the variation in $\hat{c}_i(H)$ such as the Gini-coefficient which is given by

$$\mathscr{G}(x_1, \ldots, x_n) = \frac{2 \sum_{i=1}^{n} i x_{[i]}}{(n-1) \sum_{i=1}^{n} x_{[i]}} - \frac{n+1}{n-1},$$

for the ordered data $x_{[1]} < x_{[2]} < \ldots < x_{[n]}$. Note that the Gini-coefficient $\mathscr{G}(x_1, \ldots, x_n) \in [0,1]$ is interpreted as a measure for inequality. If all values $x_1, \ldots, x_n$ are equal, $\mathscr{G}$ equals zero. In case of strong heterogeneity in $x_1, \ldots, x_n$ the Gini-coefficient approaches one.[3]

Now, we minimize the Gini-coefficient of $(\hat{c}_i(H))_{i=1,\ldots,d}$ with respect to H in order to find

$$\widehat{H} = \underset{H \in (0,1)}{\arg\min}\, \mathscr{G}(\hat{c}_1(H), \ldots, \hat{c}_n(H)).$$

The left panel of Figure 3.4 plots the Gini-coefficient in dependence of H for different interval length T. We roughly find that $\widehat{H} = 0.22$ for $T = 30$min, $\widehat{H} = 0.23$ for $T = 60$min, $\widehat{H} = 0.25$ for $T = 120$min, $\widehat{H} = 0.27$ for $T = 180$min and $\widehat{H} = 0.31$ for $T = 360$min. The rainbow-color-code of Figure 3.3 has been transferred to the right panel of Figure 3.4. In contrast to Figure 3.3 yet, we present the quantities $\hat{c}_i(\widehat{H})$ in dependence of the optimal $\widehat{H}$ for the given interval length T. In case $T = 120$min for instance, the estimates $\hat{c}_i(H = 0.25)$ range from 1.2 to 2.6 for different assets i. On an absolute scale, the variation seems to be smaller compared to Figure 3.3, where the estimates $\hat{c}_i(H = 0.5)$ lie between 2 and 4.5 for the same interval length $T = 120$min. In relative terms though, the difference between the variation in $\hat{c}_i(H = 0.25)$ and $\hat{c}_i(H = 0.5)$ is not so significant, as $\mathscr{G}(\hat{c}_1(H = 0.25), \ldots, \hat{c}_n(H = 0.25)) = 0.11$ compared to $\mathscr{G}(\hat{c}_1(H = 0.5), \ldots, \hat{c}_n(H = 0.5)) = 0.14$ for $T = 120$min.

For now, we can only speculate on reasons why the optimal $\widehat{H}$ is strikingly smaller than $1/2$ for all interval lengths T. The quantity $\hat{c}_i(H)$ relies on tick-by-tick data, so that an

[3] The coefficient of variation defined as the ratio of the standard deviation to the sample average could be employed as an alternative to the Gini-coefficient. The presented results are widely robust with respect to the chosen measure of standardized dispersion.

obvious explanation for these unexpected optimal values of H are market microstructure effects. To be more concrete, Benzaquen et al. (2016) observe similar to our results a sub-diffusive behavior for so called large tick future contracts. Large tick assets are defined such that their bid-ask spread is almost always equal to one tick, see e.g. Eisler et al. (2012). Most of the stocks in our sample can be categorized as large tick stocks based on this definition.

When referring to market microstructure effects, however, it deserves to be stressed that the value $H = 1/2$ is implied by numerous models for the efficient price process $(P_t)_{t\geq 0}$, which are backed by empirical evidence and take market microstructure effects into account. Hence, the scaling of the squared volatility through time implied by $H = 1/2$ seems suitable in many applications. We also note that the Gini-coefficient $\mathscr{G}$ in Figure 3.4 does not vary drastically when H ranges between the optimal $\widehat{H} \approx 0.25$ and the traditional $H = 1/2$, namely roughly between $\mathscr{G} = 0.12$ and $\mathscr{G} = 0.15$. Hence, the value of H does not seem to play a very significant role in explaining the heterogeneity of the value of $\hat{c}_{ij}(H)$. Nevertheless, a better understanding of the behavior of $\widehat{H}$ seems to us a challenging topic for future research.

3.5 Conclusion

Finding laws relating the trading activity (defined here as the number of trades N within a given time interval) to other relevant market quantities has been the subject of numerous investigations. The earliest contribution dating as far back as the beginning of the 1970s. Two decades later, Jones et al. (1994) suggested the relation $N \sim \sigma^2$ based on an extensive empirical study. Other landmark contributions include the relation $N \sim (\sigma P/S)^2$ of Wyart et al. (2008) and the so called $3/2$-law $N^{3/2} \sim \sigma PV/C$ of Benzaquen et al. (2016), respectively obtained using market microstructure arguments, which were supported by empirical evidence. In the first part of the paper we show that all these scaling laws can be obtained using arguments relying on dimensional analysis. The relation $N \sim \sigma^2$ follows from the assumption that N is fully explained by the squared volatility σ^2, the asset price P and the traded volume V, and the assumption that the relation between these quantities is invariant under changes of the dimensions shares $\mathbb{S}$, time $\mathbb{T}$ and money $\mathbb{S}$. The somewhat refined relation $N^{3/2} \sim \sigma PV/C$ is obtained when assuming that N depends only on σ^2, P, V and the cost of trading C, and assuming in addition, that an invariance principle known as "Leverage Neutrality" holds true. This "Leverage Neutrality Assumption" can be seen as a no-arbitrage condition enabling us to obtain a unique functional relation from the assumption $N = g(\sigma^2, P, V, C)$. Substituting the quantity C by the bid-ask spread S in the latter assumption, we derive the relation $N \sim (\sigma P/S)^2$, which is shown to be equivalent to the $3/2$-law. Alternatively, we can consider the volatility of the *relative* price change

instead of the *absolute* price change, i.e., assume $N = g(\sigma^2 P^2, V, C)$ resp. $N = g(\sigma^2 P^2, V, S)$. This assumption simplifies the analysis in that a unique solution for $g(\cdot,\cdot,\cdot)$ can be obtained without recourse to the "Leverage Neutrality Assumption". Since our *theoretical* analysis relies on a set of well-defined, but not necessarily realistic assumptions, the validity of any of the aforementioned scaling laws needs to be confirmed through an empirical analysis.

Based on data from the NASDAQ stock exchange, we provide empirical evidence that the $3/2$-law $N^{3/2} = c \cdot \sigma PV/C$ (or equivalently $N = c^2 \cdot (\sigma P/S)^2$) fits the data clearly better than $N \sim \sigma^2$. In fact, the $3/2$-law holds for a fixed asset and a fixed interval length. However, the estimated value of the constant c strongly depends on the considered asset. In the language of Benzaquen et al. (2016), this means that the $3/2$-law holds with weak universality.

Finally, we note that both our theoretical and empirical analysis relied on the assumption that the scaling of σ^2 is inversely proportional to time $\mathbb{T}$. This hypothesis is clearly debatable as it tacitly assumes diffusive price behaviors, and ignores e.g. the discrete nature of prices. A closer look at the scaling of σ^2 suggests the scaling $[\sigma^2] = \mathbb{T}^{-2H}$ for some $H \in (0,1)$ that can be seen e.g. as the Hurst parameter of a fractional Brownian motion. Repeating our dimensional arguments, the latter scaling of σ^2 yields the relation $N^{1+H} \sim \sigma^2 PV/C$. An essential drawback of this more general situation is that the parameter H is unknown. We formulate an optimality criterion for the choice of H. It should yield the most homogeneous estimates for the proportionality coefficients $\hat{c}_i(H)$. A preliminary analysis implies that, on average, the optimal $\hat{H}$ is of the order 0.25, i.e., quite different from the assumption $H = 0.5$. Although the overall effect of this passage from $H = 0.5$ to $\hat{H} \approx 0.25$ turns out to have only mild effects on the issue of universality of the corresponding laws, we believe that this phenomenon merits further investigation.

CONCLUSION

First, let me give some concluding remarks concerning the second of the two main topics discussed in this thesis. We have seen that dimensional analysis is a useful tool when applied in the context of financial markets. In this respect, it is crucial to select a meaningful set of explanatory variables to fully explain the variable of interest. One then needs to identify the units in which the considered quantities are measured. In financial applications the quantities are typically measured in monetary units, in units of time and in numbers of shares. The three implied dimensional invariances allow us to uniquely determine the functional relation between the variable of interest and at most three explanatory variables. If a fourth explanatory variable appears in an application at hand, then the assumption of leverage neutrality can be imposed to determine a unique relation. This assumption establishes how given quantities behave when the stock-issuing company changes its capital structure by paying dividends or by raising new capital. Once we understand these basic principles, the derivation of a considered relation is as simple as solving a linear system of equations and can be seen as a starting point for a more evolved analysis.

As an application of these concepts we examine two different variables of interest: the market impact of a meta-order and the trading activity. With respect to the first variable, the above line of argument allows us the derive the well known square-root law for market impact. In case of the trading activity, different sets of explanatory variables lead to different relations, which have already been established in the literature. We compare these relations based on an empirical study with data from the stock market and thereby illustrate the amazing power of dimensional analysis.

We now come to the discussion of the first chapter's topic: robust portfolio optimization. The portfolio selection problem turns out to be a good problem to conceptually understand robust decision making in stochastic optimization. In non-robust stochastic optimization the decision is based on a specific stochastic model. We outline two approaches which allow us to directly account for ambiguity with respect to this stochastic model, referred to as

reference model. Similarly as one typically models the risk aversion of a decision maker in stochastic optimization, we model the "ambiguity aversion" of a decision maker, i.e., her distrust in the reference model. In the first approach, the ambiguity is considered with respect to the model for the joint distribution of the assets' returns. In this set-up, high ambiguity aversion leads to an equally weighted portfolio investment strategy. Intuitively, this means that when an investor has no information concerning the assets in her portfolio, it is optimal to invest an equal fraction of her wealth into each asset.

We contrast this approach with portfolio optimization under dependence uncertainty. We now assume that the investor has reliable information concerning the marginal distributions of the assets' returns. Hence, the ambiguity is only on the level of the dependence structure between the asset returns. In this framework, we prove that under high ambiguity aversion, i.e., high dependence uncertainty, it is optimal to invest only in one single asset. This result can be explained as follows. In contrast to the former approach, the investor has - per assumption - perfect information regarding each individual asset. She incorporates this information into her decision by identifying the best performing asset. As she possesses no information of the dependence between the assets, in the worst possible case the assets' returns are perfectly correlated. It is therefore optimal to only invest into the best performing asset.

We therefore conclude that robust decision making can have diametrically opposed implications depending on the particular way of accounting for model ambiguity.

REFERENCES

Aas, K. and Puccetti, G. (2014). Bounds for total economic capital: the DNB case study. *Extremes*, 17(4):693–715.

Acerbi, C. (2002). Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26(7):1505–1518.

Aït-Sahalia, Y. and Yu, J. (2009). High frequency market microstructure noise estimates and liquidity measures. *The Annals of Applied Statistics*, 3(1):422–457.

Almgren, R., Thum, C., Hauptmann, E., and Li, H. (2005). Direct estimation of equity market impact. *Risk*, 18:57–62.

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance*, 51(1):169–204.

Andersen, T. G., Bondarenko, O., Kyle, A. S., and Obizhaeva, A. A. (2016). Intraday trading invariance in the E-mini S&P 500 futures market. *Available at SSRN: 2693810*.

Ané, T. and Geman, H. (2000). Order flow, transaction clock, and normality of asset returns. *The Journal of Finance*, 55(5):2259–2284.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, 9(3):203–228.

Bachelier, L. (1900). Théorie de la spéculation.

Bacry, E., Iuga, A., Lasnier, M., and Lehalle, C.-A. (2015). Market impacts and the life cycle of investors orders. *Market Microstructure and Liquidity*, 01(02):1550009.

Basak, S. and Shapiro, A. (2001). Value-at-risk-based risk management: optimal policies and asset prices. *Review of Financial studies*, 14(2):371–405.

Benzaquen, M., Donier, J., and Bouchaud, J.-P. (2016). Unravelling the trading invariance hypothesis. *Market Microstructure and Liquidity*, 02(03n04):1650009.

Bershova, N. and Rakhlin, D. (2013). The non-linear market impact of large trades: evidence from buy-side order flow. *Quantitative Finance*, 13(11):1759–1778.

Bertsimas, D., Gupta, V., and Kallus, N. (2014). Robust SAA. *arXiv preprint arXiv:1408.4445*, page 617.

Black, F. (1986). Noise. *The Journal of Finance*, 41(3):528–543.

Bluman, G. and Kumei, S. (2013). *Symmetries and differential equations*, volume 154. Springer Science & Business Media.

Bouchaud, J.-P., Farmer, J. D., and Lillo, F. (2009). How markets slowly digest changes in supply and demand. In *Hanbook of Financial Markets: Dynamics and Evolution*, pages 57–156. Elsevier, Academic Press.

Brandt, M. (2009). Portfolio choice problems. *Handbook of financial econometrics*, 1:269–336.

Brokmann, X., Sérié, E., Kockelkoren, J., and Bouchaud, J.-P. (2015). Slow decay of impact in equity markets. *Market Microstructure and Liquidity*, 01(02):1550007.

Charnes, A. and Cooper, W. W. (1959). Chance-constrained programming. *Management science*, 6(1):73–79.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–155.

Consigli, G., Kuhn, D., and Brandimarte, P. (2017). *Optimal financial decision making under uncertainty*, pages 255–290. Springer International Publishing, Cham.

Coutin, L. (2007). An introduction to (stochastic) calculus with respect to fractional Brownian motion. In *Séminaire de Probabilités XL*, volume 1899 of *Lecture Notes in Mathematics*, pages 3–65. Springer.

Curtis, W., Logan, J. D., and Parker, W. (1982). Dimensional analysis and the Pi theorem. *Linear Algebra and its Applications*, 47:117–126.

Dantzig, G. B. (1955). Linear programming under uncertainty. *Management science*, 1(3-4):197–206.

Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612.

Denuit, M., Genest, C., and Marceau, E. (1999). Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, 25(1):85 – 104.

Doan, X. V., Li, X., and Natarajan, K. (2015). Robustness to dependency in portfolio optimization using overlapping marginals. *Operations Research*, 63(6):1468–1488.

Donier, J., Bonart, J., Mastromatteo, I., and Bouchaud, J.-P. (2015). A fully consistent, minimal model for non-linear market impact. *Quantitative Finance*, 15(7):1109–1121.

Dufour, A. and Engle, R. F. (2000). Time and the price impact of a trade. *The Journal of Finance*, 55(6):2467–2498.

Edirisinghe, N. C. P. (2011). *Stochastic programming approximations using limited moment information, with application to asset allocation*, pages 97–138. Springer New York, New York, NY.

Eisler, Z., Bouchaud, J.-P., and Kockelkoren, J. (2012). The price impact of order book events: market orders, limit orders and cancellations. *Quantitative Finance*, 12(9):1395–1419.

Embrechts, P., Wang, B., and Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. *Finance and Stochastics*, 19(4):763–790.

Emmer, S., Kratz, M., and Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, 18(2).

Engle, R., Ferstenberg, R., and Russell, J. (2012). Measuring and modeling execution cost and risk. *The Journal of Portfolio Management*, 38(2):14–28.

Epps, T. W. and Epps, M. L. (1976). The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis. *Econometrica*, 44(2):305–321.

Esfahani, P. M. and Kuhn, D. (2015). Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, pages 1–52.

Fabozzi, F. J., Huang, D., and Zhou, G. (2010). Robust portfolios: contributions from operations research and finance. *Annals of Operations Research*, 176(1):191–220.

Fermanian, J.-D., Radulovic, D., and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10(5):847–860.

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and stochastics*, 6(4):429–447.

Föllmer, H. and Schied, A. (2011). *Stochastic finance: an introduction in discrete time*. Walter de Gruyter.

Foucault, T., Pagano, M., Roell, A., and Röell, A. (2013). *Market liquidity: theory, evidence, and policy*. Oxford University Press.

Fournier, N. and Guilin, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162:707–738.

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2006). Institutional investors and stock market volatility. *The Quarterly Journal of Economics*, 121(2):461–504.

Gao, R. and Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. *arXiv preprint arXiv:1604.02199*.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2):199 – 213.

Gomes, C. and Waelbroeck, H. (2015). Is market impact a measure of the information value of trades? Market response to liquidity vs. informed metaorders. *Quantitative Finance*, 15(5):773–793.

Hanasusanto, G. A., Roitch, V., Kuhn, D., and Wiesemann, W. (2015). A distributionally robust perspective on uncertainty quantification and chance constrained programming. *Mathematical Programming*, 151(1):35–62.

Hansen, L. P. and Sargent, T. J. (2001). Robust control and model uncertainty. *The American Economic Review*, 91(2):60–66.

Hautsch, N. and Podolskij, M. (2013). Preaveraging-based estimation of quadratic variation in the presence of noise and jumps: Theory, implementation, and empirical evidence. *Journal of Business & Economic Statistics*, 31(2):165–183.

Huntley, H. E. (1967). *Dimensional analysis*. Dover Publications, New York.

Ji, R. and Lejeune, M. (2015). Data-driven optimization of ambiguous reward-risk ratio measures. *Available at SSRN 2707122*.

Jones, C. M., Kaul, G., and Lipson, M. L. (1994). Transactions, volume, and volatility. *The Review of Financial Studies*, 7(4):631–651.

Kakouris, I. and Rustem, B. (2014). Robust portfolio optimization with copulas. *European Journal of Operational Research*, 235(1):28 – 37.

Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *The Journal of Financial and Quantitative Analysis*, 22(1):109–126.

Kim, J. H., Kim, W. C., and Fabozzi, F. J. (2014). Recent developments in robust portfolios with a worst-case approach. *Journal of Optimization Theory and Applications*, 161(1):103–121.

Konno, H. and Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. *Management science*, 37(5):519–531.

Kusuoka, S. (2001). On law invariant coherent risk measures. In *Advances in mathematical economics*, pages 83–95. Springer.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.

Kyle, A. S. and Obizhaeva, A. A. (2016). Market microstructure invariance: Empirical hypotheses. *Econometrica*, 84(4):1345–1404.

Kyle, A. S. and Obizhaeva, A. A. (2017a). Dimensional analysis, leverage neutrality, and market microstructure invariance. *Available at 2785559*.

Kyle, A. S. and Obizhaeva, A. A. (2017b). Market microstructure invariance: A dynamic equilibrium model. *Available at SSRN: 2749531*.

Liesenfeld, R. (2001). A generalized bivariate mixture model for stock price volatility and trading volume. *Journal of Econometrics*, 104(1):141–178.

Liu, Y., Meskarian, R., and Xu, H. (2017). Distributionally robust reward-risk ratio optimization with moment constraints. *SIAM Journal on Optimization*, 27(2):957–985.

Lux, T. and Papapantoleon, A. (2016). Model-free bounds on value-at-risk using partial dependence information. *arXiv preprint arXiv:1610.09734*.

Makarov, G. (1982). Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed. *Theory of Probability & its Applications*, 26(4):803–806.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419.

Mandelbrot, B. B. and Ness, J. W. V. (1968). Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Review*, 10(4):422–437.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Markowitz, H. (1959). *Portfolio selection: Efficient diversification of investments.* Yale University Press.

Mastromatteo, I., Toth, B., and Bouchaud, J.-P. (2014). Agent-based models for latent liquidity and concave price impact. *Physical Review E*, 89(4):042805.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, NJ, USA.

Meskarian, R., Xu, H., and Fliege, J. (2012). Numerical methods for stochastic programs with second order dominance constraints with applications to portfolio optimization. *European Journal of Operational Research*, 216(2):376 – 385.

Modigliani, F. and Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, 48(3):261–297.

Moro, E., Vicente, J., Moyano, L. G., Gerig, A., Farmer, J. D., Vaglica, G., Lillo, F., and Mantegna, R. N. (2009). Market impact and trading profile of hidden orders in stock markets. *Physical Review E*, 80(6):066102.

Nelsen, R. B. (2007). *An introduction to copulas*. Springer Science & Business Media.

Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic constrained optimization*, pages 272–281. Springer US.

Pflug, G. C. and Pichler, A. (2011). Approximations for probability distributions and stochastic optimization problems. In *Springer Handbook on Stochastic Optimization Methods in Finance and Energy*, volume 163, chapter 15, pages 343–387. Springer.

Pflug, G. C. and Pichler, A. (2014). *Multistage stochastic optimization*. Springer.

Pflug, G. C., Pichler, A., and Wozabal, D. (2012). The 1/n investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36(2):410–417.

Pflug, G. C. and Pohl, M. (2017). A review on ambiguity in stochastic portfolio optimization. *Set-Valued and Variational Analysis*, pages 1–25.

Pflug, G. C. and Römisch, W. (2007). *Modeling, measuring and managing risk*, volume 190. World Scientific.

Pflug, G. C. and Wozabal, D. (2007). Ambiguity in portfolio selection. *Quantitative Finance*, 7(4):435–442.

Pobedrya, B. E. and Georgievskii, D. V. (2006). On the proof of the π-theorem in dimension theory. *Russian Journal of Mathematical Physics*, 13(4):431–437.

Pohl, M., Ristig, A., Schachermayer, W., and Tangpi, L. (2017). The amazing power of dimensional analysis: Quantifying market impact. *To appear in Market Microstructure and Liquidity. Available at arXiv: 1702.05434*.

Pohl, M., Ristig, A., Schachermayer, W., and Tangpi, L. (2018). Theoretical and empirical analysis of trading activity. *arXiv preprint arXiv:1803.04892*.

Pratelli, M. (2011). A remark on the 1/H-variation of the fractional Brownian motion. *Séminaire de Probabilités*, 43:215–219.

Puccetti, G. and Rüschendorf, L. (2012). Computation of sharp bounds on the distribution of a function of dependent risks. *Journal of Computational and Applied Mathematics*, 236(7):1833 – 1840.

Puccetti, G. and Wang, R. (2015). Extremal dependence concepts. *Statistical Science*, 30(4):485–517.

Robert, C. Y. and Rosenbaum, M. (2010). A new approach for the dynamics of ultra-high-frequency data: The model with uncertainty zones. *Journal of Financial Econometrics*, 9(2):344–366.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42.

Rüschendorf, L. (1982). Random variables with maximum sums. *Advances in Applied Probability*, 14(3):623–632.

Rüschendorf, L. (2013). *Mathematical risk analysis*. Springer Berlin Heidelberg.

Rüschendorf, L. (2016). Risk bounds and partial dependence information. *Preprint, Univ. Freiburg*.

Schachermayer, W. and Teichmann, J. (2008). How close are the option pricing formulas of Bachelier and Black-Merton-Scholes? *Mathematical Finance*, 18(1):155–170.

Shorack, G. R. and Wellner, J. A. (2009). *Empirical processes with applications to statistics*, volume 59. Siam.

Steinbach, M. C. (2001). Markowitz revisited: Mean-variance models in financial portfolio analysis. *SIAM review*, 43(1):31–85.

Tauchen, G. E. and Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica*, 51(2):485–505.

Tintner, G. (1941). The theory of choice under subjective risk and uncertainty. *Econometrica: Journal of the Econometric Society*, pages 298–304.

Tóth, B., Lempérière, Y., Deremble, C., de Lataillade, J., Kockelkoren, J., and Bouchaud, J.-P. (2011). Anomalous price impact and the critical nature of liquidity in financial markets. *Physical Review X*, 1(2):021006.

Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems with monotone marginal densities. *Journal of Multivariate Analysis*, 102(10):1344–1360.

Wang, B. and Wang, R. (2015). Extreme negative dependence and risk aggregation. *Journal of Multivariate Analysis*, 136:12–25.

Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin: The Journal of the IAA*, 26(1):71–92.

Wets, R. (1966). Programming under uncertainty: The complete problem. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 4(4):316–339.

Wozabal, D. (2012). A framework for optimization under ambiguity. *Annals of Operations Research*, 193(1):21–47.

Wyart, M., Bouchaud, J.-P., Kockelkoren, J., Potters, M., and Vettorazzo, M. (2008). Relation between bid ask spread, impact and volatility in order driven markets. *Quantitative Finance*, 8(1):41–57.

Xu, H. and Zhang, D. (2012). Monte carlo methods for mean-risk optimization and portfolio selection. *Computational Management Science*, 9(1):3–29.

Zarinelli, E., Treccani, M., Farmer, J. D., and Lillo, F. (2015). Beyond the square root: Evidence for logarithmic dependence of market impact on size and participation rate. *Market Microstructure and Liquidity*, 01(02):1550004.

APPENDIX A

FURTHER REMARKS ON AMBIGUITY IN PORTFOLIO OPTIMIZATION

A.1 Bivariate example

Consider a portfolio consisting of $m = 2$ assets. Assume that it is known that the return of asset j is normally distributed with mean μ_j and variance σ_j^2 for $j = 1,2$. Hence, instead of the empirical distribution function $\hat{F}_n$, we now use the normal cdf in the definition of $\hat{C}_n$. Moreover, we are given n observations of the two asset returns which allow us to compute the empirical estimate $\hat{C}_n$. In accordance with section 1.4, we want to allow for the true dependence structure to be different from the one modeled by $\hat{C}_n$ and make a robust investment decision in this respect. That is we want to solve problem (1.14). The following heuristic allows us to rapidly compute a lower bound C_L and an upper bound C_U as given in (1.15), which translates into bounds for problem (1.14):

Algorithm 1: Heuristic for a bivariate empirical copula

INPUT:

U_0 ... $n \times 2$ matrix of support points $U_{i,j} = F_j(\xi_{i,j})$ for $i = 1, \dots, n$, $j = 1, 2$ of $\hat{C}_n$.
ϵ ... bound for the Wasserstein distance $d_W(\cdot, \cdot)$.
α ... confidence level of the AV@R.

OUTPUT:

U_L ... $n \times 2$ matrix of support points of the lower bound C_L.
U_U ... $n \times 2$ matrix of support points of the upper bound C_U.

$U_{\text{NEW}} = U_0$;
for $l \leftarrow 1$ **to** $\lceil (1-\alpha)m \rceil$ **do**
  Find i_1 and i_2 s.t. $U_{\text{NEW}}(i_1, 1) = U_{\text{NEW}}(i_2, 2) = l/(n+1)$;
  $U_{\text{NEW}}(i_1, 1) = U_{\text{NEW}}(i_2, 1)$;
  $U_{\text{NEW}}(i_2, 1) = l/(n+1)$;
  if $d_W(U_{NEW}, U_0) > \epsilon$ **then**
    $U_L = U_{\text{OLD}}$;
    $U_U = U_{\text{NEW}}$;
    Break
  end
  $U_{\text{OLD}} = U_{\text{NEW}}$;
end

In the following example, we set the parameter values $\mu_1 = 1\%, \mu_2 = 0.2\%, \sigma_1^2 = 2\%, \sigma_2^2 = 0.5\%, \alpha = 0.8$ and $\lambda = 10$ as well as U_0 resp. $\hat{C}_n$ as shown in Figure 1.3. Notice that the exact solution can be computed by enumeration, which is feasible only due to the small values of m and n. Figure A.1 compares the optimal portfolio weights as well as the optimal objective value of the maximin problem (1.14) to the bounds computed with algorithm 1. As it can be seen the investment decision implied by the upper and lower bound might differ significantly from the optimal one for fixed ε. The reason for this observation is shown in figure A.2, where we fix $\varepsilon = 0.003$ and plot the worst case copula $C^\star = \operatorname{argmax}_{C \in \mathscr{C}_{\varepsilon,n}(\hat{C}_n)} \text{AV@R}_\alpha\left(-x^\top \xi^C\right)$ in comparison to the bounds C^L and C^U. Figure A.2 shows that for this particular value of ε the worst case copula does not concentrate on the main diagonal in the lower tail, but C^L and C^U do so by construction. Hence, C_L and C_U differ significantly from $C^\star$ and so do the investment decision implied by these copulas. Nevertheless, it should be clear that the bounds become better for larger n.

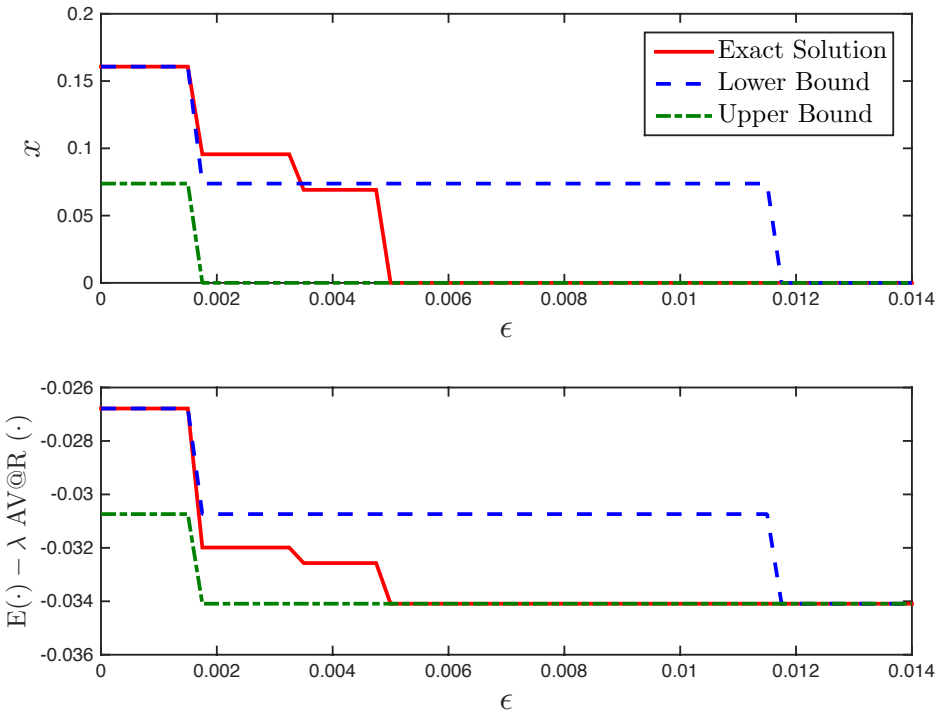


Fig. A.1 These plots compare the performance of the proposed heuristic method to the exact solution. The upper panel shows the optimal weight of the first asset in the portfolio as a function of ε. The lower panel shows the corresponding value of the objective function (1.14) as a function of ε.

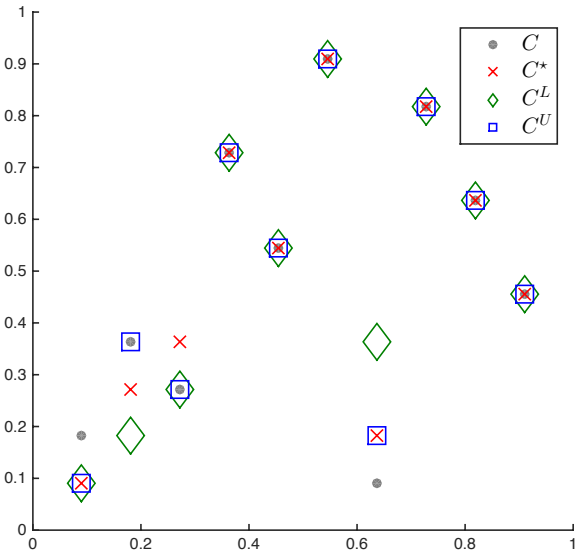


Fig. A.2 The support points of the initial copula $\hat{C}_n$ from figure 1.3 are compared with those of the modified copulas C_L and C_U, which are computed by Algorithm 1, and to those of the "true" worst case copula $C^\star$ for fixed $\varepsilon = 0.003$.

A.2 Minimaxity and saddle points

Let $f(x,y)$ be a function defined on $\mathbb{X} \times \mathbb{Y}$, where $\mathbb{X}$ and $\mathbb{Y}$ are compact topological spaces. We call the pair (x^+,y^+) a saddle point of f, if

$$f(x^+,y) \leq f(x^+,y^+) \leq f(x,y^+),$$

for $x \in \mathbb{X}, y \in \mathbb{Y}$. Assume that $x \mapsto f(x,y)$ is lower semicontinuous for every y and $y \mapsto f(x,y)$ is upper semicontinuous for every x. Let x^* be a minimizer of $x \mapsto \max_y f(x,y)$ and let y^* be a maximizer of $y \mapsto f(x^*,y)$. We say that (x^*,y^*) is a solution of the minimax problem. Notice that (x^*,y^*) does not need to be saddle point. Indeed, let $\mathbb{X} = \{x^*,x\}$ and $\mathbb{Y} = \{y^*,y\}$ and $f(x^*,y^*) = 2$, $f(x^*,y) = 1$, $f(x,y^*) = 1$, $f(x,y) = 3$. Then (x^*,y^*) is a solution of the minimax problem, but not a saddle point. On the other hand, if (x^+,y^+) is a saddle point, then it is a solution. Suppose the contrary, then (x^+,y^+) is a saddle point, but not a solution and (x^*,y^*) is a solution. Let $a = f(x^*,y^*)$, $b = f(x^+,y^*)$, $c = f(x^*,y^+)$, $d = f(x^+,y^+)$. By the saddle point property of (x^+,y^+),

$$b \leq d \leq c,$$

and by the optimality of (x^*, y^*)

$$c \leq a < d.$$

This contradiction proves the assertion. We say that a Minimax Theorem holds, if

$$\min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} f(x,y) = \max_{y\in\mathbb{Y}} \min_{x\in\mathbb{X}} f(x,y).$$

Notice that the existence of a saddle point (x^+, y^+) implies the validity of a Minimax Theorem: Notice first that

$$\min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} f(x,y) \geq \max_{y\in\mathbb{Y}} \min_{x\in\mathbb{X}} f(x,y),$$

naturally holds in any case. Thus it is sufficient to prove the reverse inequality. By the saddle point property of (x^+, y^+)

$$\min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} f(x,y) \quad \leq \quad \max_{y\in\mathbb{Y}} f(x^+,y) = f(x^+,y^+) = \min_{x\in\mathbb{X}} f(x,y^+) \leq \max_{y\in\mathbb{Y}} \min_{x\in\mathbb{X}} f(x,y).$$

However, a one-to-one relation between saddle points and solutions of the minimax problem can only be established under additional assumptions. We cite here Sion's famous Minimax Theorem.

Theorem A.1. *Suppose that f is quasi convex-concave on a topological space $\mathbb{X}\times\mathbb{Y}$, such that $x \mapsto f(x,y)$ is lower semicontinuous for every $y \in \mathbb{Y}$ and $y \mapsto f(x,y)$ is upper semicontinuous for every $x \in \mathbb{X}$. If $\mathbb{X}$ is compact, then*

$$\min_{x\in\mathbb{X}} \sup_{y\in\mathbb{Y}} f(x,y) = \sup_{y\in\mathbb{Y}} \min_{x\in\mathbb{X}} f(x,y).$$

If $\mathbb{Y}$ is compact, then

$$\inf_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} f(x,y) = \max_{y\in\mathbb{Y}} \inf_{x\in\mathbb{X}} f(x,y).$$

If both are compact, then

$$\min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} f(x,y) = \max_{y\in\mathbb{Y}} \min_{x\in\mathbb{X}} f(x,y).$$

Notice that under the assumptions of this Theorem, every solution must be a saddle point. If x^* is a solution of the minimax problem and if $y^* \in \operatorname{argmax}_y f(x^*,y)$, then

$$f(x^*,y) \leq f(x^*,y^*),$$

for all $y \in \mathbb{Y}$. Moreover, since $f(x^*,y^*) = \min_{x\in\mathbb{X}} \max_{y\in\mathbb{Y}} f(x,y) = \max_{y\in\mathbb{Y}} \min_{x\in\mathbb{X}} f(x,y)$,

$$f(x^*,y^*) = \min_{x\in\mathbb{X}} f(x,y^*) \leq f(x,y^*),$$

for all $x \in \mathbb{X}$, that is (x^*,y^*) is a saddle point. In our application, $\mathbb{Y}$ is a subset of Borel probability measures on $\mathbb{R}^m$ endowed with the weak topology. If $\mathbb{Y}$ is a Wasserstein ball, then it is weakly compact (see below). $\mathbb{X}$ is the simplex of nonnegative portfolio weights and also compact. Thus a saddlepoint theorem holds.

Proposition A.2. *Let P_0 be a Borel probability measure on $\mathbb{R}^m$. Then closed Wasserstein balls with center P_0 are weakly compact.*

Proof. Let $\mathscr{P} = \{P : d_W(P,P_0) \leq \eta\}$ be a Wasserstein ball with center P_0. Recall Prohorov's Theorem: If a set of probability measures on a Polish space is weakly closed and tight, it is weakly compact. Closed Wasserstein balls are weakly closed and we have to prove the (uniform) tightness of such a ball, that is: for every ε there is a compact set K such that $P(K) \geq 1-2\varepsilon$ for all $P \in \mathscr{P}$. Since P_0 is Borel, there is a compact set K' such that $P_0(K') \geq 1-\varepsilon$. Let

$$h(x) = \begin{cases} 1 & x \in K' \\ \max\left(1 - \frac{\varepsilon}{\eta} \cdot d(x,K'), 0\right) & x \notin K' \end{cases}.$$

Notice that h is Lipschitz with Lipschitz constant ε/η and therefore

$$\left| \int h(x)\, dP(x) - \int h(x)\, dP_0(x) \right| \leq \frac{\varepsilon}{\eta} d(P,P_0) \leq \varepsilon.$$

Let $K = \{x : d(x,K') \leq \varepsilon/\eta\}$. Then K is compact and

$$P(K) \geq \int h(x)\, dP(x) \geq \int h(x)\, dP_0(x) - \left| \int h(x)\, dP(x) - \int h(x)\, dP_0(x) \right| \geq 1-2\varepsilon,$$

for all $P \in \mathscr{P}$. □

Remark A.1. *With the same argument one sees that closed Wasserstein balls of copulas are weakly compact, since the set of all probability measures on $[0,1]^n$ with uniform marginals is weakly closed and the intersection of a weakly closed and a weakly compact set is weakly compact.*

APPENDIX B

FURTHER REMARKS ON DIMENSIONAL ANALYSIS

B.1 The pendulum

In the setting of Theorem 2.1, somehow surprisingly, the market impact does not depend on the stock price, although – a priori – the price is included in our analysis. There is a simple explanation: The stock price is the only quantity in our analysis involving the "dimension" $\mathbb{U}$ of money. Hence, in the setting of Theorem 2.1 it cannot play a role, because the market impact also does not involve the "dimension" $\mathbb{U}$ of money. In the following, we give a more detailed discussion of this argument in the form of an analogy to the case of the pendulum. In this example, the period also does not depend on the mass of the pendulum which - a priori - is considered as an explaining variable.

Consider a pendulum with length l (measured in meters), mass m (measured in grams) and period (measured in seconds). Assume that the period depends only on l, m and the acceleration g of gravity (measured in meters per seconds squared). That is, we assume that there is a function $f : \mathbb{R}_+^3 \to \mathbb{R}_+$ such that

$$\text{period} = f(l, m, g).$$

From Table B.1, we get the matrices

$$D = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \quad \text{and} \quad c = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

which represent the dimensions of the quantities (l,m,g) and the period respectively in the unit system meter, gram and seconds. As D has full rank, it follows from Corollary 2.4 that

$$\text{period} = \text{const} \cdot l^{y_1} m^{y_2} g^{y_3},$$

for some const > 0, where the unique solution of the linear system $Dy = c$ is given by $y = (1/2, 0, -1/2)^\top$. Thus,

$$\text{period} = \text{const} \cdot \sqrt{\frac{l}{g}}. \tag{B.1}$$

Why does this solution not involve the variable m? The answer is given by looking at the second row of D and the second coordinate of c, which forces y_2 to equal zero. This is perfectly analogous to the role of the variable P, i.e., the price of the stock, in the setting of Theorem 2.1.

Next, we shall illustrate the difference between Theorem 2.1 and Theorem 2.2 by discussing an analogous variation of the assumptions in the case of the pendulum. The crucial assumption in the reasoning above was that the period of the pendulum is completely determined by its length, its mass and the acceleration due to gravity. However, it is conceivable (and, in fact, the case) that the period also depends on other variables, e.g. the amplitude a (measured in meters) of the observed swing. In other words, we might also start from the weaker assumption

$$\text{period} = f(l,m,g,a). \tag{B.2}$$

The matrix containing the dimensions of the observed quantities is now given by

$$\widetilde{D} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 \end{pmatrix}.$$

It follows from Corollary 2.5 that the general form of the relation (B.2) is given by

$$\text{period} = l^{y_1} m^{y_2} g^{y_3} a^{y_4} h(l^{x_1} m^{x_2} g^{x_3} a^{x_4}),$$

for some function $h : \mathbb{R}_+ \to \mathbb{R}_+$, where $x = (x_1, x_2, x_3, x_4)^\top$ is a solution of the homogeneous system $\widetilde{D}x = 0$ and $y = (y_1, y_2, y_3, y_4)^\top$ a solution of the inhomogeneous system $\widetilde{D}y = c$.

	l	m	g	a	period
length	1	0	1	1	0
mass	0	1	0	0	0
time	0	0	-2	0	1

Table B.1 A labelled overview of the matrix summarizing the dimensions of the quantities considered to determine the period of a pendulum.

Choosing

$$x = \left(\frac{1}{2}, 0, -\frac{1}{2}, 0\right)^\top \qquad \text{and} \qquad y = (1,0,0,-1)^\top,$$

we obtain

$$\text{period} = \sqrt{\frac{l}{g}}\, h\left(\frac{l}{a}\right). \tag{B.3}$$

In the setting of (B.2), dimensional analysis does not allow to determine the function h. In order to do so, we need some additional information. In physics we have the possibility of experiments. Already Galileo noticed the – at first glance surprising – experimental result that the period of the pendulum does *not* (at least not strongly) depend on the amplitude. Using this insight from experimental physics, we conclude that $h \equiv \text{const}$ is a physically reasonable choice in the general solution (B.3).

We conclude this discussion by making the analogy to the case of Theorem 2.1 and Theorem 2.2 above. Dimensional analysis alone cannot decide whether the special solution given by Theorem 2.1 is the "true" relation between the market impact and the relevant variables, or whether some other explanatory variables as provided, e.g., by (2.14), (2.15), or (2.17) yield the "true" relation. To answer this question one has to take recourse either to economic theory or to empirical analysis. This is analogous to the above discussed situation of the pendulum where *physical experiments* yield that the special case (B.1) of the more general solution (B.3) is in fact the "true" relation (as long as the amplitude remains within reasonable bounds).

To round up this discussion, we give an example how dimensional analysis can lead astray, if applied blindly. Start with the (silly) assumption that the period of the pendulum depends *only* on the mass m, the acceleration g, and the amplitude a so that

$$\text{period} = f(m,g,a),$$

for some function $f : \mathbb{R}_+^3 \to \mathbb{R}_+$. Repeating verbatim the analysis preceding (B.1) we obtain

$$\text{period} = \text{const} \cdot \sqrt{\frac{a}{g}}, \tag{B.4}$$

as the unique solution satisfying the invariance properties of dimensional analysis. But, of course, the solution (B.4) is far from physical reality. The reason is that we have chosen *a wrong set of explanatory variables*. In other words, dimensional analysis only yields reasonable solutions if the set of explanatory variables is well chosen and really contains essentially all the information necessary to determine the quantity of interest.

B.2 Proof of the Pi-Theorem

Proof of Theorem 2.3: We pass to logarithmic coordinates by using the following notation: Given $Z \in \mathbb{R}_+$ we shall write $\widetilde{Z} = \log(Z)$. On the logarithm scale $\widetilde{U}$ satisfies

$$\widetilde{U} = g\left(\widetilde{W}_1, \ldots, \widetilde{W}_n\right), \tag{B.5}$$

for some function $g : \mathbb{R}^n \to \mathbb{R}$.

If the left and the right hand sides of (B.5) do not depend on L_1, then it is sufficient to work with the units $(L_2, \ldots, L_m)$. On the other hand, if $a_1 \neq 0$ and $b_{11} = \cdots = b_{1n} = 0$ then $g \equiv 0$. If there exists $i \in \{1, \ldots, n\}$ such that $b_{1i} \neq 0$, we assume without loss of generality that $b_{11} \neq 0$. Putting $\widetilde{V} := -\frac{a_1}{b_{11}}\widetilde{W}_1 + \widetilde{U}$ and $\widetilde{X}_{i-1} := -\frac{b_{1i}}{b_{11}}\widetilde{W}_1 + \widetilde{W}_i$, $i = 2, \ldots, n$ we have

$$\begin{aligned}\widetilde{V} &= -\frac{a_1}{b_{11}}\widetilde{W}_1 + g\left(\widetilde{W}_1, \frac{b_{1i}}{b_{11}}\widetilde{W}_1 + \widetilde{X}_1, \ldots, \frac{b_{1n}}{b_{11}}\widetilde{W}_1 + \widetilde{X}_{n-1}\right) \\ &= f\left(\widetilde{W}_1, \widetilde{X}_1, \ldots, \widetilde{X}_{n-1}\right),\end{aligned}$$

for some function f. Let $\lambda \in \mathbb{R}$ and put $L_1^* := e^{\lambda} L_1$. Since the dimensions of $\widetilde{V}$ and $\widetilde{X}_{i-1}$, $i = 2, \ldots, n$ are given in terms of the units $(L_1, \ldots, L_m)$, the quantities $\widetilde{V}, \widetilde{X}_1, \ldots, \widetilde{X}_{n-1}$ remain unchanged upon passing to the system of units $(L_1^*, L_2, \ldots, L_m)$. On the other hand, $\log([W_1]) = -\lambda b_{11} + b_{11}\widetilde{L}_1^* + \sum_{i=2}^m b_{i1}\widetilde{L}_i$ so that in the system of units $(L_1^*, L_2, \ldots, L_m)$ it holds

$$\widetilde{V} = f\left(\lambda b_{11} + \widetilde{W}_1, \widetilde{X}_1, \ldots, \widetilde{X}_{n-1}\right).$$

Since λ was taken arbitrary, f does not depend on the first component, that is,

$$\widetilde{V} = f\left(\widetilde{X}_1, \ldots, \widetilde{X}_{n-1}\right). \tag{B.6}$$

By repeating the argument $\operatorname{rank}(B)-1$ times, we obtain

$$\widetilde{U}-\sum_{j=1}^{n} y_j\widetilde{W}_j = h\left(\sum_{j=1}^{n} x_{1j}\widetilde{W}_j,\ldots,\sum_{j=1}^{n} x_{kj}\widetilde{W}_j\right). \tag{B.7}$$

In fact, since $x^{(i)}$ is a solution of the homogeneous system $Bx=0$, the quantity $\sum_{j=1}^{n} x_{1j}\widetilde{W}_j$ is dimensionless. Notice that $Bx=0$ has $k=n-\operatorname{rank}(B)$ linearly independent solutions. Similarly, since y is a solution of the inhomogeneous system $By=0$, the left hand side in (B.7) is a dimensionless quantity. Hence, there is a function $F:\mathbb{R}_+^k\to\mathbb{R}_+$ such that

$$U\cdot W_1^{-y_1}\cdots W_n^{-y_n} = F(\pi_1,\ldots,\pi_n),$$

with $\pi_j := W_1^{x_{1j}}\cdots W_n^{x_{nj}}$, $j=1,\ldots,k$. □

Proof of Corollary 2.4: First notice that if $k=0$ then $n=\operatorname{rank}(B)$. Repeating the argument leading to (B.6) $\operatorname{rank}(B)-2$ times, we can find $(z_1,\ldots,z_{n-1})^\top\in\mathbb{R}^{n-1}$, a quantity X with dimension $[X]=L_m^{\alpha_m}$ and a function $f:\mathbb{R}\to\mathbb{R}$ such that

$$\widetilde{Y} := \widetilde{U}-\sum_{j=1}^{n-1} z_j\widetilde{W}_j = f(\widetilde{X}).$$

Let us denote by $L_m^{c_m}$ the dimension of Y. We can assume without loss of generality that $\alpha_m\neq 0$. As in the proof of Theorem 2.3, we have

$$\widetilde{V} := \widetilde{U}-\sum_{j=1}^{n-1} z_j\widetilde{W}_j-\frac{c_m}{\alpha_m}X = -\frac{c_m}{\alpha_m}X+f(\widetilde{X}) = g(\widetilde{X}),$$

for some function g. Let $\lambda\in\mathbb{R}$ and put $L_m^* = e^{\lambda}L_m$. Since $\widetilde{V}$ is dimensionless, its value does not change when passing to the unit L_m^*. On the other hand, $\log([X]) = -\alpha_m\lambda+\alpha_m L_m^*$. Hence, with respect to the unit L_m^* we have

$$\widetilde{V} = g(\alpha_m\lambda+X).$$

Since λ was taken arbitrary, the function g must be a constant. Thus, there is $\text{const}>0$ such that $U=\text{const}\cdot W_1^{y_1}\cdots W_n^{y_n}$, since the right hand side of the latter equation has the dimension of U, if and only if $By=a$. □

Proof of Corollary 2.5: The result follows from a direct application of Theorem 2.3. □

B.3 Proofs of Sections 3.2 and 3.4

In this section, we provide formal arguments for the results presented in Sections 3.2 and 3.4. The proofs are based on the above particular cases of the Pi-Theorem.

Proof of Proposition 3.1. Combining relation (3.1) and the dimensions of the quantities σ^2, P, V and N, we obtain that the matrix B as well as the vector a are given by

$$B = \begin{pmatrix} 0 & -1 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & -1 \end{pmatrix} \quad \text{and} \quad a = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}.$$

Table 3.1 illustrates how B and a relate to the considered quantities and their dimensions. As the matrix B has full rank, i.e., $\operatorname{rank}(B) = 3$, applying Corollary 2.4 yields

$$N = c \cdot \sigma^{2y_1} P^{y_2} V^{y_3},$$

for some constant $c > 0$, where $y = (y_1, y_2, y_3)^\top$ is the unique solution of the linear system $By = a$ which is given by $y = (1,0,0)^\top$. □

Proof of Relation (3.4). Combining relation (3.3) and the dimensions of the quantities σ^2, P, V and C as well as N, the matrix B as well as the vector a become

$$B = \begin{pmatrix} 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & -1 & 0 \end{pmatrix} \quad \text{and} \quad a = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}.$$

The vector $x = (-1,1,1,-1)^\top$ is a solution of the homogeneous system $Bx = 0$, and the vector $y = (1,0,0,0)^\top$ is a solution of the inhomogeneous system $By = a$. Thus, relation (3.4) follows from Corollary 2.5. □

Proof of Theorem 3.2. Combining the dimensions of the quantities considered in relation (3.5) and the "Leverage Neutrality Assumption", we obtain that the matrix B as well as the vector a are given by

$$B = \begin{pmatrix} 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -1 & 0 & -1 & 0 \\ 2 & -1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad a = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \end{pmatrix}.$$

As the matrix B has full rank, i.e., $\operatorname{rank}(B) = 4$, applying Corollary 2.4 yields

$$N_t = c \cdot \sigma_t^{2y_1} P_t^{y_2} V_t^{y_3} C_t^{y_4},$$

for some constant $c > 0$, where $y = (y_1, y_2, y_3, y_4)^\top$ is the unique solution of the linear system $By = a$ which is given by $y = (1/3, 2/3, 2/3, -2/3)^\top$. □

Proof of Corollary 3.3. Considering the dimensions of the quantities σ_B, V, C, we obtain that the matrix B as well as the vector a are given by

$$B = \begin{pmatrix} -2 & 1 & 0 \\ 2 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix} \quad \text{and} \quad a = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}.$$

As the matrix B has full rank, i.e., $\operatorname{rank}(B) = 3$, applying Corollary 2.4 yields

$$N = c \cdot V^{y_1} \sigma_B^{y_2} C^{2y_3},$$

for some constant $c > 0$, where $y = (y_1, y_2, y_3)^\top$ is the unique solution of the linear system $By = a$ which is given by $y = (1/3, 2/3, -2/3)^\top$. This shows (3.14). □

Proof of Corollary 3.4. As explained before the statement of Corollary 3.4, the conditions (3.5) and (3.15) are equivalent. Thus, it holds

$$N^{3/2} = c \cdot \frac{\sigma_B V}{C}.$$

Since $C = SV/N$, the corollary follows. □

Proof of Proposition 3.5. The proof is the same as that of Theorem 3.2 except that in the present case the matrices B and a are given by

$$B = \begin{pmatrix} 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -2H & 0 & -1 & 0 \\ 2 & -1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad a = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 0 \end{pmatrix}.$$

The unique solution y of the linear system $By = a$ is $y = 1/(1+H) \cdot (1/2, 1, 1, -1)^\top$. Applying Corollary (2.4) gives the desired result. □

B.4 Integer part of Brownian motion

With the notation from Section 3.4, we want to show that as $T \searrow 0$

$$\mathbb{V}\mathrm{ar}\left(\log(\check{P}_{t+T}) - \log(\check{P}_t)\right) \approx c\sqrt{T},$$

for some constant $c > 0$. Recall that $\left(\log(\check{P}_t)\right)_{t\geq 0}$ is given by

$$\log(\check{P}_t) = \mathrm{int}(W_t),$$

where $(W_t)_{t\geq 0}$ is a standard Brownian motion and $\mathrm{int}(x)$ denotes the integer closest to the real number x, i.e., $\mathrm{int}(x) = \sup\{n \in \mathbb{Z} : n \leq x + 0.5\}$.

To present the idea in its simplest possible form, note that for fixed $t > 0$, say $t = 1$ and T small, it is straightforward to verify that

$$\left(\log(\check{P}_{t+T}) - \log(\check{P}_t)\right)^2 = \left(\mathrm{int}(W_{t+T}) - \mathrm{int}(W_t)\right)^2 = \begin{cases} 0 & \text{with probability of order 1,} \\ 1 & \text{with probability of order } T^{1/2}, \\ >1 & \text{with probability smaller than } T. \end{cases}$$

So that $\mathbb{V}\mathrm{ar}\left(\log(\check{P}_{t+T}) - \log(\check{P}_t)\right)$ is of order $T^{1/2}$, as $T \searrow 0$, rather than of the usual order T. In the above sketchy argument we used the fact that, for every $t > 0$,

$$\lim_{h\to 0} \frac{1}{h}\, \mathbb{P}\left(\min_{n\in\mathbb{Z}} |W_t - n| \leq h\right) \geq c,$$

for some constant $c > 0$.

To furnish a more precise result, we make - contrary to our usual assumption $W_0 = 0$ - the assumption that the Brownian motion starts from a random variable W_0 which is uniformly distributed on $[-1/2, 1/2]$. Then, we can formulate the following more quantitative result for fixed $t = 0$.

Proposition B.1. *Assume that W_0 is uniformly distributed on $[-1/2, +1/2]$. Then,*

$$\liminf_{T\to 0} \sqrt{\frac{\pi}{2T}}\, \mathbb{V}\mathrm{ar}\left(\log(\check{P}_T) - \log(\check{P}_0)\right) = 0.$$

Proof. Note that

$$\mathbb{V}\mathrm{ar}\left(\log(\check{P}_T) - \log(\check{P}_0)\right) = \mathbb{E}\left[\left(\log(\check{P}_T) - \log(\check{P}_0)\right)^2\right],$$

where $\log(\check{P}_0)$ is in fact zero as we assumed that $W_0 \sim \mathrm{Uni}(1/2, 1/2)$. In the following $(B_t)_{t\geq 0}$ denotes a standard Brownian motion starting at $B_0 = 0$ such that $W_T = B_T + W_0$. Then,

$$\begin{aligned}
\mathbb{E}\left(\left(\log(\check{P}_T) - \log(\check{P}_0)\right)^2\right) &= \int_{-0.5}^{0.5} \mathbb{E}\left(\left(\mathrm{int}(B_T + x)\right)^2\right) dx \\
&= \int_{-0.5}^{0.5} \sum_{i=1}^{\infty} i^2 \left(\mathbb{P}\left(\frac{2i-1}{2} - x \leq B_T \leq \frac{2i+1}{2} - x\right)\right. \\
&\qquad \left. + \mathbb{P}\left(-\frac{2i+1}{2} - x \leq B_T \leq -\frac{2i-1}{2} - x\right)\right) dx \\
&= \int_{-0.5}^{0.5} \sum_{i=1}^{\infty} i^2 \left(\Phi\left(\frac{i+0.5-x}{\sqrt{T}}\right) - \Phi\left(\frac{i-0.5-x}{\sqrt{T}}\right)\right. \\
&\qquad \left. + \Phi\left(\frac{i+0.5+x}{\sqrt{T}}\right) - \Phi\left(\frac{i-0.5+x}{\sqrt{T}}\right)\right) dx \\
&= \sum_{i=1}^{\infty} i^2 \left(\sqrt{\frac{2T}{\pi}}\left(\exp\left(-\frac{(i+1)^2}{2T}\right) + \exp\left(-\frac{(i-1)^2}{2T}\right) - 2\exp\left(-\frac{i^2}{2T}\right)\right)\right. \\
&\qquad \left. + (2i+2)\Phi\left(\frac{i+1}{\sqrt{T}}\right) + (2i-2)\Phi\left(\frac{i-1}{\sqrt{T}}\right) - 4i\Phi\left(\frac{i}{\sqrt{T}}\right)\right) \\
&= \sqrt{\frac{2T}{\pi}}\left(1 + 2\sum_{i=1}^{\infty} \exp\left(-\frac{i^2}{2T}\right)\right) - 4\sum_{i=1}^{\infty} i\Phi\left(-\frac{i}{\sqrt{T}}\right)
\end{aligned}$$

We now use that fact for $x \to \infty$, $\Phi(-x) \approx \phi(x)/x$, where $\phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$ is the probability density function of the standard normal distribution (we thank Friedrich Hubalek for pointing this out to us). It follows that for small T

$$i\Phi\left(-\frac{i}{\sqrt{T}}\right) \approx \sqrt{\frac{T}{2\pi}} \exp\left(-\frac{i^2}{2T}\right),$$

which concludes the proof. □

ABSTRACT

This thesis consists of two parts. The first one is devoted to robust portfolio optimization. We review tractable reformulations of the mean-risk portfolio selection problem under model ambiguity. In particular, we study the portfolio selection problem when all probability distributions contained in a Wasserstein-neighborhood of some reference model are taken into account. It is then an original contribution of this dissertation that we introduce the portfolio selection problem under dependence uncertainty. We assume the marginal return distributions of the individual assets are known and that the model ambiguity lies solely in the dependence structure between the assets. We show theoretically and empirically that under high model ambiguity (in the respective sense) the two approaches have diametrically opposed implications: Portfolio diversification is optimal under high ambiguity with respect to the joint distribution, whereas portfolio concentration is optimal under high dependence uncertainty.

In the second part of this dissertation, we use dimensional analysis to study relations between financial quantities. Firstly, we apply this concept, which is well known in physics, to derive the following remarkable fact. If the market impact of a meta-order only depends on four well-defined and financially meaningful variables, then there is only one possible functional form of this dependence. In particular, the market impact is proportional to the square-root of the size of the meta-order. This result is based on three scaling invariances with respect to the units in which the considered quantities are measured as well as on the assumption of leverage neutrality. The latter is a restrictive assumption on the behavior of the considered quantities when the corresponding stock is paying dividends. Secondly, we apply the same line of argument in the context of trading activity. Different combinations of the relevant explanatory variables result in different proportionality relations, which have been proposed in the literature. Hence, the question which of the derived relations describe the reality boils down to which set of variables indeed fully explains the variables of interest - in our case the number of trades. Hence, we perform an empirical analysis to examine the fit of the considered relations to data from the stock market.

ZUSAMMENFASSUNG

Die vorliegende Arbeit besteht aus zwei Teilen. Im ersten Teil behandeln wir robuste Portfolio Optimierung. Dabei wird in die Berechnung der optimalen Anlagestrategie mittels stochastischer Modellierung der Aktienrenditen auch die Unsicherheit bezüglich des dazu verwendeten Modells einbezogen. Üblicherweise wird diese sogenannte Modellunsicherheit hinsichtlich der gemeinsamen Verteilung der Aktien betrachtet. Als Erweiterung der aus der Literatur bekannten Ansätze führen wir Portfoliooptimierung unter Unsicherheit bezüglich der Zusammenhangsstruktur zwischen den Aktien ein. Konkret lautet unsere Annahme, dass nur die Randverteilungen der einzelnen Aktien bekannt sind. Wir vergleichen die verschiedenen Implikationen der beiden genannten Ansätze: Während Diversifizierung optimal ist bei höchstmöglicher Unsicherheit bezüglich der gemeinsamen Verteilung, stellt sich eine Konzentration des Portfolios in eine einzige Aktie als optimal heraus, falls die Modellunsicherheit nur die Zusammenhangsstruktur betrifft.

Im zweiten Teil der Arbeit wenden wir Dimensionsanalyse, eine aus der Physik bekannte Methode, zur Studie von Wechselwirkungen zwischen wichtigen Variablen am Finanzmarkt an. Bei einer sogenannten Meta-Order wird eine Handelsentscheidung in mehreren kleinen Tranchen umgesetzt. Die Preisauswirkung einer solchen Transaktion versuchen wir durch vier relevante Variablen zu erklären. Wir beweisen unter anderem, dass diese Preisauswirkung proportional zur Quadratwurzel der Anzahl gehandelter Aktien ist. Unsere Resultate basieren auf drei Skaleninvarianzen bezüglich der Einheiten, in denen die gegebenen Variablen gemessen werden, sowie auf einer Annahme bezüglich des Verhaltens der Variablen wenn Dividenden ausgeschüttet werden.

Weiters wenden wir dieselben Argumente an, um die Handelsaktivität zu analysieren. Verschiedene Annahmen bezüglich der Auswahl der Variablen, von denen die Handelsaktivität abhängt, ergeben verschiedene Zusammenhänge. Daher hängt die Richtigkeit der hergeleiteten Gesetzmäßigkeiten davon ab, ob die gewählten erklärenden Variablen die Handelsaktivität tatsächlich zur Gänze beschreiben. Diese Frage wird anhand einer empirischen Studie diskutiert.