Essays on Non-stationarity in Finance

Dissertation

zur Erlangung des Grades

Doktor der Wirtschaftswissenschaft (Dr. rer. pol.)

der Juristischen und Wirtschaftswissenschaftlichen Fakultät
der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von
Jakob Krause

Halle (Saale),
Februar 2020

1. Gutachter: Prof. Dr. Jörg Laitenberger

2. Gutachter: Prof. Dr. Gregor Weiß

Tag der Verteidigung: 10.11.2020

Natura quidem suas habet consuetudines, natas ex reditu causarum, sed non nisi $\omega\zeta$ $\varepsilon\pi\iota$ τo $\pi o\lambda\nu$.
[Nature has established patterns originating in the return of events [latin in original], *but only for the most part [greek in original].*]

Gottfried Wilhelm Leibniz in a letter to Jacob Bernoulli, Dec 3, 1703

Acknowledgement

I thank Professor Jörg Laitenberger for providing me with the necessary academic freedom, personal support and a managable teaching load that have made this thesis possible.

I thank the University of Halle for generous funding for conference attendances.

I thank all my colleagues, especially Yassine, Steffen, Pit, Hannes, Max, Luise and Doris for being great intellectual sparring partners and/or for making a highly enjoyable working environment and/or for being general entertainment factors in not always easy times.

I thank my parents and grandparents for their continuous encouragement and support throughout my study time.

I thank Miriam for being who she is.

Contents

List of Figures

List of Tables

1 Introduction

Moving from a Ptolemaean to a Copernican worldview has been one of the most important developments in the history of modern science and western thought (see Kuhn (1957)). The proposition that our place in the universe is not 'special' and that we are subject to the same laws of motion as the astronomical bodies around us is deeply thought provoking and has been viewed as one of the most important catalysts of the renaissance and the scientific revolution: Kuhn argues *'The significance of 'De revolutionibus [orbium coelestium]' lies [...] less in what it says itself than what it caused others to say [...] it is a revolution-making rather than a revolutionary text [...].'* In addition to its important conceptual contribution, it also serves as one of the prime early examples where observations are used to reject a theory and propose a new one that explains the observations in a better way.

Today, in any quantitative scientific field, researchers are interested in the derivation of underlying principles from data. By relying on quantification and statistical analysis it is possible to deduce aspects of a potential underlying mechanism. However, when one is performing such an analysis, one is limited by the aspects that are visible to us through our statistical techniques. Statistical techniques only work properly in the situations for which they have been designed and their limitations are determined by their specific set of assumptions. In order to arrive at viable statistical conclusions it is necessary to build the statistical techniques used in a data analysis on structural properties of the data set itself. Typically, the more structure one imposes on the object of study ex ante, the easier the job of analyzing it. This leads to an incentive to oversimplify the analysis by imposing too much assumptions up front. In addition, it leads to a deep conceptual relationship between statistical techniques and the problems for which they have been designed.

The machinery of modern statistics has been developed with physics in mind. The corresponding set of assumptions that one can reasonably impose on a physical system ex ante is extensive: One can credibly argue that neither time nor location play a role when performing the same controlled experiment since the laws of nature are typically considered to be universal, i.e. stable across space and time. Due to the conceptual contribution of 'De revolutionibus', which questions any exceptionality of our current perspective, this concept is now known as the 'Copernican Principle' (see Bondi and Gold (1952)). In such a setup, every data point is a manifestation of the same underlying principles. Furthermore, under the assumption of universal laws of nature it is possible to repeat experiments and, as a consequence, it is possible to collect arbitrarily large data sets for any statistical analysis. Therefore, any statistical technique leading to definite conclusions with arbitrarily large data sets can be used in order to arrive at definite conclusions of (some aspects of) the underlying principles governing the outcome of any experiment. Hence, there is a very tight connection between the repeatability of an experiment, universal underlying laws, and the possibility of deducing those principles via statistical techniques (see Uzan (2003)).

In physics it has been realized early on that this methodological setup contains a blind spot which may be a detriment to a thorough understanding of our environment. By assuming that time and location of an experiment are irrelevant, we close ourselves to even the possibility of seeing the opposite. The 'Copernican Principle' might have conceptually enabled the scientific revolution but it also contains a universality assumption which is an institutionalized blind spot. Consequently, it is not surprising that the possibility that some aspects of the laws of nature are not universal has been discussed in physics literature. In fact, it was initiated by none other than Paul Dirac (see Dirac (1937), Dirac (1938)). This idea lead to a decade-long research inquiry culminating in a variety of theoretical and experimental results proving Dirac's suspicion was right (see Uzan (2003) for a review of the literature summarizing 70 years of experimental and theoretical developments on this topic). In view of these results it is possible that - even in physics - past experimental evidence might not be representative of the current state of the underlying principles governing the system and that the 'Copernican Principle' does not hold. While it is credible and reasonable to assume that the change in the laws of nature is negligible for practically all situations of interest, it is not clear to which degree this effect can be ignored in other fields. Furthermore, since most of our statistical techniques have been developed for physics applications it is also necessary to explore the impact of this effect of potential variability on those techniques and their applicability in those other fields. The objective of this thesis is to start with this investigation. In this endeavour I start from the fundamental assumption that the object of study is genuinely changing over time, i.e. a conceptual violation of the 'Copernican Principle'. In addition, I assume that one can develop an opinion about the degree to which the object changes, but the argument developed here works under any arbitrary dynamic of this type. As a special case this includes the opinion that the object does not change, i.e. classical 'asymptotic theory' leading to definite conclusions with arbitrarily large data sets is conceptually imbedded in the framework of this thesis. Since scientific exploration starts with observation, the interesting question in this context is how to optimally observe such a changing system. The main contribution of this thesis is to introduce a methodology that is able to identify an optimal data quantity which balances the statistical benefits of more data with representativeness issues caused by instability of the underlying laws. The other parts of this thesis address generalizations and implications of the data identification method and its conceptual impact on modern finance theory. The modern economic and financial system is undoubtedly one of the most profound innovations of humanity and has objectively increased the standard of living for billions of people worldwide over hundreds of years. Given the profound role this system plays in governing our lives, it is of paramount importance to explore potentially faulty reasoning in its intellectual underpinnings. The objective of the next section is to document what I believe is such a mistake which is directly related to the argument developed so far.

1.1 The role of time in Finance and Economics

1.1.1 Ergodicity

The usefulness of statistics is based on the presumption that the past is representative of the present and the future. If one accepts the Copernican Principle, then the past is fully representative of the present. In this context it is useful to distinguish between temporary and non-temporary effects. In 1884 Ludwig Boltzmann introduced the corresponding statistical assumption, called the 'ergodicity' property, in order to precisely analyze, in mathematical terms, the properties of a closed thermodynamical system, i.e. a system that does not exchange any energy with its environment (see Boltzmann (1884)). Such a system will typically settle in a state of maximum entropy called the 'thermodynamical equilibrium' after some time. In such a state the properties of the system are characterized solely by the properties of the gas molecules (velocities and positions in phase space, i.e. the space of possible states of the physical system, here the 6-dimensional real numbers). If one is interested in the analysis of those properties this greatly simplifies the problem. Boltzmann literally called the ergodicity assumption a 'trick' [Kunstgriff] in his paper. Since entropy is always increasing and a small contained system eventually reaches its 'state of maximum entropy', all the temporary effects can be accounted for by simply waiting a while. This argument justifies an assumption that the state of the system is independent of its initial condition and, if one also accepts that the fundamental principles governing the system are stable over time, then the current state of maximum entropy is representative of the past and the future. Hence, the concept of thermodynamical equilibrium reduces the number of parameters that are relevant to fully describe a system and the assumption of ergodicity allows us to learn the properties of a given system over time. Formally, ergodicity can be defined as follows (after Kirstein (2015) who, in turn, relies on Lebowitz and Penrose (1973)):

Definition

A system is called **ergodic** if the following equality holds for all T and $t \leq T$:

$$\lim_{t \to -\infty} \frac{1}{T-t} \int_t^T f(\varphi_m(x)) \, dm = \frac{\int_S f(x)\mu(dx)}{\mu(S)} \,, \tag{1}$$

where S is the energy surface of the system in phase space, φ_m is the state of the system at time m, μ is the measure which describes the volume in phase space and f is an arbitrary μ-integrable real valued function.

Intuitively, in ergodic systems time averages (left) are equal to ensemble averages (right), or, in other words, the state space at one point in time (right) can be deduced by observing the history of the system (left).

1.1.2 Ergodicity and Economic Methodology

Given the considerations so far, it is somewhat surprising that the ergodicity assumption made it into the economic toolbox at all. The degree to which the past is representative of the present is very different in physics and in economics and, consequently, the criticality of the assumption is very different. Given the degree of the problem one would expect that economists should, by sheer transitivity, give more attention to the problem than physicists. Unfortunately, this is not the case. In fact, Paul Samuelson famously defended the ergodic assumption as follows:

'[...] interesting [...] assumption implicit and explicit in the classical mind. It was a belief in unique long-run equilibrium independent of initial conditions. I shall call it the 'ergodic hypothesis' by analogy to the use of this term in statistical mechanics. [...] Now, Paul Samuelson, aged 20 [...] as an equilibrium theorist he naturally tended to think of models in which things settle down to a unique position independently of initial conditions. Technically speaking, we theorists hoped not to introduce hysteresis phenomena into our model, as the Bible does when it says 'We pass this way only once' and, in so saying, takes the subject out of the realm of science into the realm of genuine history. [...] we envisaged an oversimplified model with the following ergodic property: no matter how we start [...] after a sufficiently long time it will become [...] a unique ergodic state.'

(see Samuelson (1968), p.12). In this quote Samuelson equates the ergodicity assumption with the *possibility* of economics to be a science. This is a comment he, reportedly, made repeatedly (see North (1999)). For a person who is responsible like no other for transposing physical and statistical methodology into economics (see, Lo (2017) and Kirstein (2015)) his indifference towards the cautionary qualifications in the field from which he obtains his tools might be surprising. However, from a practical perspective it is highly understandable why Samuelson's position has such an appeal: In an ergodic system, arguments can be 'objectively' settled, markets can 'efficiently' incorporate information and one can 'optimize' decisions along various dimensions. Consequently, in ergodic systems, one can answer questions simply by looking at the data and one can credibly argue that any question can be settled if only there would be enough data.

At this point it has to be noted that even Samuelson himself apparently did not fully trust his own work on efficient markets: In a 2010 account of the hedge fund industry (Mallaby (2010)) reports in a chapter entitled 'Paul Samuelson's Secret' that Samuelson was invested heavily in a prominent hedge fund ('Commodities Corporation') specialising first in fundamental research, which is consistent with the weak form of the efficient market hypothesis, and later (with an additional investment commitment of Samuelson) in what we today would call 'chart trading', which under Samuelson's own Efficient Market Hypothesis would not be considered a viable investment business: *'Do I really believe what I have been saying? I would like to believe otherwise. But a respect for evidence compels me to* *incline toward the hypothesis that most portfolio decisionmakers should go out of business, take up plumbing, teach Greek [and Latin], or help produce the annual GNP by serving as corporate executives. Even if this advice to drop dead is good advice, it obviously is not counsel that will be eagerly followed.'* (see Samuelson (1974)). Given the unconvincing intellectual history of the idea (see Kirstein (2015)) and the personal conduct of even its most ardent proponents, a variety of economists have tried to work without the ergodicity assumption (most notably Keynes (1936)) founding the school of Postkeynesianism in the process, see, e.g., Davidson (2012) but see, e.g., also North (2005)). Their post-ergodic position, however, is typically unmarketable since their ergodic counterparts have numbers and quantitative models to back up their claims and, as such, can fulfill better the 'demand for precision' which is one of the features of Economics in the 20th century (Bernstein (1998)). A more compelling method to overcome the ergodic assumption would be working out the viability of and eventually establishing a quantitative non-ergodic model of economic activity. A quantitative non-ergodic model would illustrate the shortcomings of the perceived precision of the ergodic model while simultaneously competing with it in terms of being 'quantitative'.

1.1.3 The role of data - Efficient and Adaptive Markets

One of the implications of a non-ergodic environment is that one cannot fully trust data from the past. Consequently, due to the amount of data being limited it is impossible to fully determine the current state of the world around us based on statistical analysis alone. In this context, any definition that is based on actions assumed to be determined by knowledge received through some statistical techniques must be checked for their viability. For example, in Finance and Economics the most prominent definition of this kind is 'market efficiency' which states that prices reflect 'available information'. Here, 'available information' is typically considered to be quantifyable statistical properties. If there is a limit on 'statistical perception' this definition cannot reliably differentiate between given situations. In fact, as part of this thesis I will show that this definition is not even well-defined when one replaces 'available information' by 'optimal available information' since the data set yielding the 'optimal available information' may be not unique (see chapter

2).

Hence, in a non-ergodic world 'market efficiency' is conceptually flawed. Furthermore, given the recent failures of current quantitative techniques (e.g. in the 2008 financial crisis as well as earlier, such as when a nobel laureate staffed hedge fund brought the financial system to the brink of collapse in 1998, see Lowenstein (2000)) that are conceptually based on ergodicity, there has been much support for an adaptive further development of economic and financial methodology (see, e.g., Colander *et al.* (2010), Mirowski (2013)).

In this thesis I will rely conceptually on the 'Adaptive Market Hypothesis' where the term 'adaptive markets' refers to *'the multiple roles that evolution plays in shaping behavior and financial markets'* (see Lo (2017)). In quantitative circles this literature is relatively new but the call to base our understanding of finance and economics on concepts from biology rather than physics can be traced back far in the history of economics (see, e.g., Niman (1991)). In this thesis I assume that there exists some fundamental force that changes core aspects of an economic system over time. There can be a variety of reasons for this evolution to take place: Adaption of market participants (e.g. through a revision of expectations (see, e.g., Lucas (1976)) or simply through individual adaptive behavior (see, e.g., Hommes (2005))), changes in institutions and regulation (see, e.g., North (2005)), or technical innovation (see, e.g., Schumpeter (1949), Romer (1990)). It has to be noted that in the following I only assume the existence of such a force and provide no general model for it. Instead, the model has to be developed on a case-by-case basis. However, an application can be found in chapter 2.

1.1.4 The need for objectification

In the presence of ongoing structural change it is impossible to determine the current state of any system out of discrete observations. However, one should not confuse a limited understanding with no understanding at all. In perception, all of our senses are limited and we still use them to go about our daily lives. In the context of Economics those senses have not had time to adapt to the environment of a global financial system (see Lo (2017)). In a situation, such as economic behavior, in which our subjective judgments based on personal perception and observation are comparatively more inadequate than in our daily lives the need to rely as much as possible on objective information is understandable. Even the most ardent Post-Keynesians do not argue against models and objectification *per se* (see Davidson (2012)). The interesting question in this contest is *'What are the limits to our understanding of the world around us?'* (see North (1999)). The objective of this thesis is to probe the statistics angle of this question, i.e. the possible degree of quantification which is needed in the face of flawed human judgment. As such, the arguments in this thesis also deal with the distinction between rationality and ecological rationality (see, e.g., Gigerenzer and Todd (1999), Gigerenzer and Selten (2002), and Gigerenzer (2015)). This strand of literature introduces a spectrum of rationality with a fluid environment, bounded rationality, and decisions via heuristics on one end and a static environment, rationality and decisions via constrained optimization on the other end, i.e. the decision paradigm is depending on structural properties of the environment. In this context it would be very nice to know where to draw the line, i.e. in which situations one can reasonably argue for one decision paradigm over another based on 'stability' properties of the environment. The arguments in this thesis deal with the question how near we can get to a 'rational' environment in which all decision-relevant parameters can be inferred.

1.1.5 Ergodicity and Stationarity

So far I dicussed the notion of fundamental change in terms of the ergodicity assumption since this is typically the concept used in the literature on economic history. However, since I am interested in quantitative measurement I need a concept that allows me to manipulate the stochastic properties of a system over time, i.e. a concept that allows for structured violations of the ergodicity assumption while also remaining recognizable to the methodology typically used in stochastics. Non-stationarity serves this purpose: It is a stronger concept than ergodicity (i.e. a process can be stationary but not ergodic, see Kirstein (2015) for an example) and it is typically modeled as a parametric property, i.e. there is a parameter that can be manipulated in order to arrive at a degree of nonstationary behavior. In this context it is useful to rely on a class of stochastic processes whose stationarity properties can be easily manipulated by a small set of parameters. Semi-martingales, in general, and time-inhomogeneous diffusion processes, in particular, serve that purpose: Most of the stochastic properties of semi-martingales are governed by their 'characteristics' which are appropriate generalsations of the well-known Lévy-Khintchin triplet (drift coefficient, diffusion coefficient, jump measure) of Lévy Processes (see Jacod and Shiryaev (2003)). Restricting myself to the class of semi-martingales has the additional advantage that a stalemate of the mathematical finance techniques, stochastic calculus, is still accessible in this setup since semi-martingales are the largest class of stochastic integrands for which one can construct stochastic integrals in a convenient way (see Protter (2005)). Within the class of semi-martingales I restrict myself further: Dealing with more than one aspects of the characteristics would entail dealing with a very complex filtering problem which would not be directly related to the data identification argument introduced here. Consequently, I restrict myself to driftless, time-inhomogeneous semi-martingales with deterministic characteristics without jumps or simpler, independent increment time-inhomogeneous diffusion processes (see Jacod and Shiryaev (2003) and Protter (2005) for a background on the related concepts).

1.2 Objective of the Thesis

The objective of this thesis is to provide quantitative statistical and modelling methods for financial and economic questions in a framework in which the stationarity and ergodicity assumption is violated. As pointed out above, a non-ergodic environment implies that one cannot rely on market efficiency. After introducing a method to optimally extract statistical knowledge from a changing environment by means of an optimal data set choice and an argument that is used for the generalisation of this concept, I use market inefficiency as a modelling tool in the context of electricity markets.

1.3 Outline of the Thesis

1.3.1 Representativeness vs Convergence: Optimal Data Selection in Non-Stationary Systems

In statistics the advantage of more data is quantitatively measureable in a better estimator convergence. However, under the assumption of a non-stationary environment data from the past is only partially representative of the current situation. By deriving convergence rates for independent but not identically distributed (i.-non-i.d.) normal random variables, this paper aggregates those two sources of errors and arrives at a functional form of the aggregated error. The combined error function can be minimised which leads to the identification of optimal data sets. I then discuss the existence and uniqueness of optimal data sets and support my derivations with appropriate simulations. In addition, I analyse the impact of a rule found in the foundational document for banking regulation (Basel 3 framework) which, apparently, is based on a non-ergodic worldview and illustrate the severity of its consequences.

1.3.2 A Convergence Speed Dependent Data Quantity Definition for Weighted Observations

This part serves as a preparation for the generalisations of the methodology introduced in section 2. Section 2 indentifies an optimal data window, i.e. an optimal data quantity, in order arrive at minimum variance estimators. Since there are different methods to downweight data from the past, it is interesting to ask whether similar optimality criteria can be constructed for those other weighting methods. In order to pursue this question one first needs to find convergence errors for those alternative weighting methods for which it is necessary to define an implicit data quantity of a specific weighting method. This paper introduces this connection. As such, this paper is a substantial correction of an argument that can already be found in the literature (see Meucci (2012)). I document that this paper contains an unjustified step leading to an erroneous conclusion implying confidence bands that are *unjustifiably narrow*. In addition, I provide a formal argument to identify the correct candidate within the set of possible data quantity functions constructed in Meucci (2012). In an appendix to the paper, I perform a simulation similar to the one in section 2 in order to show that the error aggregation and possible data identification mechanism introduced there can be extended beyond the case of classical data windows. Analytic results are currently out of reach.

1.3.3 Introducing Stylized Facts of Electricity Futures Through a Market Impact Model

This part of the thesis provides an alternative way to introduce some of the distinctive qualitative features of electricity futures by embracing market specific inefficiencies and problems. Specifically, it provides an alternative rationale for the 'Samuelson Effect' (see Samuelson (1965)) postulating that the volatility of futures is increasing throughout their life span due to an increase of relevant 'short-term' information (see Benth *et al.* (2008a)). In this part I induce an increase in futures volatility and other characteristic behavioral aspects of electricity futures out of an objective information process with constant volatility which is increasingly amplified by the impact of the behavior of market participants trading towards their target. Since electricity is currently not storable on an industrial scale supply and demand have to be cleared at every point in time. For this reason quantity risk and the corresponding notions of market impact and optimal liquidation behavior play a major role in electricity markets. This paper provides a stochastic model introducing the stylized facts of electricity futures out of those considerations. Conceptually, this paper provides one example how to induce non-stationary futures behavior out of a stationary information process by utilizing market specific inefficiencies.

1.4 Further Research

1.4.1 Uncertainty and Risk

Economics and Finance are based on individual decisions. Decisions are influenced by our observations and understanding of the world around us. In Finance and Economics one (typically) distinguishes between the notions of risk and uncertainty. Risk refers to a situation where the probability distribution of the problem in question is known. Uncertainty, on the other hand, refers to a situation where the distribution (Knight (1921)) or even the possible states (see Keynes (1921) and North (1999)) are not known. When one confronts economic subjects with decisions in those two frameworks, one can find significant differences in behavior (see Ellsberg (1961)), i.e. the distinction is decision-relevant. Uncertainty and non-ergodicity are closely related and one of the outcomes of the methodology presented in section 2 is that one always operates under limited statistical perception but one does so under varying degrees. Consequently, one can base decisions (like, e.g., portfolio management decisions) based on the degree of uncertainty given in the particular circumstances: The more data one has, the more one can rely on techniques that require estimates in order to optimally act in a given context (e.g. Markowitz portfolio optimisation). However, the less data one has, the less knowledge one has about the distribution of the problem and, consequently, the more one has to base the decision on some paradigma regading how to act under uncertainty, i.e. the decision is viewed out of a perspective of a degree of ecological rationality (see Gigerenzer (2015)). This question goes beyond the scope of this thesis and is currently under preparation.

1.4.2 Optimal Sampling Frequencies for Low-Frequency Signal Inference

In Finance and Economics one frequently encounters arguments that the qualitative properties of time series are different on different time scales. In Quantitative Finance an entire subfield, high frequency finance, exists precisely for this reason. If one only has access to a limited time series, say, 3 years of data, and one is interested in the extraction of a long-term signal, i.e. a signal on low frequencies, say, a yearly volatility, it could be beneficial to set up a bias variance argument similar to the one in section 2 where one does inference on higher frequencies (e.g. monthly) in order to have access to more data. However, in this context one has to take care of the qualitative difference (bias) that can be found on other time scales. In order to set up this problem properly, one needs to move away from self-similar processes, i.e. one needs to employ, e.g., fractional processes. However, with fractional processes one loses the possibility to employ the usual mathematical finance toolbox and, thus, it is not clear to which extent the results of this approach can be integrated into a practical, say, portfolio management problem.

1.4.3 Decidability & Replication Crisis

One basic objective of this thesis is to find optimal data sets. A natural follow up question is for which purpose those data sets can be used, i.e. to decide whether the data set can be used for the justification of some statement. In a non-ergodic world, the truth-value of statements can change over time. Consequently, reality is shifting over time and one can only be interested in the truth-value of a statement at some point in time. Given a limited amount of data one has to find a method to translate the number of data points and the signal within them in a decision topology of statements. This is a challenging task. However, if one acknowledges that the truth-value of statements can shift over time one also has a new way to look at the so-called 'replication crisis'. Currently, a variety of fields grapple with the finding that the results published even in the best journals are not replicable and sometimes not even reproducible (see, e.g., Open Science Coll. (2015), Christensen and Miguel (2018)). Currently, the blame for this perceived failure to replicate is targeted primarily on statistical techniques and the misincentives imbedded into them (see Ioannidis (2005), Bailey *et al.* (2014)). While I acknowledge that current statistical techniques carry misincentives, it is still conceivable that additional forces are at play, namely that the underlying signal has changed over time. Given the abundance of studies of relationships between economic measures and measures of social well-being, e.g. suicides, it is not unfathomable that certain psychological studies are influenced by the Zeitgeist and that the same study on the same demographic simply cannot replicate results due to changes in culture and other society-wide issues. This view has been formulated in psychology (see Greenfield (2017)). However, this argument is interesting in the context of this thesis since the failure to replicate can serve as a measure of representativeness. When studies fail to replicate results, this can be an indicator that aspects of the society are transforming.

2 Representativeness vs Convergence: Optimal Data Selection in Non-Stationary Systems

Abstract

Starting from the premise that economic and social systems change fundamentally over time and that, consequently, data from the past is only partially representative of the current situation this paper aims to identify data sets yielding minimal bias estimators. As an application an impact study of a paragraph in the banking regulation framework is performed as an illustration of the proposed methods to identify such data sets. [1]

2.1 Introduction

It is widely recognized that parameter instability is a crucial issue when analyzing financial and economic time series (see, e.g., Stock and Watson (1996), Pesaran and Timmermann (2007), Giacomini and Rossi (2009), Giraitis *et al.* (2013), Inoue *et al.* (2017)). In order to handle such instability, instead of using all observations, it is quite common to reduce the number of data points in an estimation. In Finance, e.g., it is very common to rely on rolling windows and other data-weighting techniques that attempt to strike a balance between using as much objective information as possible while, simultaneously, not relying too much on outdated information. In this context, however, it is of crucial importance to clarify to which degree one should limit the possible input to some statistical techniques. So far, the arguments that are used to rationalize the size of rolling windows or, more generally, 'data quantity' are either the experience of the researcher, or arguments based on statistical significance (see, e.g., Christoffersen (2011), Alexander (2009)) as well as a 'standard in the literature' which can be traced back to either of the first two options. Given the subjective nature of the first method and the incompatibility of some asymptotic argument with any genuine 'parameter instability', which is considered to be the driving force of our economic activities (see Schumpeter (1949)), a quantitative method is needed

[1] An earlier version of this paper has received a YITP price at the XIX Quantitative Finance Workshop 2018. I want to thank the committee and the sponsor (ACRI) for their interest and support and the staff at Scuola Normale Superiore for their hospitality during the corresponding research visit.

in order to justify the size of the data set used in an analysis. The introduction of such a quantitative criterion is the primary objective of this paper.

My data identification method is based on a perception of statistical instability which will be formalized in the concept of a 'representativeness metric'. As a special case, this perception can be that the economic environment is perfectly stable, i.e. that the past is fully representative of the present (see Samuelson (1968) for a defence of the related ergodic hypothesis) and in this situation it is natural to use all the possible data that is available since any additional observation reduces estimator convergence errors. Given that it is commonplace that rolling window estimation is utilised in Empirical Finance and Economics studies, there is a fundamental disconnection between empirical practice, where data sets are restricted on a regular basis, and economic theory assuming that the past is fully representative of the current situation.[2]

The problem of an optimal data reduction in the presence of parameter instability consists of two parts: A detection of a change in the stochastic underpinnings of a time series and, subsequently, a rationalisation of the corresponding data reduction. In the context of this paper I circumvent the first problem by introducing a generic formal tool, the representativeness metric, to keep track of changes in parameter values over time. However, my approach allows for more: Instead of tying the representativeness metric to actual movements in parameters (which will be the definition for the formal derivation) one can also use this tool to keep track of changes in the institutional environment that have yet to materialise in actual movements of the parameters like, e.g., credit-buildup in a volatility estimation (see Danielsson *et al.* (2018)) and more generally, indicators of economic change. Thus, if one is motivated by movement of parameters, natural choices for the representativeness metric are given by (but not limited to) jump detection techniques (see, e.g., see Pesaran and Timmermann (2007), Giacomini and Rossi (2009), Giraitis *et al.* (2013), Inoue *et al.* (2017), Aït-Sahalia and Jacod (2014)) and techniques to infer ongoing

[2]Note, that in Physics the question whether the laws of nature are stable across space and time has been debated since the 1930s (see Dirac (1938), Uzan (2003), Webb *et al.* (2011), Webb *et al.* (2001)). Given that physicists cared about the stability of their fundamental assumptions for a period of almost a century, economists may benefit from attention to this question by sheer transitivity and this paper provides a tool to perform inference in such an unstable system.

structural change (see, e.g., Phillips *et al.* (2017) and Giraitis *et al.* (2013)) while other choices include indicators of economic change like liquidity metrics or measures of systemic risks (see Danielsson *et al.* (2018)). Consequently, the objective of this paper is the second question introduced above, i.e. which degree of data reduction is justified in a situation where one already has arrived at an understanding of whether and, if so, to which degree the underlying parameter or the system governing the parameter has changed. In order to set up an optimization one has to define an appropriate target that can be optimized: The choice of my target function is the first main difference to the current econometric literature on the problem: Instead of minimizing forecasting errors, i.e. considering the problem

$$\min_{data} E_T\left[(y_{T+1} - \tilde{y}_{T+1,data})^2\right] \ , \tag{2}$$

where $\tilde{y}_{T+1,data}$ denotes a forecast for some time series $(y_t)_t$ at time $T+1$ using the information in *data* and where $E_T[\cdot]$ denotes the conditional expectation operator given the information at time T, I am interested in optimal inference, i.e.

$$\min_{data} E_T\left[(\sigma_T - \tilde{\sigma}_{T,data})^2\right] \tag{3}$$

for some statistical quantity σ_T and its estimator $\tilde{\sigma}_T$. The Econometric Literature mentioned above minimizes forecasting errors while I am interested in minimizing estimation errors. For the first problem one needs to introduce a model for the forecast $\tilde{y}$ and, subsequently, performs a goodness of fit calibration which depends on the model parameters of the forecasting model. The second problem addresses the minimization of the distance between two statistical properties σ_T and $\tilde{\sigma}_{T,data}$. The difference between the two problems is profound: While the second optimization attempts to approximate the current statistical state to an optimal degree, the first addresses model dependent optimal forecasts of a time series.

To the best knowledge of the author, the literature on econometrics has focussed exclusively on the first problem. However, it is noted (see Giraitis *et al.* (2014)) that forecasting in the second framework is possible by setting the forecasts to the end-of-observed-sample values. A variation of the second problem has been discussed in the literature on acoustics and signal processing under the name of 'optimal segment length' (see Dahlhaus and Giraitis (1998))). However, to the knowledge of the author this literature focusses on the question of optimal data intervals *around* a data point (see Theorem 2.3 in this paper) and not at the end of the data set which is the question that is typically the one in which one is interested in financial and economic settings. In addition, the techniques used in this literature (non-stationary processes with an evolutionary spectral representation) are more technical and highly different from the methodology used here (convergence speeds of moments of mixture distributions). See also the literature review below. Furthermore, in Giraitis *et al.* (2014) it is argued that the class of models from this strand of literature, i.e. the methodology in Acoustics and Signal Processing, *' [have] not really been influential in applied macroeconometric analysis [...]'* which warrants the conclusion that it is not a common technique in the Economic and Finance Literature at large. The second main difference to the Econometrics Literature is that I am situating myself in a setup that is suitable for quantitative finance purposes. Below I will utilize a semimartingale setup which is the largest class of processes for which one can conveniently define stochastic integrals (see Protter (2005)) which are a powerful tool in Quantitative and Mathematical Finance. To sum up, I consider the so-called 'bias-variance' tradeoff within the realm of 'estimation techniques', i.e. optimal inference, instead of 'learning techniques', i.e. optimal predictions, (see Geman *et al.* (1992), Breiman (2001)) within a framework that is suitable to answer quantitative finance questions and that is typically formally excluded in econometric analysis (see below).

For the sake of completeness I want to give a short overview of a prominent strand of literature dealing with inference of parameters of either non-stationary stochastic processes or time series exhibiting structural change. Naturally, there is a deep relation between those two questions. In the following I will primarily rely on the article Dahlhaus (2012) reviewing the literature on locally stationary processes (which have been utilised extensively in financial econometrics, see, e.g., Stărică and Granger (2005) since their introduction in Dahlhaus (1997)), i.e. one popular class of 'non-stationary processes'. However, note that time-inhomogeneous brownian motions constitute continuous-time 'random walks', i.e. unit root processes. Unit root processes are notoriously difficult to deal with in a classic time series context and are typically excluded from the analysis. This is also the case in the context of this review article (see Dahlhaus (2012), condition on α [Notation from the paper] after equation (1), pg.3). Thus, 'non-stationarity' in the literature on inference of non-stationary processes and structural change refers to a time-dynamics of the autoregression parameter while 'non-stationarity' in the context of this paper refers to non-stationarity of the volatility of the (continuous time) error terms. See, especially eq. (76) in Dahlhaus (2012) which formulates explicitly a condition on stationarity necessary to achieve the main representation used in this literature.

Nonetheless, for the sake of completeness let us consider the setup in this review article: Dahlhaus (2012) primarily discusses processes with with a time varying spectral representation, i.e. processes of the form

$$X_t = \int_{-\pi}^{\pi} exp(i\lambda t)\bar{A}_t(\lambda)d\xi(\lambda), t \in Z \tag{4}$$

with an orthogonal increment process $\xi(t)$ and a time varying transfer function $\bar{A}_t(\lambda)$ (see Dahlhaus (2012) for details and Kayhan *et al.* (1994) eq.(1) - (8) for a concise derivation of the formula above in a special case). Originally, this setup of analysis has been introduced in Priestley (1965). The main technical tool used in the literature summarised by the review article are 'infill asymptotics' where the transfer function $\bar{A}_t(\lambda)$ is replaced with some function $A(\frac{t}{T}, \lambda)$ (see Dahlhaus (2012), especially eq. (78), for details). Thus, time structure between the X_t is introduced by means of the transfer function $A(\cdot, \cdot)$. However, *' [...] processes which can be described with this infill asymptotics are processes which locally at each time point are close to a stationary process but whose characteristics (covariances, parameters, etc.) are gradually changing in an unspecified way as time evolves.'* (see Dahlhaus (2012), pg. 2), i.e. the framework using infill asymptotics uses local structural assumptions, mainly differentiability of the parameter function of the autoregression parameter over time, which is not needed for the data identification method developed in this paper. As such the modeling framework considered in this paper is substantially different from the one in Dahlhaus (2012) and, for this reason, this framework is not considered further. However, Dahlhaus (2012) references additional work on processes with time-varying parameters that does not use infill asymptotics (see pg. 2). All the papers mentioned in this context, i.e. Subba Rao (1970), Hallin (1986), Grenier (1983), Kayhan *et al.* (1994), consider only stochastic processes that have connections to time series modeling. Specifically, all those papers either exclude the unit root case explicitly or implicitly (by using some form of the 'Wold Representation' (see, e.g.,Greene (2012)) which requires a covariance stationary time series, i.e. a constant autocovariance structure which is, in general, not fulfilled by time inhomogeneous brownian motions [3] and are, as a consequence, not applicable within the framework considered in this paper). Additional reviews on forecasting in non-stationary environments (see Giacomini and Rossi (2015)) and general structural breaks in time series (see Casini and Perron (2019)) are available. However, to the knowledge of the author, there are no results on the identification of optimal data windows for time-inhomoegeneous brownian motions, their generalizations, or their discrete analogue, i.e. 'unit root processes'.

While the primary objective of this literature review is to demonstrate a degree of novelty of the methodology developed in this paper it has to be noted that the literature on financial econometrics struggles with non-stationarity in a broader sense. One prominent paper, Mikosch and Stărică (2005), on non-stationarity in return time series argues that non-stationarity can be used in order to explain certain time series puzzles (specifically, the 'IGARCH effect'). However, the authors come to the following conclusion (Mikosch and Stărică (2005), pg. 388f): *'As for the question whether there is [a statistical property] in the absolute log-returns or not, we believe that, because one cannot decide about the stationarity of a stochastic process on the basis of a finite sample, that question will certainly keep the academic community busy in the future.'* As such the analysis boils down to the question whether one believes that there is non-stationarity in the data. The representativeness metric is designed to circumvent this problem altogether. It serves as a formal tool which allows to incorporate any subjective opinion regarding the degree of non-stationarity (including 'no non-stationarity') of the data generating process.

In order to construct data sets based on an opinion regarding the dynamics of the underlying process I proceed as follows: In section 2.2 I introduce a class of non-stationary

[3] By the Levy Characterisation of the brownian motion (see, e.g., Karatzas and Shreve (1998)) it is known that $cov(B_s, B_t) = min(s,t)$ for a standard brownian motion. However, for a time-inhomogeneous brownian motion $\tilde{B}$ it follows that $cov(\tilde{B}_s, \tilde{B}_t) = E[\tilde{B}_s \tilde{B}_t]) = E[\tilde{B}_s(\tilde{B}_s + (\tilde{B}_t - \tilde{B}_s))] = E[\tilde{B}_s^2] = Var(B_s^2) = \sigma_s^2 \cdot s$ where σ_s^2 measures the average variance of the process from its starting point up to s. For a non-stationary process this means that the covariance structure changes over time. Thus, such a process is not necessarily covariance stationary.

stochastic processes whose stationarity properties can be easily manipulated. In addition, a metric measuring the degree of nonstationarity is introduced. In section 2.3 I introduce the representativeness and estimator convergence errors that are present in the nonstationary setup from section 2.2. In section 2.4 I combine the errors and discuss the existence and uniqueness of a minimal bias data set. Section 2.5 is devoted to applications. First, I apply the framework to a setup with different jump sizes in order to illustrate the mechanic developed beforehand and afterwards I perform an impact analysis of a paragraph in the banking regulation framework. In section 2.6 I draw conclusions.

2.2 Setup

The objective of this section is to fix a class of processes whose stationarity properties can be conveniently manipulated. To this end, I will rely on a subclass of semimartingales which are a prominent tool in quantitative finance since they are the largest class of processes for which one can conveniently define stochastic integrals (see Protter (2005)). The stochastic properties of general semimartingales can be expressed through their 'characteristics' which will be introduced below (see Jacod and Shiryaev (2003)). The notion of 'characteristics' of a semimartingale is designed to extend the three terms drift, variance and Lévy measure that characterise the distributional properties of Lévy Processes. Let $(\Omega, F, P, (F_t)_t)$ be a stochastic basis fulfilling the usual hypothesis (see Protter (2005)).

Definition A stochastic process $X = \{X_t : t \geq 0\}$ is said to be a Lévy process if it satisfies the following properties:

1. $X_0 = 0$ almost surely.
2. For any $0 \leq t_1 < t_2 < ... < t_n < \infty$, the corresponding increments $X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, ..., X_{t_n} - X_{t_{n-1}}$ are independent.
3. For any $s < t$, $X_t - X_s$ is equal in distribution to X_{t-s}.
4. For any $\varepsilon > 0$ and $t \geq 0$ it holds that

$$\lim_{h \to 0} P\left(\mid X_{t+h} - X_t \mid > \varepsilon\right) = 0\,.$$

The distribution properties of a Lévy process are determined by its characteristic function, which is given by the well-known Lévy-Khintchin formula (see, Protter (2005), Ch.1 Thm. 43):

Theorem If $X = \{X_t : t \geq 0\}$ is a Lévy process, then its characteristic function $\psi_X(\theta)$ is given by

$$\psi_X(\theta)(t) := E\left[e^{i\theta X_t}\right] \tag{5}$$

$$= \exp\left\{t\left(ai\theta - \frac{\sigma^2\theta^2}{2} + \int_{\{|x|\geq 1\}}(1-e^{i\theta x})\nu(dx) + \int_{\{|x|<1\}}(1-e^{i\theta x}+i\theta x)\nu(dx)\right)\right\}, \tag{6}$$

where $\nu(dx)$ is a Lévy measure, i.e. ν fulfills $\nu(0) = 0$ and $\int_R \min(1, x^2)\nu(dx) < \infty$, where R denotes the real numbers.

Since the characteristic function uniquely determines the probability distribution of the corresponding random process each Lévy process is uniquely determined by its *Lévy-Khintchine-* or *characteristic triplet* (a, σ, ν). Given the characteristic function it is clear (see, Jacod and Shiryaev (2003), pg. 75) that the process

$$M_t := \frac{\exp(i\theta X_t)}{\exp(\psi_X(\theta)(t))} \tag{7}$$

is a martingale. For more general processes eq. (7) is the condition that is used in order to *define* the characteristics. In the following I will consider a subclass of semimartingales. Semimartingales are defined as follows:

Definition (see, Jacod and Shiryaev (2003), pg. 43, Def I.4.21.): A semimartingale is a process X of the form

$$X = X_0 + M + A\,, \tag{8}$$

where X_0 is finite-valued and F_0-measurable, where M is a local martingale, i.e. a process that satisfies the martingale property for some sequence of stopping times, such that $M_0 = 0$ and where A is an adapted process with finite variation such that $A_0 = 0$.

Concerning its characteristics, for a semimartingale X the objective is to find two processes $a(t)$ and $\sigma(t)$ and a random measure ν such that

$$\psi_X(\theta)(t) = \exp\left\{ i\theta a(t) - \frac{\theta^2\sigma^2(t)}{2} + \int_{0^c} e^{i\theta x} - 1 - i\theta 1_{|x|<1}\nu([0,t]\times dx)\right\} \tag{9}$$

is the characteristic function of X and

$$M_t := \frac{\exp(i\theta X_t)}{\exp(\psi_X(\theta)(t))} \tag{10}$$

is a martingale. Here, 0^c denotes the complement of 0. In the context of non stationary processes the jump measure nu can be different at different points in time, thus one has to allow for an additional degree of freedom related to time which is typically introduced via the notation used above (see Karatzas and Shreve (1998)). For stationary processes eq. (9) collapses to eq. (6) with $\sigma^2(t) = \sigma^2 \cdot t$ and $a(t) = a \cdot t$ and an according transformation of the jump measure. It can be shown that for every semimartingale there exists a unique set of predictable processes $a(t), \sigma(t)$ and a random measure ν that fulfills those conditions. This set of objects is called the characteristics of the semimartingale X (see Jacod and Shiryaev (2003), Def II.2.6 and Proposition II.2.9).

If one considers semimartingales with independent increments, one can show that the corresponding set of characteristics has to be indistinguishable (see Karatzas and Shreve (1998), Def. 1.3) from a deterministic process (see Jacod and Shiryaev (2003), Theorem II.4.15)[4], i.e. the characteristics of such a process have to be deterministic functions of time. This is a convenient setup to work in since now the distributional properties of the process are parametric and can be inferred using classical statistical techniques. In this class of processes I restrict myself further: Since in data sets we observe only the combined behavior of all the characteristics one would have to first separate this combined signal into its parts, i.e. one would have to decide which attribute of an increment was caused by the drift or the diffusion part or the jump measure. Testing for jumps in a discretely observed process is a challenging problem (see, e.g., Aït-Sahalia and Jacod (2009) and Aït-Sahalia and Jacod (2014)) and since the objective of this paper is to provide a rationale for optimal data identification rather than a contribution to this strand of literature, I

[4]In Jacod and Shiryaev (2003) the authors use the terminology 'version', which is not defined in the book, for the same property that is called 'indistinguishability' in Karatzas and Shreve (1998). See Jeanblanc, Yor and Chesney (2009) for a formal identification of the two properties.

will circumvent this problem by assuming that the available observations are the result of a single characteristic. To this end, I assume that there is no drift and that there are no jumps and that the only characteristic that is present is given by the variance of the diffusion part. This restriction leaves me with driftless, continuous semimartingales with independent increments.

Since the objective of this paper is to construct optimal data sets out of a conception of stochastic instability, I also need a device measuring this instability. From a mathematical perspective, it is convenient to tie the instability measure to the dynamics of the parameter governing the stochastic behavior:

Definition Let X be a driftless continuous semimartinagle with independent increments and characteristics $\sigma(t)$ defined on $(-\infty, T]$. Let $t_1 < t_2 \leq T$. The quantity

$$R(t_1, t_2) = \int_{t_1}^{t_2} |\frac{\partial \sigma(s)}{\partial s}| d\xi \tag{11}$$

is called the *representativeness metric* of the process X on the interval $[t_1, t_2]$. For notational purposes one has to assume that σ is differentiable. However, in discrete samples it is not decidable whether the dynamics of a continuous parameter is differentiable or not. As such, it is natural to consider the finite approximation of the integral above whenever necessary, i.e. finite differences as approximations of the derivative and a discrete-valued measure ξ assigning a weight of 1 for elements in the finite approximation and zero otherwise. In this regard, the differentiability assumption is not critical for the data identification method developed in this paper.

This is a formulation of the representativeness metric that does not take into account changes in the institutional environment (as mentioned in the introduction) but only the time dynamics of the actual parameters involved. For formal derivations this is a useful definition. However, in applications one has to have a *model* of the representativeness metric (see section 2.5). Formally, the representativeness metric serves as a distance measure on the space of stochastic processes from 'stationary behavior' when considering the process at different points in time. For a stationary process in our model class, i.e. a brownian motion with characteristics σ, the metric is zero since σ is constant over time and neither a derivate nor a finite sum of differences will be nonzero. On the other hand, a process with a jump of size κ at time s in its characteristics exhibits representativeness $R(s-\varepsilon, s_1) = \kappa, \forall \varepsilon > 0, s_1 \geq s$. Note that the second process can also have a representativeness metric of zero if s is not in the time interval under consideration. Thus, the representativeness metric measures deviations from stationary behavior within the time frame defined by its arguments.

Given that the objective of the paper is optimal inference based on a perception of instability, it is clear that the degree of data reduction should be higher for data sets generated by processes that exhibit a higher degree of non-stationarity, i.e. higher levels of $R_{(\cdot,\cdot)}$ over some time frame (for an example, see section 2.5.1). Furthermore, the representativeness metric has the purpose of measuring what is genuinely 'new' about a situation. In a mathematical context 'new' means that the values of the parameters have to be at levels that have not been attained in the history of the sample. Consequently, for the formal derivation it is necessary for identification purposes to assume a monotone dynamics of $\sigma(t)$ in order to ensure that 'novelty' as measured by the representativeness metric corresponds to 'novelty' in terms of stochastic behavior. Hence, for the next two sections I assume a monotone behavior of $\sigma(t)$ and the representativeness metric will not appear again until the application in section 2.5. This type of monotonicity assumption, which can be veiled by the assumption that there is just one isolated structural break, is common in order to have proper identification mechanism (see Pesaran and Timmermann (2007), sect 3.6) and constitutes a generalization of the setup that is used in this paper since it allows for both continuous change and jumps (see section 2.5).

However, when considering applications then the situation is typically different: When looking at specific risk metrics, the same market at different times can exhibit the 'same risk', i.e. the monotonicity assumption used in the formal derivation is violated. If one would assume that past periods of, e.g., low volatility are comparable to current periods of low volatility then one could argue in favor of, e.g., state-space conditioning (see, e.g., Meucci (2012)) in order to shore up the size of data sets. However, given that the underlying structure of the market is changing over time, it is important to distinguish between periods where the measurable risk criteria are similar, i.e. in applications one encounters situations where the mathematical notion of 'novelty' does not correspond to the economic notion of 'novelty'. Consequently, it is useful to define the representativeness metric on the basis of indicators of economic change and adaptation of expectations. Prime examples for this are liquidity metrics or extreme values which indicate an update of expectations. A simple version of the latter will be utilised in section 2.5.

2.3 Individual Errors

In terms of the setup from the last section, the objective of this paper is to estimate $\sigma(T)$ by choosing an optimal data set of discrete observations $\{x_i\}_i \subset [0,T]$ of the process X utilising knowledge on $R(\cdot,\cdot)$. For the purpose of deriving the optimality condition I will operate under perfect knowledge, i.e. for the purpose of the following two sections I assume full knowledge of the history of the process up to time T, i.e. $\{\sigma(t)\}_{t\leq T}$. The optimization is set up between data quantity which carries the incentive to use as much data as possible and data quality which does the opposite. The objective of this section is to derive estimator convergence and representativeness errors as a (data) interval dependent properties.

2.3.1 Representativeness Error

When one uses non-representative data to estimate some quantity, one makes an error. The objective of this subsection is to derive the extent of the error for the variance estimator, i.e. for the characteristic of the process that is still present in our setup. The situation one encounters is the following: Given one is situated at the end of a data set that spans the interval $[0,T]$ one is interested in deriving the difference between the current state of the characteristics, i.e. $\sigma(T)$, and the value of the estimator one would have of this quantity using some data set $[t_1,t_2] \subset [0,T]$ which I will denote by $\hat{\sigma}_{[t_1,t_2]}$. For convenience I will assume $t_2 = T$. I am now interested in the moments of $\hat{\sigma}_{[t_1,T]}$ as a function of t_1.

Since I operate under the assumption of knowing $\{\sigma(t)\}_{t\in[0,T]}$, I can immediately proceed to calculate $\hat{\sigma}_{[t_1,T]}$ by appealing to results on mixture distributions since $\hat{\sigma}_{[t_1,T]}$ is the standard deviation of the random variable created by mixing the normal distributions with local standard deviation $\sigma(t)$ over the time interval $[t_1,T]$. For this purpose t is considered to be equally distributed on the interval $[t_1,T]$. The result can be summarised as follows.

Proposition For a driftless semimartingale with independent increments and without jumps

$$X(T) = \int_0^T \sigma(s) dW(s) \tag{12}$$

the variance of the random variable $X_{[t_1,T]} := X(T) - X(t_1)$, where $t_1 > 0$, is given by the mixture distribution of the Wiener process W with diffusion term $\sigma(t)$ and t. More specifically, X is locally distributed according to a centralised normal distribution with parameter $\sigma(t)$ and t is uniformly distributed in the time interval $[t_1, T]$. Consequently, the distribution of X restricted to the interval $[t_1, T]$ can be calculated as

$$f_{X_{[t_1,T]}}(x) = \int f(x|\sigma^2(t)) f_{\sigma^2}(t) dt \tag{13}$$

$$= \frac{1}{T - t_1} \int_{t_1}^{T} \frac{1}{\sqrt{2\pi\sigma^2(t)}} \exp\left\{ -\frac{x^2}{2\sigma^2(t)} \right\} \sigma^2(t) dt \ , \tag{14}$$

where f is the density of a normal distribution and f_σ^2 denotes the distribution of σ^2 with respect to its latent parameter t. Specifically, the variance is given by

$$\hat{\sigma}^2_{[t_1,T]} = V_{X_{[t_1,T]}} = E_t\left[V(x|\sigma^2(t))\right] + V_t\left[E(x|\sigma^2(t)\right] \tag{15}$$

$$= \frac{1}{T - t_1} \int_{t_1}^{T} \sigma_B(t)^2 dt + 0 \ , \tag{16}$$

where $E_t[\cdot]$ and $V_t[\cdot]$ denote expectations and variances with respect to the distribution of t.

Proof The results above are basic properties of mixture distributions, see, e.g., Lindsay (1995).

Given this result it is possible to define the difference between the statistically perceived variance, i.e. the estimator if estimated using the mean of a uniform distribution, within an interval and the current state of the process.

Definition

Let $t_1 < T$, $t_1 \in [0, T]$. The **representativeness error** β_σ of the parameter σ that is caused by using data from the interval $[t_1, T]$ is given by

$$\beta_\sigma(t_1, T) = \left| \frac{1}{T - t_1} \int_{t_1}^{T} \sigma^2(s) \ ds - \sigma^2(T) \right| \ . \tag{17}$$

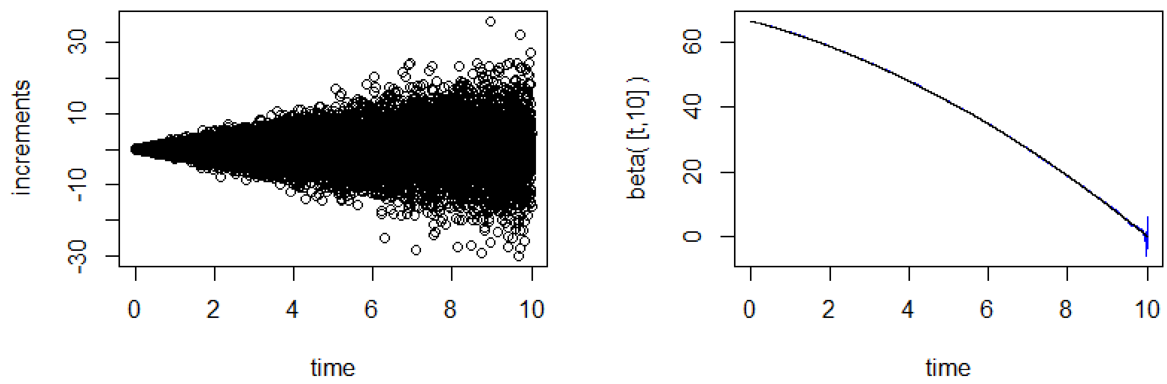

Figure 1: **Example - Nonrepresentativeness Error**: In the left panel an example data set for the situation from the example is displayed. In the right panel the corresponding average simulated (blue) and theoretical (black) representativeness error $\beta_B(x, 10)$ is shown. Here, the simulated errors are the average results of a monte carlo simulation of 100.000 data sets comparable to the one on the left. It is visible that for small data quantities (in the vicinity of 10) estimator convergence is still an issue and leads to more scattered errors.

Example

Let $[0, T] = [0, 10]$ and $\sigma(t) = t$. Then the representativeness error is given by

$$\beta_\sigma(t, 10) = \left| \frac{1}{10 - t} \int_t^{10} s^2 ds - 10^2 \right| \tag{18}$$

which is easily solvable and can be contrasted against a simulation (see Figure 1). It can easily be seen that the representativeness error is low when only using data that is in the immediate vicinity of 10 and becomes larger the more data is used.

In the example as well as in the definition it is easily discerned that the representativeness error is increasing in the amount of data that is used to estimate $\sigma(T)$ and consequently constitutes an incentive to use as few data as possible when estimating parameters in a non-stationary environment. However, when performing any statistical analysis some data is necessary to achieve any form of estimator convergence. To calculate the quantitative benefits of additional data in the context of mixture distributions is the objective of the next subsection.

2.3.2 Convergence Error

The value of more data is given by an increased estimator convergence. Consequently, it has to be assessed how beneficial additional data is in the context of mixed normal distributions. Since data is discrete and the ultimate answer to the question how much data should be used in an estimation will also be discrete any argument related to estimator convergence will naturally only depend on natural numbers (data quantity). Thus, the calculations in this section will involve largely finite sums. However, given the prevalence of continuous time models in quantitative finance I set up the problem in a continuous time frame-work and will embed the discrete considerations of this section into this framework. The first problem that has to be addressed in this context is to define the object to which the estimator is converging. In the context of the mixture distribution above this is given by $\hat{\sigma}_{[t_1,T]}$, i.e. when calculating the estimator convergence the heterogeneity will also play a role in the target of the convergence. The following theorem extends the convergence speed results for i.i.d. normal random variables to i.-non-i.d. normal random variables.

Theorem

Let $(X_i)_{i\in N}$ be a family of independent and $N(0,(\sigma_i^2)_{i\in N})$ distributed random variables and let

$$S_n^2 := \frac{X_1^2 + ... + X_n^2}{n} \tag{19}$$

be the sample variance. Here, σ_i denotes $\sigma(t_i)$.

Then the convergence speed of S_n towards $\frac{1}{n}\sum \sigma_i^2$ in quadratic norm, i.e. the variance of the error, is

$$E\left[\left(S_n^2 - (\frac{1}{n}\sum \sigma_i^2)\right)^2\right] = 2\frac{\frac{1}{n}\sum \sigma_i^4}{n}\ . \tag{20}$$

Proof Using only elementary properties of normal distributions regarding higher moments (see Shao (1998)) one calculates

$$E\left[\left(S_n^2-(\frac{1}{n}\sum\sigma_i^2)\right)^2\right]=E\left[\left(\frac{1}{n}(X_1^2+...+X_n^2)-(\frac{1}{n}\sum\sigma_i^2)\right)^2\right] \tag{21}$$

$$=E\left[\left(\frac{1}{n}(X_1^2+...+X_n^2)\right)^2\right] \tag{22}$$

$$-2(\frac{1}{n}\sum\sigma_i^2)E\left[\left(\frac{1}{n}(X_1^2+...+X_n^2)\right)\right]+(\frac{1}{n}\sum\sigma_i^2)^2 \tag{23}$$

$$=E\left[\left(\frac{1}{n}(X_1^2+...+X_n^2)\right)^2\right]-(\frac{1}{n}\sum\sigma_i^2)^2 \tag{24}$$

$$=\frac{1}{n^2}E\left[\left(X_1^2+...+X_n^2\right)^2\right]-(\frac{1}{n}\sum\sigma_i^2)^2 \tag{25}$$

$$=\frac{1}{n^2}\left(\sum_i E\left[X_i^4\right]+2\sum_{i,j,i\neq j}E\left[X_i^2X_j^2\right]\right)-(\frac{1}{n}\sum\sigma_i^2)^2 \tag{26}$$

$$=\frac{1}{n^2}\left(3\sum\sigma_i^4+2\sum_{i,j,i\neq j}\sigma_i^2\sigma_j^2\right)-\left(\frac{1}{n^2}\left(\sum\sigma_i^4+2\sum_{i,j,i\neq j}\sigma_i^2\sigma_j^2\right)\right) \tag{27}$$

$$=2\frac{1}{n^2}\sum\sigma_i^4 \tag{28}$$

$$=2\frac{\frac{1}{n}\sum\sigma_i^4}{n}\,. \tag{29}$$

Remark Since every data interval only contains a finite number of data points this statement is sufficient. However, given an equidistantly spaced data set with frequency η and assuming that the $(X_{t_i})_{i\in\{0,\dots,n\}}$ spans the interval $[t_1,T]$ and includes the endpoints, i.e. $X_{t_m}=X_{t_1+\frac{m}{n}(T-t_1)}$, the statement above can be restated to

$$E\left[\left(S_n^2-(\frac{1}{n+1}\sum\sigma_i^2)\right)^2\right]=2\frac{\frac{1}{\eta(T-t_1)}\int_{t_1}^T\sigma_s^4\zeta(ds)}{\eta(T-t_1)} \tag{30}$$

for the discrete measure ζ assigning a weight of 1 at the points t_i and zero otherwise.

Definition The convergence error $\alpha_\sigma(t_1,T)$ for a finite family of centered i-non-i.d. normal random variables $(X_{t_i})_{i\in\{0,\dots,n\}}$ that includes the endpoints of the intervals and that is equally spaced with frequency η is defined by

$$\alpha_\sigma^2(t_1,T):=E\left[\left(S_n^2-(\frac{1}{n+1}\sum\sigma_i^2)\right)^2\right]=2\frac{\frac{1}{\eta(T-t_1)}\int_{t_1}^T\sigma^4(s)\zeta(ds)}{\eta(T-t_1)}\,. \tag{31}$$

2.4 Error Combination and Minimal Bias Data Sets

The objective of this section is to combine the errors derived in the last two sections. When aggregating different sources of errors one can rely on classical results from the field of Error Analysis (see Ku (1966)) which is a field concerned with the question in which way individual results and their corresponding errors (typically experimental results with some measurement error) are propagated when one looks at functional relationships of the individual results. Specifically, given the individual errors $\alpha_\sigma(t,T)$ and $\beta_\sigma(t,T)$ and assuming that they are independent, it is known that the average measurement error of some function of the errors, $f(\alpha_\sigma(t,T),\beta_\sigma(t,T)) = \alpha_\sigma(t,T) + \beta_\sigma(t,T)$, is given by

$$sd(f(\alpha_\sigma(t,T),\beta_\sigma(t,T))) = \sqrt{\frac{\partial f}{\partial \alpha_\sigma(t,T)} \cdot \alpha_\sigma^2(t,T) + \frac{\partial f}{\partial \beta_\sigma(t,T)} \cdot \beta_\sigma^2(t,T)} \quad (32)$$

$$= \sqrt{\alpha_\sigma^2(t,T) + \beta_\sigma^2(t,T)}\,. \quad (33)$$

Here, $sd(\cdot)$ denotes the standard deviation (measurement error) of some quantity. If the time dynamics of the characteristics process is deterministic then the representativeness error is deterministic and, consequently, it is independent of the other error by construction and the conditions of the statement above are satisfied. If one would like to have the time dynamics of the process $\sigma(t)$ to be random one can, in principle, model the dynamics of $\sigma(t)$ as a subordinator, i.e. an almost surely increasing Lévy Process (see (Clark, 1973)) in order to have both, a random measurement quantity as well as the correspondence between 'statistical novelty' and 'novelty' as measured by the representativeness metric laid out in section 2.2.

Given those considerations the overall error $\gamma_\sigma(t,T)$ associated with a data set spanning the interval $[t,T]$ is given by

$$\gamma_\sigma(t,T) = \sqrt{\alpha_\sigma^2(t,T) + \beta_\sigma^2(t,T)}. \quad (34)$$

Given this functional form it is natural to restrict the space of possible data sets. It is natural to ask which of the most recent observations can be used to optimally estimate the current value of some parameter, i.e. to find the optimal data set within the set $\{[x,T]; x \leq T\}$. The optimal data set is defined as the data set minimizing $\gamma_\sigma(t,T)$, i.e. by

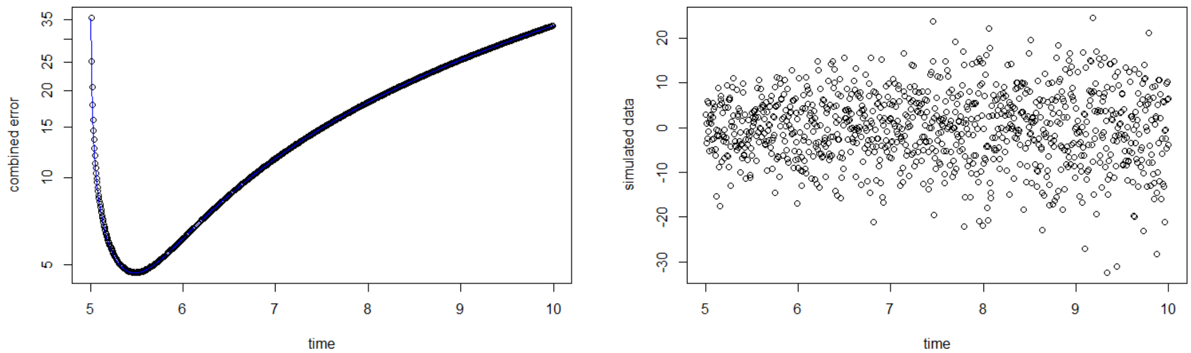

Figure 2: **Simulation:** Depicted are the theoretical and empirical $\gamma_\sigma(5, x)$, i.e. in this graphics we consider the problem from the left and try to infer $\sigma(5)$ with data from the interval $[5, x]$, for the situation where $\sigma(t) = t$ and the data set spans the time interval $[5, 10]$. The results of the simulation are based on 10.000 simulations per time step and a data density of $\eta = 200$. The blue line is the theoretical magnitude of the error combination procedure from the definition. The panel on the right depicts one sample data set.

the condition

$$argmin_{x\in[0,T]}\gamma_\sigma(x,T) = \sqrt{\alpha_\sigma^2(x,T)+\beta_\sigma^2(x,T)} \tag{35}$$

$$= \sqrt{2\frac{\frac{1}{\eta(T-x)}\int_x^T \sigma^4(s)\zeta(ds)}{\eta(T-x)} + \left(\frac{1}{T-x}\int_x^T \sigma^2(s)\,ds - \sigma^2(T)\right)^2}. \tag{36}$$

For an illustration comparing this theoretical prediction with a simulation, see Figure 2.

2.4.1 Existence of the optimal data set

Given the combined error it is immediately possible to show the formal existence of an optimal data set:

Theorem

Under the assumption that $\sigma(t)$ is an integrable function there exists an $x^* \in [x, T-2\eta]$ that minimizes the expression above. Here, η refers to the data frequency.

Proof Let $g(x) = \sqrt{\alpha_\sigma^2(t,T) + \beta_\sigma^2(t,T)}$. Since $\sigma(t)$ is integrable all integrals in the expression above are continuous functions. Hence, g is a combination of continuous functions and the functions in the denominator are non-zero for $x \leq T - 2\eta$. Consequently, g is continuous. It is also bounded on the compact interval $[0, T - 2\eta]$ and therefore has a minimum in this interval (see, e.g., Rudin (2013)).

The 2η in the statement above is chosen since one needs to have at least two data points in order to estimate a variance and one can consequently ignore the singularity in the first term of the expression.

2.4.2 Uniqueness of the optimal data set

The purpose of this subsection is to illustrate that, in general, it cannot be guaranteed that the optimal data set is unique. See Figure 3 for an example where two data sets exhibit the same combined error. Here, one data set is longer with a higher representativeness error and a smaller estimator convergence error while the other one is shorter with a smaller representativeness error and a higher estimator convergence error. From a numerical perspective this nonuniqueness can make the length of the data set a very unstable quantity, since any data driven algorithm can change from one minimum to the other with only small perturbations. However, this is exactly the insecurity one would like to mimic in the presence of structural breaks: In a setup where structural breaks are possible one cannot be sure whether the informational content of some new data point is based on a change in the underlying structure and, consequently, constitutes a sample from a new distribution or rather due to an outlier drawn from the old distribution. Hence, if there is an action tied to the value of estimators (as, e.g., in the case of banking regulation, see below) this change in the data set and the corresponding change in the estimators can translate into a very fluid business strategies. An example of a repeated oscillation between two data sets of different lengths can be seen in Figure 5 in the summer of 2013. However, it is also briefly present in the financial crisis and in 2015.

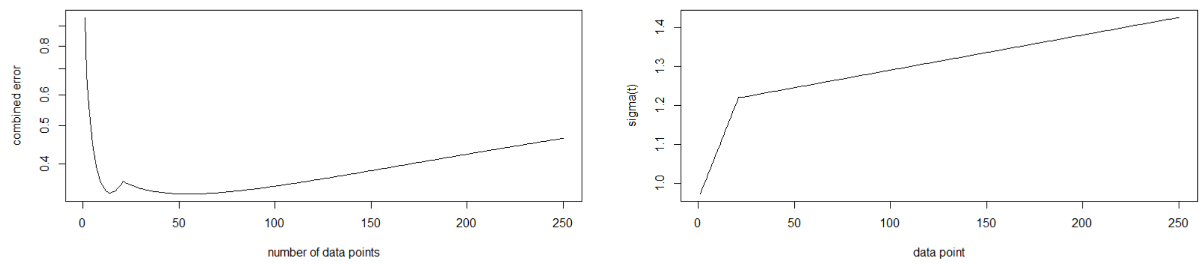

Figure 3: **Nonuniqueness of optimal data sets**: This figure shows the combined error $\gamma_\sigma(0, x)$ (left) of data sets generated by observations from a process with characteristics $\sigma(t)$ as shown on the right.

2.5 Applications

The objective of this section is to go through two simple examples for which the method introduced above can be applied. As mentioned in the introduction the problem of an optimal data reduction consists of two parts: First, one has to have an opinion of the degree of non-stationarity and then one has to derive an optimal data set based on the degree of non-stationarity. The first example compares single jumps of different sizes in the characteristics. The purpose of the example is to support the basic intuition that a smaller jump in the characteristics should lead to 'less data reduction' after the jump. In the next subsection I perform an impact analysis of a paragraph in the Basel 3 framework.

2.5.1 Jumps and Data Reduction

In this section I briefy demonstrate how the results are infuenced by the degree of nonstationarity. To this end, consider a data set of length 500 with a single structural break of size κ in the underlying characteristic σ at data point 250, i.e. before the break the value of the characteristic is given by $\sigma = 1$ while after the break it is at $\sigma = 1 + \kappa$. We expect that larger κ's lead to increased data reduction and less usage of prebreak data while smaller κ's have the opposite effect. Implementing eq. (36) for this particular example yields results that are fully compatible with this intuition (see figure 4). Here the values of the κ's are set to 0.002 (black), 0.01 (red), 0.05 (blue), and 0.1 (green). When the jump is small enough

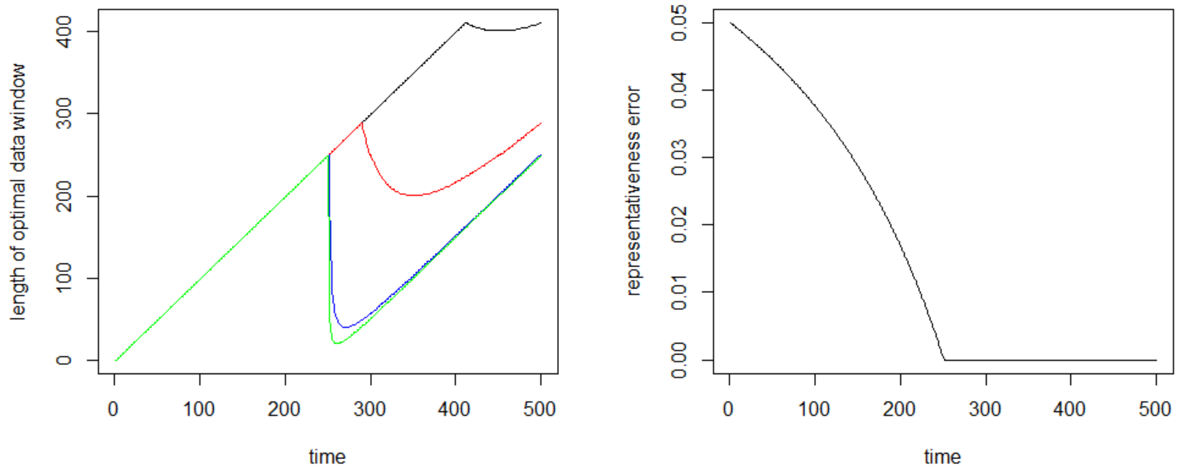

Figure 4: **Optimal Data Length for Jumps of different sizes**: This figure displays the degree of data reduction due to different jump sizes $\kappa \in \{0.002, 0.01, 0.05, 0.1\}$. In the right panel the representativeness error for the last data point is shown, i.e. $\beta_\sigma(x; 500)$ for $\kappa = 0.1$. For $x > 250$ the error is zero since only post break data is used while for $x < 250$ some non-representative data is in the sample. The left panel displays the length of the data window at the respective point in time. It can be seen that prior to the break all data is used while the data window collapses after the break. The collapse is more severe the larger the jump is.

almost all of the prebreak data is still used at the end of the sample (black line) while larger jumps lead to higher degrees of data reduction.

2.5.2 Application: Impact Analysis of a Paragraph in Basel III

The objective of this section is to apply the methodology introduced above on a risk management application for which the sensitive modeling choice of the methodology is largely taken out of my hands. To this end one has to make two adjustments to the formal methodology above: First, 'novelty' has been defined as genuine new values of parameters in order to ensure that stochastic novelty and novelty as measured by the representativeness metric coincide. In applications that is not a necessary requirement: Different periods of, say, low volatility realise themselves in different institutional frameworks and, consequently, are 'novel' in their own regard. Hence, I will simply drop this assumption. One also has to provide an implementable algorithm of the methodology above: In Equation (36) the right term is modeled via the representativeness metric. In the left term I will use the sample fourth moment.

Quantitative Risk Measures like Value-at-Risk or Expected Shortfall are employed when determining the capital requirements of financial institutions. The following paragraph from the Basel 3 framework is dealing with the data sets that have to be used in order to calculate quantitative risk measures, specifically under which conditions they can be contracted:

'The supervisory authority may also require a bank to calculate its Expected Shortfall using a shorter observation period if, in the supervisor's judgement; this is justified by a significant upsurge in price volatility. In this case, however, the period should be no shorter than 6 months.' Basel-Committee (2016)

This paragraph contains all the ingredients that are part of the methodology introduced in this paper: Under a significant inflow of new information (measured by a 'significant upsurge in price volatility') the supervisory authority can require banks to reduce the reliance on older data. The interesting question here is to which degree the triggering of this paragraph can influence the dynamics of risk metrics. In order to approach this question one first needs to establish a connection between the parameter discussed in the methodology, i.e. variance, and the risk metric mentioned here: I assumed the data generating process is locally gaussian, i.e. that the aggregate distribution over an interval is a centered gaussian mixture distribution. Centered Mixture distributions exhibit excess kurtosis and since excess kurtosis is a measure of extreme values Westfall (2014) it is adequate to conclude that the expected shortfall of a gaussian mixture distribution is bounded from below by the expected shortfall of the gaussian distribution with the same variance as the gaussian mixture distribution. However, the expected shortfall of a gaussian distribution is a deterministic function of its variance. More specifically, the expected shortfall of a normal distribution is given by the value-at-risk, i.e. a quantile, times a quantile dependent constant. Consequently, the variance estimator can be used to establish a lower bound on the expected shortfall of the mixture distribution.

I also need to specify the representativeness metric that is used. Here, I again rely on the fourth moment since extreme events naturally signify a structural break (see Westfall (2014)) and since an increase in a higher even moment is also highly related to an increase in variance. In order to avoid very short time series I also use a normalising constant of 100 and arrive at the representativeness metric

$$R(t_1, t_2) = \sum_{k=t_1}^{t_2} \frac{r_k^4}{100} \,. \tag{37}$$

Note that the representativeness metric serves as a model of change of the underlying parameter of a time series. There is no clear connection between the occurence of extreme events in some time series and some form of non-stationarity. Here, the occurence of extreme events serves as an *indicator* of non-stationarity. The better the indicator the more weight it can receive. Since my goal is a simple illustration of the concept I am acknowledging the imperfections of using extreme events and fourth moments as indicators of non-stationarity by severely downweighting them (a weight of zero would lead to using the full historical data set for every point in time while larger weights would increase the representativeness error which would yield shorter data sets).

For this particular choice of representativeness metric the corresponding data windows can easily be calculated by constructing time point dependent backward looking representativeness errors and convergence errors (see eq. (36)). Subsequently, the estimated variance can be used in order to calculate the lower bound on the value-at-risk mentioned above. For a DAX data set of daily closing prices spanning the time frame from April 1st 2001 to August 31st 2017 the results are depicted in Figure 5. In order to achieve a certain comparability I also depict the results for the Value-at-risk for a fixed rolling window estimation for which the rolling window is fixed to the average of the adaptive data window, i.e. 372 data points.

It can be noted that the average value of the risk measure is lower for the adaptive data method whereas the volatility is higher. The author is not aware of a discussion of the interesting question regarding what the optimal tradeoff between the mean and variance of different risk metrics is. Recent papers have, however, focused on the mean

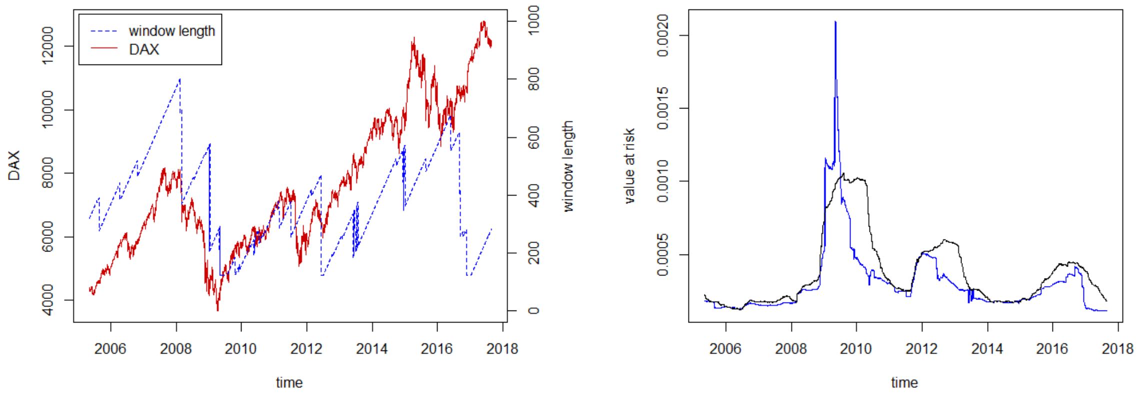

Figure 5: **Empirical Example** Depicted are the optimal data windows (left panel, blue) as well as the levels of the DAX (left panel, red). In the right panel the corresponding 97.5% value-at-risk estimates that are based on the adaptive data method (blue) and a fixed rolling window (black) are depicted.

as a quality criterion (see, e.g., Lau (2015b)) since it can be interpreted as bound capital and consequently it is desirable for it to be as low as possible. However, since especially in times of crisis it is extremely difficult to acquire new capital it is not clear whether an average low capital requirement is actually desirable.

In addition, it is also clear from the analysis that the data reduction takes place at the beginning of the financial crisis of 2007 and that the increase in risk assessment is much steeper than for the rolling window estimation. This is not surprising since 'a significant increase in price volatility' is typically the outcome of a beginning financial crisis and a data reduction in this scenario leads to a relative increase of crisis data in the data set used for the estimation of the risk metric. It is fair to assess that triggering this statute at the beginning of a financial crisis would lead to a significant increase in capital requirements in a market environment that is already in turmoil. As such, triggering this paragraph would add to the problem of procyclicality that is inherent to tying the capital requirements of banks to the risks of their assets (see, e.g., Danielsson *et al.* (2001), Pennacchi (2005)) Consequently, one could argue that the regulator has a tool in hand that he could use to effectively shut down the financial system. The flipside of the argument, however, is equally disturbing: If at the beginning of a financial crisis the only thing that keeps a bank alive from a quantitative risk management perspective are the smoothing effects from data prior to the crisis (which is, arguably, not relevant anymore) this is not comforting.

2.6 Conclusion

Data plays a crucial role in our understanding of the world around us. In Economics and Finance observations from the past have always been eyed with suspicion by empirical researchers and have regularly been either down-weighed or excluded from their studies. At the same time Economic and Finance theory have worked under the assumption that the past is fully representative of the present while, for a much longer period, physicists have devoted significant ressources to discuss the validity of this assumption. Given that data restriction is common practice and given the influence that empirical research has on policy issues it is of utmost importance to derive a quantitative criterion for data selection purposes. Otherwise, it is a degree of freedom which can be exploited. For a special situation this paper introduced a method to solve this problem. In the process I did not address the question which 'representativeness metric' is good for a particular situation since I believe it to be a problem specific issue which has to be argued for on a case-by-case basis. If one is interested, e.g., in the current risk metrics of some oil futures then any action by the OPEC or Russia regarding their oil exporting policy should appear in my related model of representativeness. However, this particular model does not play a role when one is interested in the risk metrics of, say, a european life insurance company since the oil market and the european life insurance market can assumed to be unrelated. Instead, in a life insurance context, one might want to build a model of representativeness based on, e.g., progress in medicine and geriatrics.

However, in order to arrive at optimal data sets one has to acknowledge that economic reality shifts over time and that data from different points in time is qualitatively different. This assertion is incompatible with a variety of concepts that lie at the heart of not only Finance and Economics but the scientific method in general: In Finance, a shifting reality can only mean that the concept of 'efficiency' is not clearly defined because it is based on a perfect statistical perception which, in view of the results in this paper, is impossible to achieve. In fact, I even showed that the 'optimal degree' of statistical perception is not unique. Given how deeply rooted efficiency and the related notion of rationality are in Economics it is somewhat elusive what would be a good substitute in a world that is not stable over time. However, very recently, in the aftermath of the 2007-2009 financial crisis the finance literature has moved on towards adaptive markets where the term 'adaptive' refers to *the multiple roles that evolution plays in shaping behavior and financial markets* (see Lo (2017)) and the method presented here can serve as a quantitative modeling device which can carry evolutionary information.

More generally, a 'shifting reality' also sheds a light on the so-called replication crisis (see Open Science Coll. (2015)). Currently, the blame for the missing degree of replicability of studies in many fields of the social sciences is placed at the feet of statistical techniques (see, e.g., Ioannidis (2005) and in a financial context also Bailey *et al.* (2014)) and while I acknowledge the misincentives that are inherent to some statistical techniques, it is also perfectly conceivable that the underlying object that is studied is changing over time and that, as a consequence, some experiments do not replicate because the behavior of subjects is shaped by some form of 'Zeitgeist' and cultural evolution. This point has already been raised in the psychology literature (see Greenfield (2017)) but it is worthy noting since the underlying concept here, again, is ergodicity which is a notion that social scientists are rarely aware of.

In addition to these conceptual questions the technique presented here spawns a host of formal problems that can be followed up on: In order to keep the presentation to its essential parts I restricted myself to a subclass of semimartingales but it is natural to perform the analysis for more general semimartingale processes. In this context, however, one has to deal with a variety of complicated filtering problems since one has to deal with the attribution problem pointed out in section 2.2. In this paper I worked under the assumption that I know precisely the location and size of structural breaks since I wanted to focus on the problem of an optimal data identification. In the statistics literature, however, jump detection is a notoriously difficult problem with a large number of possible solution techniques. Hence, it is natural to combine different possible representativeness metrics or, more generally, one can treat the representativeness metric as a stochastic process in its own right (e.g. by assuming it to be a subordinator). However, in this instance the two individual errors will not be independent anymore and one would have to develop additional methods for the error aggregation step in order to solve the corresponding data selection problem. If one is in a situation where the data generating process is not self-similar and one has qualitatively different behavior on different time-scales it is also conceivable to leave the class of semimartingales and consider, e.g., fractional processes to set up a problem of an optimal sample frequency where one is interested in the extraction of a long-term signal from a data set of limited length and an increase in sampling frequency leads to an increase in data quantity but introduces a bias caused by the non-self similarity. Lastly, in this paper I restricted myself to processes that are locally normal distributed and argued on the basis of the convergence speed of mixture distributions which also leaves room for a variety of generalisations regarding distributional assumptions.

3 A Convergence Speed Dependent Data Quantity Definition for Weighted Observations

The results of this part of the thesis have been published as Krause (2019b).

Abstract

Data Quantity plays a crucial role in the estimation of risk measures since 'more data' leads to a better estimator convergence and consequently to a better risk assessment. The objective of this paper is to use this relationship between data quantity and estimator convergence to formally derive a measure of data quantity for estimators based on weighted observations. For the case of a variance estimation and using exponentially weighted observations this procedure leads to analytical formulas for the implied measure of data quantity. As such, this paper specifies the theoretical underpinnings of measures of data quantity which have been present in the literature (Effective Number of Scenarios) and, as an application, demonstrates the effect of the specific measure of data quantity on risk assessment.

3.1 Introduction

It is widely recognized that parameter instability is a crucial issue when analyzing financial time series (see, e.g., Stock and Watson (1996), Pesaran and Timmermann (2007), Giacomini and Rossi (2009), Inoue *et al.* (2017)). In a situation where one is faced with making an evidence-based business or investment decision it is crucial to conceptually include this instability in the analysis and the corresponding decision making process. In principle, there are two formal methods to cope with statistical instability: On the one hand one can attempt to accurately model the instability and analyze a given situation in light of the chosen model. Given the tendency of economic and financial systems to induce 'unintended consequences' this approach should be used with caution and only for contained situations of managable complexity. The alternative is to restrict the data input to some statistical technique which estimates a decision-relevant quantity by excluding or downweighting data one does not deem representative or relevant for the problem at hand. In Quantitative Finance we typically rely on the latter method which gives rise to popular techniques such as rolling-window estimation and data weighting procedures that attempt to strike a balance between using as much objective information as possible while fading out information that is deemed irrelevant for the current situation. In this context, however, it is of crucial importance to clarify to which degree one should limit the possible input data to some statistical technique or, more precisely, which degree of data quantity is justified in a particular setup. This question is relevant for a variety of reasons: First, it is a degree of freedom that has to be chosen and therefore one would like to have an argument that goes beyond 'experience of the researcher' or 'standard in the literature' which are the common arguments in the finance literature (see Inoue *et al.* (2017)). Second, 'data quantity' defines the amount of evidence one can utilise in statistical techniques. As such, it is directly related to concepts like confidence intervals and estimation errors. In a world that obeys the statistical assumptions an infinite amount of data would lead to definite conclusions. As such, one has to be aware that the reliance on any form of data reduction conceptually acknowledges fundamental change as well as the awareness that statistical techniques do not yield incontrovertable evidence and have to be augmented with additional arguments in order to justify a decision.

In order to determine an appropriate degree of data reduction one first has to have a concept of data quantity. While it is clear how much data is used when estimating some quantity with data from a rolling window of, say, 500 data points, it is not immediately clear how much data one is effectively using when one relies on, say, an exponential weighting scheme with a data decay parameter of 0.99 in which the final observation receives a normalized weight of 1 and every observation before the final one receives a normalized weight of 0.99^n where n denotes the distance to the final observation. Suitable candidates for such measures of data quantity which are called 'Effective Number of Scenarios' (ENS) have been derived in Meucci (2012) who also coined the term ENS. However, what is currently missing from the literature is a framework that specifies the relevant parameters and concepts that are needed to pick an appropriate candidate within this general class of candidate measures. To introduce this framework and document the assumptions one has to put into place in order to arrive at an appropriate choice is the objective of this paper.

I start from the premise that the value of more data is given by a lower estimation error (as measured by some loss function). Consequently, there is a natural relationship between data quantity and estimation errors. Hence, it is natural to define the effective number of scenarios of some weighting scheme by the (rounded) number of the data quantity of a window estimator that yields the same average convergence error. The formal introduction of this definition is the objective of section 3.2. In section 3.3 I calculate an explicit example and derive the ENS-formula for a class of exponentially weighted variance estimators. In section 3.4 I document that the choice of the a particular ENS measure can have significant effects on the size of confidence intervals of risk metrics. In section 3.5 I conclude.

3.2 Convergence Rate Based ENS

The objective of this section is to provide an analytical condition that can be used in order to calculate the ENS of weighted observations. The condition introduced here is based on convergence errors. Consequently, in order to connect the properties of different estimators of the same quantity via convergence errors it is useful to assume that they are unbiased. For unbiased estimators the target of two alternative estimation techniques of the same quantity is the same for any data quantity and consequently the difference to the true value can be used as a quantitative criterion. Consequently, I assume to have access to a class of unbiased estimators $\hat{\sigma}^2(\omega, \cdot)$ of the same quantity σ^2 that is defined across different weight vectors ω, i.e.

$$\hat{\sigma}^2 : W \times X^n \to R \ , \quad (\omega, (X_1, ..., X_n)) \mapsto \hat{\sigma}^2(\omega, (X_1, ..., X_n)) \ , \tag{38}$$

where X^n denotes a set of n i.i.d. random variables $X_1, ..., X_n$. In this setup I am interested in assigning the ENS as a function of the weights. However, the following definition should be seen as being conditional on the estimation technique.

Definition: The **Effective Number of Scenarios (ENS)** is a function denoted by $ENS^{\hat{\sigma}^2}(\cdot)$ that maps a vector of weights $\omega \in W$ to a real number, i.e.

$$ENS^{\hat{\sigma}^2} : W \to R, \ \omega \mapsto ENS^{\hat{\sigma}^2}(\omega) \ . \tag{39}$$

Here, $\omega = (\alpha_1, \alpha_2, ...) \in W$ denotes a vector of weights, i.e. it fulfills $\sum_{i=1}^{\infty} \alpha_i = 1$ and $\alpha_i \geq 0$. From now on I will drop the superscript.

One needs to impose the condition that $ENS(\omega_n) = n$, where $\omega_n \in W$ denotes the vector $(\frac{1}{n}, ..., \frac{1}{n}, 0, ...)$ which has the entry $\frac{1}{n}$ in the first n components and zero otherwise. This condition is at the heart of assigning the ENS for weighted estimators since one can clearly say how many data points are used when employing equally weighted independent observations. In statistics, the value of more data is given by a lower estimation error. Consequently, there is a fundamental relationship between estimation errors and data quantity. It is natural to define the effective number of scenarios of a given weighting scheme by the (rounded) number of data quantity of a window estimator that yields the same average convergence error. In this context one has to specify a loss function $d(\cdot, \cdot)$ that measures the convergence error. Hence, the ENS notion already depends on a loss function d and the convergence speed of the estimators $\hat{\sigma}^2$.

Definition: Two vectors of observation weights ω_1 and ω_2 exhibit the same ENS_d, i.e.

$$ENS_d(\omega_1) = ENS_d(\omega_2) \tag{40}$$

if the following condition is fulfilled:

$$E\left[d\left(\hat{\sigma}^2(\omega_1, X), \sigma^2\right)\right] = E\left[d\left(\hat{\sigma}^2(\omega_2, X), \sigma^2\right)\right] . \tag{41}$$

Thus, if one can show that

$$E\left[d\left(\hat{\sigma}^2(\omega_n, X), \sigma^2\right)\right] = E\left[d\left(\hat{\sigma}^2(\omega_2, X), \sigma^2\right)\right] . \tag{42}$$

where ω_n denotes the vector with equal weights one can conclude that $ENS_d(\omega_2) = n$. The most common choice of the loss function is given by a quadratic function, i.e. $d(a, b) = (a - b)^2$, which will be used from now on.

Since the ENS definition is only based on expectations it is suitable to find the function $ENS(\cdot)$ via Monte Carlo simulation. While not necessary, it is also helpful to assume that the unbiased estimator based on ω_n is efficient, i.e. has minimum variance. This ensures that a weighted estimator based on a data set of n observations always has a lower ENS than n. Also note that the same weighting scheme can lead to a different ENS in a situation where the efficiency of the weighted estimators are not similar: The more efficient an estimator the less data is needed in order to reach a certain convergence error. Consequently, if one would have a situation where one has two estimators of the same quantity that are based on the same weighting scheme and have different levels of efficiency then using them on the same data set would yield a higher ENS for the more efficient estimator.

3.3 The ENS formula for the exponentially weighted variance estimator

3.3.1 Infinite Number of Observations

Definitions

For specific choices of ω and σ and d it is possible to provide analytical results for the function $ENS_d(\cdot)$. In this section I calculate $ENS_2(\cdot)$ for the (weighted) sample variance estimator. Here ENS_2 stands for an ENS definition based on a quadratic loss function. The subscript is ommited from now on. In order to simplify the analysis further, let $X_1, X_2, ...$ be a countable collection of i.i.d. $N(0, \sigma^2)$ distributed random variables and let ω_α denote the following vector of weights

$$\omega_\alpha = (\alpha^0, \alpha^1, \alpha^2, ...) \ ; \ \alpha \in [0, 1]. \tag{43}$$

Here, α is called a 'data decay parameter'.

As for the estimators, let

$$\hat{\sigma}(\omega, X^\infty) = \lim_{n\to\infty} \hat{\sigma}(\omega, X^n) := \begin{cases} \frac{1}{n}\sum_{i=1}^{n} x_i^2 & \omega = \omega_n \\ (\frac{1-\alpha}{1-\alpha^n})\sum_{i=0}^{n-1} \alpha^i x_{i+1}^2 & \omega = \omega_\alpha \end{cases}. \tag{44}$$

Note that those estimator are unbiased since I assume to have knowledge on the mean. The variance estimator using equal weights also is efficient, i.e. it has minimum variance in the class of unbiased estimators, since it fulfills the Cramer-Rao lower bound (see Shao (1998) and Appendix A).

Derivation of the ENS function

It can easily be shown that the sample variance for i.i.d. $N(0, \sigma^2)$ distributed random variables is given by

$$E\left[\left(\hat{\sigma}^2(\omega_n) - \sigma^2\right)^2\right] = 2\frac{\sigma^4}{n} \ . \tag{45}$$

This result is a special case of the calculation given in section 2 (eq. (21) - (29)). Consequently, if one can argue that

$$E\left[\left(\hat{\sigma}^2(\omega_n) - \sigma^2\right)^2\right] = E\left[\left(\hat{\sigma}^2(\omega_\alpha) - \sigma^2\right)^2\right] \tag{46}$$

for some α, then $ENS(\omega_\alpha) = n$.

$$E\left[\left(\left\{(1-\alpha)\sum_{i=0}^{\infty}\alpha^i x_{i+1}^2\right\} - \sigma^2\right)^2\right] \tag{47}$$

$$= E\left[\left\{(1-\alpha)\sum_{i=0}^{\infty}\alpha^i x_{i+1}^2\right\}^2 - 2\sigma^2\left\{(1-\alpha)\sum_{i=0}^{\infty}\alpha^i x_{i+1}^2\right\} + \sigma^4\right] \tag{48}$$

$$= E\left[\left\{(1-\alpha)\sum_{i=0}^{\infty}\alpha^i x_{i+1}^2\right\}^2\right] - \sigma^4 \tag{49}$$

$$= (1-\alpha)^2\left\{\sum_{i=0}^{\infty}\alpha^{2i}E\left(x_{i+1}^4\right) + 2\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}\alpha^{i+j}E\left(x_{i+1}^2 x_{j+1}^2\right)\right\} - \sigma^4 \tag{50}$$

$$= \sigma^4\left(3\frac{(1-\alpha)^2}{(1-\alpha^2)} + 2\frac{\alpha(1-\alpha)}{(1-\alpha^2)} - 1\right) \tag{51}$$

$$= 2\sigma^4\frac{1-\alpha}{1+\alpha} \ . \tag{52}$$

Thus, using equations (46) and (53) it can be concluded that

$$\frac{1+\alpha}{1-\alpha} = n \ . \tag{53}$$

This function passes an initial screening: For α close to 1 (i.e. almost no down-weighting of past observations) the corresponding number of scenarios n is infinitely large. On the other hand, for an $\alpha = 0$ the corresponding estimator is based solely on x_1^2 and consequently the number of scenarios should be equal to 1. This is also the case.

3.3.2 finite number of observations

Next, consider the more general situation, where one only has access to a finite number of observations $X_1, ..., X_n$, i.e. where one is interested in $\hat{\sigma}(\cdot, X^n)$. In this situation the objective is to calculate

$$E\left[\left(\left\{\frac{1-\alpha}{1-\alpha^n}\sum_{i=0}^{n-1}\alpha^i x_{i+1}^2\right\} - \sigma^2\right)^2\right] \tag{54}$$

$$= E\left[\left\{\frac{1-\alpha}{1-\alpha^n}\sum_{i=0}^{n-1}\alpha^i x_{i+1}^2\right\}^2 - 2\sigma^2\left\{\frac{1-\alpha}{1-\alpha^n}\sum_{i=0}^{n-1}\alpha^i x_{i+1}^2\right\} + \sigma^4\right] \tag{55}$$

$$= E\left[\left\{\frac{1-\alpha}{1-\alpha^n}\sum_{i=0}^{n-1}\alpha^i x_{i+1}^2\right\}^2\right] - \sigma^4 \tag{56}$$

$$= \left(\frac{1-\alpha}{1-\alpha^n}\right)^2\left\{\sum_{i=0}^{n-1}\alpha^{2i}E\left(x_{i+1}^4\right) + 2\sum_{i=0}^{n-1}\sum_{j=i+1}^{n-1}\alpha^{i+j}E\left(x_{i+1}^2 x_{j+1}^2\right)\right\} - \sigma^4 \tag{57}$$

$$= \sigma^4\left(\frac{1-\alpha}{1-\alpha^n}\right)^2\left\{3\frac{1-\alpha^{2n}}{1-\alpha^2} + 2\frac{(\alpha^n-1)(\alpha^n-\alpha)}{(\alpha-1)^2(\alpha+1)}\right\} - \sigma^4 \tag{58}$$

$$= 2\sigma^4\frac{(\alpha-1)(\alpha^n+1)}{(\alpha+1)(\alpha^n-1)}\,. \tag{59}$$

Consequently it can be concluded that

$$n = \frac{(\alpha+1)(\alpha^n-1)}{(\alpha-1)(\alpha^n+1)}\,. \tag{60}$$

Given this equation one can make the following observations:

1. For $n \to \infty$ this formula simplifies to the expression derived in section 3.1.

2. This function is a member of the class of candidate data quantity measures that has been derived in Meucci (2012). In the paper it is argued that any function from the class

$$ENS_d(\omega_\alpha, n, \gamma) = \left(\frac{(1-\alpha)^\gamma(1-\alpha^{n\gamma})}{(1-\alpha^n)^\gamma(1-\alpha^\gamma)}\right)^{-\left(\frac{1}{\gamma-1}\right)} \tag{61}$$

can serve as a reasonable candidate for the definition of effective number of scenarios since they all fulfill minimal reasonability assumptions. In this context the reasonability is associated to a variety of desirable properties of which, in this paper, only the condition $ENS(\omega_n) = n$ is used. Using only elementary algebra it can be seen that the definition resulting from the calculation above corresponds to the choice $\gamma = 2$ in the class of functions presented in Meucci (2012).

3. The result derived above is based on the convergence rate of the sample variance estimator for i.i.d. normal distributed random variables and a quadratic loss function. Consequently, it uses the structural properties of the distribution as well as those of the estimator. For other estimators and/or other distribution assumptions and/or other estimation methods of the same quantity the corresponding convergence rates can be different and, as a consequence, the definition of effective number of scenarios might differ as well. This observation leaves room for a variety of follow up questions.

3.4 The effect of different ENS definitions

The objective of this section to illustrate the practical effects of the alternative definition of ENS introduced in this paper. To this end, I want to rely on the procedure for the calculation of confidence intervals of risk metrics presented in Meucci (2012) and perform a side-by-side analysis. The procedure consists of the following three steps:

1. Specify an empirical scenario probability f_X, i.e. an empirical density of risk factors X, e.g. P&L's or returns, with corresponding probabilities p_j:

$$f_X \approx \{x_j, p_j\}_{j=1,\ldots,J} \ .$$

2. Specify a weighting scheme and calculate the ENS $\tilde{J}_1$, $\tilde{J}_2$ and round them to the nearest integer. Here, $\tilde{J}_1$ and $\tilde{J}_2$ denote the ENS definitions that correspond to the choices of $\gamma = 1$ and $\gamma = 2$, respectively, in the set of possible candidate measures defined by eq. (61). In this context $\gamma = 1$ corresponds, in the limit, to the exponential of the Shannon entropy (see Meucci (2012)) while $\gamma = 2$ is the choice based on the considerations above.

3. Repeat the following a number of times: Draw $\tilde{J}_{\{1,2\}}$ indepedent equally weighted scenarios from the empirical distribution and calculate the respective risk numbers $\{\tilde{\sigma}^2(s)\}_{s=1,\ldots,S}$.

After going through those steps one has access to a distribution of the $\tilde{\sigma}^2(s)$. The confidence band is then specified by an appropriate upper and lower quantile of the distribution of $\{\tilde{\sigma}^2(s)\}_{s=1,\dots,S}$ (see Meucci (2012)).

In the following I carry out the analysis for two situations. In the first one I rely on a simulated data set while in the second I use an actual financial data set.

3.4.1 Approximately normally distributed data

In order to start one needs to specify the empirical distribution. In this section I use a 10000 point sample from the normal distribution, i.e.

$$f_X = \left\{ x_j, \frac{1}{2\pi} \exp\left(-\frac{x_j}{2}\right) \right\}_{j=1,\dots,10000} . \tag{62}$$

The next step of the method requires me to specify a weighting scheme. To this end, I will view the sample above as ordered in time. In view of the results in this paper I will choose exponential weighting with a data decay parameter of $\alpha = 0.99$. The corresponding number of scenarios is given by

$$\tilde{J}_1 = e^{-\sum_{j=1}^{J} p_j \ln p_j} = 270.467 \approx 270 \tag{63}$$

$$\tilde{J}_2 = \frac{(\alpha+1)(\alpha^n - 1)}{(\alpha-1)(\alpha^n+1)} = 199 . \tag{64}$$

The next step of the method requires us to repeatedly, say 1000 times, draw equally weighted samples of size $\tilde{J}_1$ and $\tilde{J}_2$ from the original sample of 10000 data points and calculate the standard deviation for every of the runs.

Given the corresponding histogram (see Figure 6, left) it is easily visible that the lower number of independent draws from the distribution in the case of the ENS definition from this paper leads to more scattered estimates while the larger number of independent draws $\tilde{J}_1$ yields estimates that are less scattered.

3.4.2 Volatility Estimation for a DAX data set

In this subsection I perform a similar analysis to the one in the last subsection. I estimate the standard deviation of daily returns using DAX data set spanning the time frame from 2001 to 2016. One can observe the same qualitative behavior as in the last subsection (see,

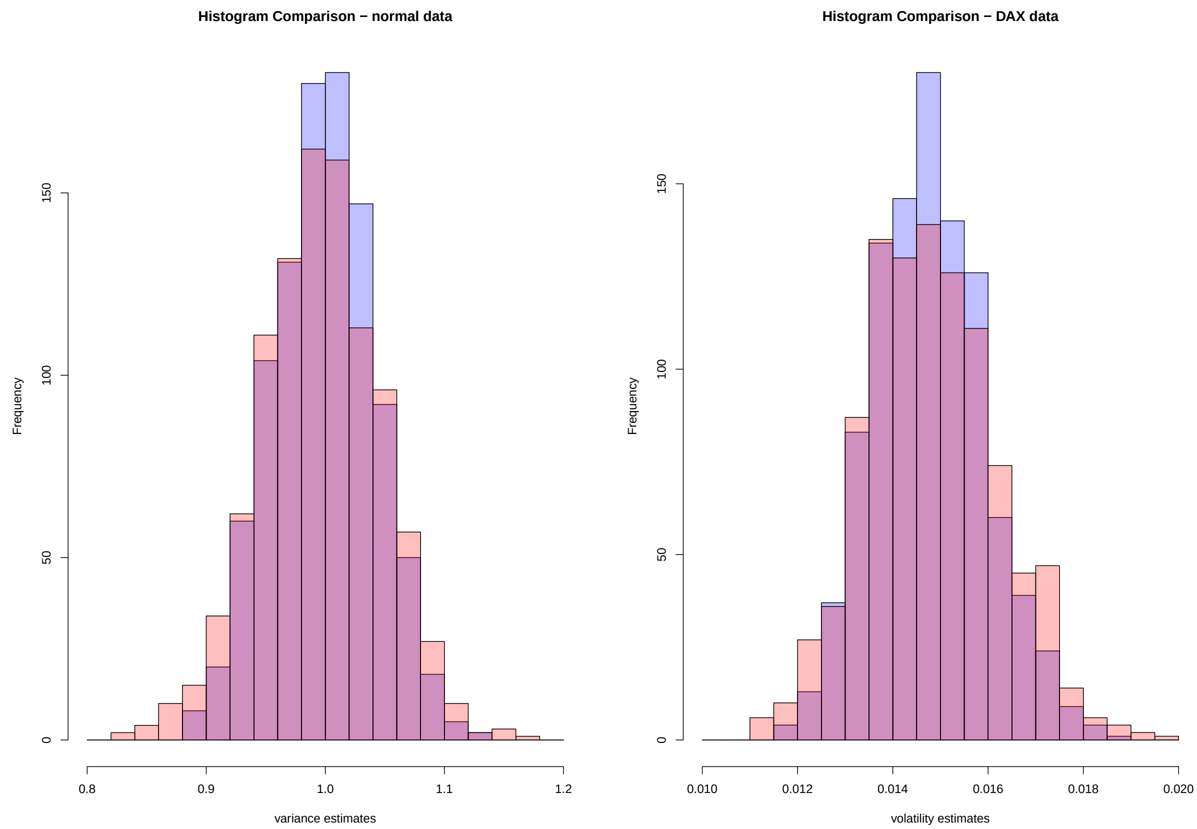

Figure 6: **Effect of different ENS definitions**: This graphics depicts the histogram of the $\tilde{\sigma}_s$ based on the different ENS methods following the procedure laid out in Meucci (2012) for normally distributed data (left) and a DAX data set (right). The blue bars reflect the exponential of the Shannon entropy, i.e. $\gamma = 1$ in the class of appropriate candidates given by eq. (61), while the red bars reflect the ENS measure derived in this paper, i.e. $\gamma = 2$.

Figure 6, right), i.e. tighter estimates in the case of $\tilde{J}_1$ and more scattered estimates in the case of $\tilde{J}_2$. Consequently, the corresponding confidence intervals are smaller for $\tilde{J}_1$ than for $\tilde{J}_2$. As such, this simulation provides an example that data quantity measures can have a significant influence on the size of confidence intervals and can constitute a critical choice in an analysis process.

3.5 Conclusion

The objective of this paper was to introduce a formal framework to specify measures of data quantities (ENS) for weighted estimations based on statistical considerations. This question is crucial in order to quantify the significance of results when calculating, e.g., risk metrics with weighted estimators. I documented the assumptions that are necessary to identify the appropriate candidate in the general class of data quantity measures derived in Meucci (2012). In order to derive the explicit results in section 3.3 I needed to impose i.i.d. normally distributed random variables, exponential weights, access to a set of unbiased estimators and a quadratic loss function to measure the estimation error and only under all those conditions one can derive a valid ENS formula. The host of assumptions leads to the conclusion that one has to argue for an appropriate ENS formula based on the situation at hand, i.e. one has to go through a seperate argument for every estimator and weighting scheme. In addition, the results in this paper can be used in order to generalise the concept of convergence rates and consistency to weighted estimators. This paper also underlines the need to consider the role of data quantity definitions for the assessment of statistical confidence: In times of 'big data' one can easily be fooled into believing that one has access to an abundance of data and that, consequently, convergence errors are negligible. While it is true that there currently is a flood of data that can be used for analysis it has to be acknowledged that economic time series are typically highly correlated, codependent and that their informational value changes over time and it is of paramount interest to assess the role of those interconnections to the notion of data quantity and the corresponding results drawn from statistical analysis. To this end it would be highly desirable to approach a relaxation of the independence assumption in the argument in this paper. The last subsection shows that a wrong measure of data quantity can lead to an unjustified sense of security.

Appendix A

Suppose X is a normally distributed random variable with i.i.d. realisations $X_1, ..., X_n$, known mean $\mu = 0$ and unknown variance σ^2. Consider the statistics

$$S = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \tag{65}$$

Then S is unbiased, since $E[S] = \sigma^2$.

The variance of S is given by

$$\begin{aligned}
Var(S) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) && (66)\\
&= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i^2\right) && (67)\\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i^2) && (68)\\
&= \frac{1}{n} Var(X_1) && (69)\\
&= \frac{1}{n} E\left[E\left[X_1^4\right] - E\left[X_1^2\right]^2\right] && (70)\\
&= \frac{1}{n}\left[3(\sigma^2)^2 - (\sigma^2)^2\right] && (71)\\
&= \frac{2(\sigma^2)^2}{n}\ . && (72)
\end{aligned}$$

The steps follow from the i.i.d. assumption and the centralised moment formula for normal distributions, specifically $E\left[X^4\right] = 3\sigma^4$ (see Shao (1998)).

The Fisher Information (see Shao (1998)) of a sample of n observations is given by

$$\begin{aligned}
I(n, \sigma^2) &= n \cdot \left(-E\left[\frac{\partial}{\partial \sigma^2}\left(\frac{\partial}{\partial \sigma^2}\log\left[\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{X^2}{2\sigma^2}}\right]\right)\right]\right) && (73)\\
&= n \cdot \left(-E\left[\frac{\partial}{\partial \sigma^2}\left(\frac{X^2}{2(\sigma^2)^2} - \frac{1}{2(\sigma^2)}\right)\right]\right) && (74)\\
&= n \cdot \left(-E\left[-\frac{X^2}{(\sigma^2)^3} + \frac{1}{2(\sigma^2)^2}\right]\right) && (75)\\
&= \frac{n}{2(\sigma^2)^2}\ . && (76)
\end{aligned}$$

Since eq. (72) is the inverse of eq. (76) the Cramer-Rao lower bound which states that

$$Var(S) \geq \frac{1}{I(n, \sigma^2)} \tag{77}$$

is attained and, as a consequence, the estimator is efficient.

Appendix B

This paper serves different purposes regarding its role in this thesis and as a contribution to the literature: Within this thesis the main objective is to generalise the methods of a optimal data set choice laid out in section 2. However, this is not a question that is currently discussed in the literature in which data quantity for weighted observations plays a role. In this strand of literature data quantity is used in order to rationalise the size of confidence intervals and statistical significance. Since the paper has to be marketable it has to be imbedded into literature on the topic that is currently citable. As a consequence I wrote the paper tailored to the literature on risk management and only allude to its role within the scope of this thesis.

The objective of this appendix is to perform an analysis similar to the one carried out in section 2, i.e. to apply the results derived here in order to approach the problem of an optimal data quantity or, in this case, an optimal weighting scheme. Since the only analytic formulas have been derived for the case of a data decay parameter, i.e. exponential weighting, I restrict myself to this case.

Analysis

Given the functional form of the combined error for the window case one has to express all terms of the combined error equation (eq. (36)) in terms of the data decay parameter, i.e. one has to express convergence errors as well as representativeness errors in terms of the data decay parameter 'α'. In order to have a connection between stochastic novelty and representativeness I again assume a monotone dynamics of σ_t over time (see the argument in section 2). The convergence error is now called ξ_σ instead of α, since α is reserved for the data decay parameter. β_σ is again the bias or representativeness error.

Representativeness Error

As before the representativeness error in this setup is defined as

$$\beta_\sigma(\alpha, t, T) = \mid \frac{1}{T-t}\int_t^T \alpha^{T-s}\sigma_s^2 ds - \sigma_T^2 \mid \ . \tag{78}$$

Given the assumption of a monotone dynamics of σ_t it can be concluded that $\beta_\sigma(\alpha, t, T)$ is increasing in α. An α of 0 discards all information from the past, i.e. only relies on current information and consequently, leads to a representativeness error of 0. An α of 1 corresponds to an equal weighting of past observations, i.e. one uses all information from the interval $[t, T]$ in an equal manner and, consequently, has the highest representativeness error.

Convergence Error

In general the convergence error, which is now called ξ instead of α, is given by

$$\xi_\sigma(\alpha, t, T) = E\left[\left(\hat{\sigma}^2_{\alpha,t,T} - \sigma^2_{\alpha,t,T}\right)^2\right] \tag{79}$$

$$= E\left[\left(\frac{1-\alpha}{1-\alpha^T}\sum_{t=1}^{T}\alpha^{T-t}\sigma_t^2 - \sigma^2_{t,T,\alpha}\right)^2\right] \ , \tag{80}$$

where $\sigma_{\alpha,t,T}$ denotes the true value of the mixture distribution induced by α on the data set spanning the time frame from t to T. Unfortunately, the combination of a sum containing an exponential term within a squared term is mathematically unapproachable to the author. However, given the behavior of convergence errors for i.-non-i.d. mixture distributions derived in part 2 one can identify the following expression as a candidate for the solution :

$$\xi_\sigma(\alpha, t, T) = 2\frac{\sigma^4_{\alpha,t,T}}{ENS(\alpha, t, T)} \ . \tag{81}$$

Using the ENS definition derived in part 3 this candidate, in turn, can be checked via simulations.

Error Combination and Simulation

Given the functional form of the individual errors one can combine the errors analogously to the procedure laid out in section 2. The corresponding function of the combined error

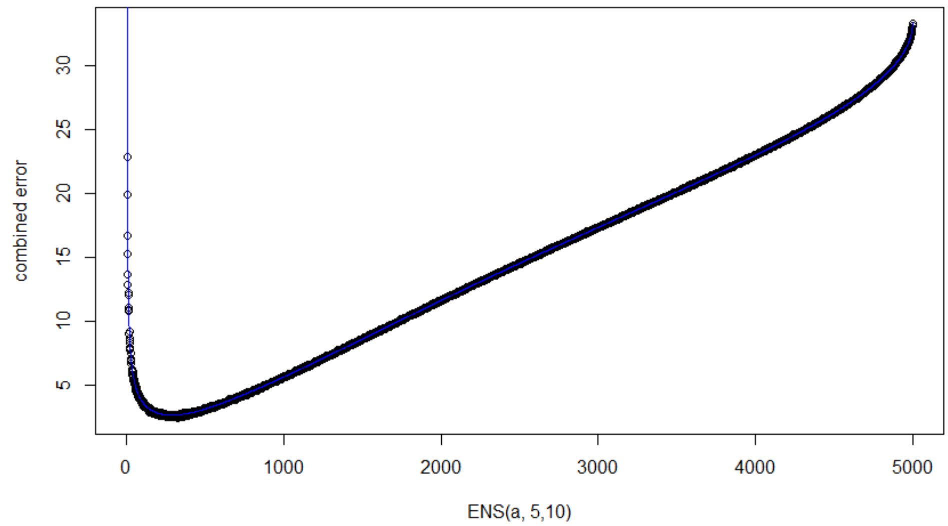

Figure 7: **Combined Error using exponential weights**: Depicted is the empirical combined error (black dots) against the theoretical prediction (blue solid line) as a function of the implied effective number of scenarios $ENS(\alpha, 5, 10)$.

can be checked via simulation. Here, I rely on the same setup as in section 2, i.e. a data set whose generating characteristics is given by $\sigma(t) = t, 5 \leq t \leq 10$ and where my goal is to infer the overall error $\gamma_{\alpha,5,x}$ as a function of x. The results are depicted in Figure 7.

Remark

The minimal error associated to the optimal data window in section 2 or the optimal data decay parameter derived in section 3 are not necessarily equal, i.e. for a data window the optimal data quantity could be 250 while for the same situation the optimal data decay paramter could have an implied ENS of less or more than 250. As a rule of thumb one can expect that the rolling window method will yield lower optimal errors in the situation of rare, severe structural breaks since the estimator will be unbiased for post structural break data while an exponential weighting scheme that utilises the whole history of the process will immediately, i.e. for any $\alpha > 0$, be biased since it uses pre-structural break data. Say, one has a sample of 500 data points with a structural break at 250. Then the last 250 data points will be homogeneous and the window estimator will be unbiased for that data set while the weighted estimator with an ENS of 250 will have a significant bias due to its reliance on data from before the structural break.

4 Introducing Stylized Facts of Electricity Futures Through a Market Impact Model

The results of this part of the thesis have been published as Krause (2020).

Abstract

The objective of this paper is to provide an alternative way to introduce some of the more prominent stylized facts of electricity futures. For non-storable commodities forward looking information is not necessarily incorporated in the price history. In contrast to contemporary electricity finance models I introduce a mechanic that is based on the trading behavior of market participants and their corresponding market impact by exploiting characteristic initial positions and quantity risk considerations. For large time to maturities I end up with a market influenced by hedging pressure whereas for small time to maturities quantity risk comes into play and yields an increased volatility. In addition, prices can also be negative. The model is accompanied by an empirical analysis that shows that parameters which are typically only relevant on small time scales have a significant dynamics in daily returns over the period of years which allows access to the tool box usually reserved for higher frequencies.

4.1 Introduction

The debate over the origin of futures risk premia has been intense since Keynes connected a downward sloping shape of the futures curve (backwardation) with the desire of producers to hedge their production (see Keynes (1930)). The counterposition is taken by a long list of authors starting with Working (1949) who relate the risk premium to issues of storage and inventories rather than hedging pressure and risk transfer. This position mostly implies an upward sloping shape of the futures curve (Contango).

The underlying assumption in mathematical finance models is that one has access to the underlying and that it is possible to store the underlying in some form. Consequently, those models are built in the spirit of the theory by Working (1949). In some commodity markets, like precious metals or agricultural products, this assumption is mostly fulfilled while in others it is not and it is dubious to use classical spot models in the description of futures markets on non-storable commodities. In fact, it has been noted that it is *fundamentally wrong* to do so (see Benth and Meyer-Brandis (2009)). However, it is still very prominent to use spot price models in, e.g., electricity markets since they are able to produce the market specific stylized facts one would like to model (see Benth *et al.* (2008a)). The objective of this paper is to introduce those stylized facts in an alternative way by exploiting nonstorability and the symptoms of the related problem of continuous supply and demand matching.

Electricty Futures exhibit a variety of characteristic stylized facts: Prices can be negative, they can experience a significant drift and their risk profile and volatility change over their life span (see Benth *et al.* (2008a)). So far, those features have been modeled utilising either Long-Term/Short-Term spot models (see Schwartz and Smith (2000)) or Heath-Jarrow Morton approaches (see Heath *et al.* (1992)). However, the Heath-Jarrow Morton approach has a low explanatory power in electricity markets (see Koekebakker and Ollmar (2005)) and for spot price models the critique mentioned above applies.

In order to proceed I have to make sure that the qualitative assumptions of the Keynes (1930) theory are fufilled, i.e. I have to argue *for* limited liquidity, i.e. against market efficiency. Qualitatively one can say that the documented stylized fact of hedging pressure in electricity markets (see, e.g., Benth *et al.* (2008b) among others) is not compatible with market efficiency and, more specifically, requires limited liquidity: In order for the hedging pressure argument to hold one needs to assume that market participants are not able to achieve their desired position size in arbitrarily short time frames, i.e. hedging pressure only works in a world where liquidity and market efficiency are limited and the visible asset prices are the outcome of an optimization between the hedging premium (market impact) and portfolio risk considerations, i.e. optimal liquidation (see the seminal paper by Almgren and Chriss (2001)). This qualitative argument leads to the hypothesis that instead of borrowing the models from other commodity and financial markets it might be helpful to look at the electricity market through the lense of high frequency finance, i.e. to check whether the parameters that are typically discussed in this field of literature are present in electricity futures markets and whether the returns of futures can be related to the position size. To check this hypothesis empirically is the first objective of this paper and is investigated in Section 2. Subsequently, a model is set up that is able to capture the known stylized facts of electricity futures utilising market specific information and non-storability. In addition, a strategy is provided how this model can be used to infer aspects of the trading behavior of market participants. In section 4 I conclude.

4.2 Liquidity and Daily Returns

This section has the objective of providing evidence for an analogy between intraday block trading and electricity portfolio liquidation over the time frame of several years. I relate return anomalies of a buy and hold trading strategy to the time dynamics of liquidity metrics. In principle electricity producing companies are faced with the following problem: For a date in the far future the electricity price risk is concentrated entirely in the hands of very few energy producing companies through their production facilities. Consequently, they have enormous positions and should have an incentive to reduce their exposure by giving out discounts. It is an intricate problem to find a balance between position reduction and market impact but in the market microstructure literature there are models that deal with this type of problem (see, e.g., the seminal paper by Almgren and Chriss (2001)). For large times to maturity it can be expected that this position reduction influences the prices more severely than for short time to maturities since the positions are bigger and the associated price risk is more concentrated on the sell-side. Since the energy producing companies own the large positions they want to find buyers. Consequently, it can be expected that the prices are lower for large time to maturities. One point that makes the problem more intricate in the context of electricity markets is also that one encounters severe quantity risk (see Pérez-González and Yun (2013)), i.e. that the amount that is liquidated is constantly changing and one has to trade towards a moving target since the market has to be cleared at every point in time and electricity demand is fluctuating. This observation implies that liquidity concerns play again a role also for short time to maturities. Since limited liquidity is typically related to an increasing risk it can be expected that the effect on the prices is again negative and one has to compensate the counterparty in order to trade which implies lower prices.

4.2.1 Data Description

The data set used in this study consists of price and volume data from the Norwegian Energy Exchange 'NASDAQ OMX Commodities' and its predecessor 'Nord Pool'. The futures contracts are financially settled quarterly contracts stretching from 2004 to 2013. To be precise, the handles of the contracts are ENOQX-YY , where $X \in \{1, 2, 3, 4\}$ and $YY \in \{06, 07, 08, 09, 10, 11, 12, 13\}$. The data set also includes yearly contracts ENOYR-YY, where YY refers to the years from 2006 to 2013 and FWYR-YY where YY refers to the years 1999 to 2005. Throughout this time period NASDAQ OMX and its predecessor Nord Pool repeatedly changed the product specifications of the contracts. Most notably contracts with maturity prior to 2006 are quoted in NOK/MWh while the remaining contracts are quoted in Euro/MWh. I do not use exchange rates to convert prices prior to 2006 to EUR due to the nature of my research question which is concerned with the behavior of single contracts over their lifespan (cf. Lau (2015b)). For the quarterly contracts there are between 500 (2 trading years) and 700 (2 trading years and 9 trading months) data points available per individual contract while for the yearly contracts individual time series can have up to 1256 data points (5 trading years). Differences in the length of the quarterly time series stem from the contract specificifation rules applied by Nasdaq (see (Benth *et al.*, 2008a)): At the beginning of each calender year, say Jan 1 2020, the quarterly contracts of the whole 'cal + 2', i.e. 2022, become tradable. This implies that the Q1 contracts are tradable for 2 years (with approximately 250 trading days per year) while Q2 contracts are traded for 2 years and a quarter and so on. In the analysis below only the contracts up to 'YY' = 12 are used since the time series for the 2013 contracts are incomplete. Volume and bid-ask spread data is only available for the EN contracts, i.e. for the contracts with delivery beginning in 2006, i.e. 7 yearly contracts and 28 quarterly contracts.

Throughout the whole of the paper I adopt the language from the electricity literature and speak about 'futures' instead of 'swaps' and use 'time to maturity' instead of 'time to delivery' (see (Benth *et al.*, 2008a)) .

4.2.2 Returns of Liquidity Exploitation Strategies

The simple strategy below is based on the following two intuitions. First, Benth *et al.* (2008b) show that their notion of 'market power' changes over the life span of the contracts and document first order price effects. Thus, if one is always on the right side of market power, one should be able to earn a return. Since the paper was published in 2008 it is interesting to test whether the effect has been robust to revelation which is known to be an issue in financial markets (see the seminal paper on the 'wandering weekday effect' by Doyle and Chen (2009)). In addition, since the authors utilise data from Germany between 2002 and 2006, a time frame that included a significant change in legislative circumstances targeting the phase out of nuclear power, testing the hypothesis on a different data set is of importance. I find that the effect is persistent in the sense that there are profitable liquidity exploitation strategies. This is an indication that this finding is not a 'statistical arbitrage' but serves an economic purpose that is prevalent on electricity markets in general.

The second intuition is the asymmetry in the initial distribution of market participants layed out in the introduction. A party liquidating a large block of assets has to optimize between their market impact and the risk on their book. This typically involves a high market impact at the start of the liquidation period and therefore it would be beneficial to buy the contracts at inception and earn a premium for liquidity provision which is in line with the hedging pressure arguments by Benth *et al.* (2008b).

Consequently, a simple strategy exploiting those intuitions is the following:

i) buy the contract at inception

ii) sell the contract at *some* time to maturity (2 times)

iii) buy the electricity in the spot market

The mean profitability of this strategy is depicted in Figure 7 (left column) depending on the time to maturity where the position is liquidated and switched against a short position. Only the quarterly contracts are shown. Due to cascading effects the pattern for yearly contracts looks similar. As mentioned before the findings are consistent with the intuitions. It is, however, very interesting to observe that the contracts for the different quarters show very similar behavior that is not connected to a time to maturity dependent event but rather connected to a date, namely the first of January in the year prior to maturity which is highlighted by a vertical line. Note that the time to maturity of the first of January is different for the different classes of quarterly contracts: For the Q1-contracts it is, typically and in line with a variety of implied option pricing models, at 252 trading days. For the respective other quarters the time to maturities are at 252 + 0.25*252 = 315 for the Q2 contracts, 252 + 0.5*252 = 378 for the Q3 contracts and 252+0.75*252 = 441 for the Q4 contracts. When one turns to possible explanatory variables one finds that there is a similar delayed shape in a variety of (il)liquidity metrics. Liquidity is an elusive concept and in the following I use a variety of different liquidity metrics. While it is clear what is meant by trading volume and bid - ask spreads an additional measure of liquidity that I will employ in the following is the Amihud illiquidity metric (see Amihud (2002)). In its simplest form it is defined as

$$Amihud_t = \frac{|r_t|}{Vol_t} \; , \tag{82}$$

where r_t denotes the return at time t and Vol_t the corresponding trading volume. Consequently, 'liquid markets' are markets in which the absolute value of returns are low and trading volume is high while illiquid markets are characterised by either high absolute returns or low trading volumes. It is a prominent measure of liquidity that can be easily calculated from data that is readily available for most markets for long time frames and has been used extensively in the literature on market microstructure (see , e.g. Foucault *et al.* (2013) and the references therein). I take the average 5 day amihud illiquidity, which is a common adaption that has been proposed in the original study by Amihud, and I also use the log of the trading volume instead of the absolute number since in electricity markets the trading volume changes several orders of magnitude throughout the life span of the contracts.

Given the liquidity metrics one can analyse their dynamics and relate them to the returns of the liquidity exploitation strategy. In Figure 8 one can see that the leveling out of the Amihud illiquidity metric falls together with the peak of the return of the strategy in the left column. In Figure 9 it can be seen that average and median autocorrelation continuously changes in level throughout the life span of the contracts. For large time to maturities autocorrelation is negative which is consistent with market microstructure literature (see, e.g., Zhang *et al.* (2005)) and for short-time to maturities it is slightly positive. Since positive autocorrelation is not compatible with the stylized facts of asset returns (see Cont (2001)) this observation already indicates that liquidity might play a role for low time to maturities as well. However, the quintessential point is that the autocorrelation is changing over time and since autocorrelation is one of the simple indicators of liquidity concerns it can be expected that the degree of liquidity in the market is also changing over time.

Those observations point to the assessment that electricity futures do not only exhibit time to maturity dependent behavior (a feature that can be introduced by means of a short-term component in spot price models as in Schwartz and Smith (2000) and its generalisations) but also show aspects of seasonal behavior, i.e. informative signals that are located around certain dates with corresponding similar behavior in liquidity metrics. This is also a point not sufficiently addressed by current electricity finance models.

4.2.3 Test Design

In this section I am interested in the qualitative statement whether liquidity metrics show a significant directional dynamics throughout the life spans of the contracts. In order to detect non-stationary behavior one needs to have a time series of the parameters in question. For returns and trading volume the original time series can be used. However, the centralised moments and liquidity measures have to be estimated. A rolling window of 50 subsequent data points is used for this purpose.[5]

In a next step it is tested whether the time series of the parameter in question reveals a drift by employing a (modified) Kendall-Mann test. This test is a drift test which can be modified to achieve robustness with respect to the existence of autocorrelation (see Hamed and Rao (1998)). In a last step the number of contracts who show a significant drift for the parameter in question is counted and it is tested whether the distribution of positive and negative drifts is symmetric by using a binomial test with parameter $p = 0.5$. The intuition behind this last step is that if there is no time-to-maturity dependent behavior then the

[5]The results are stable with respect to the length of the rolling window. 75 and 100 data points yield similar results (not shown).

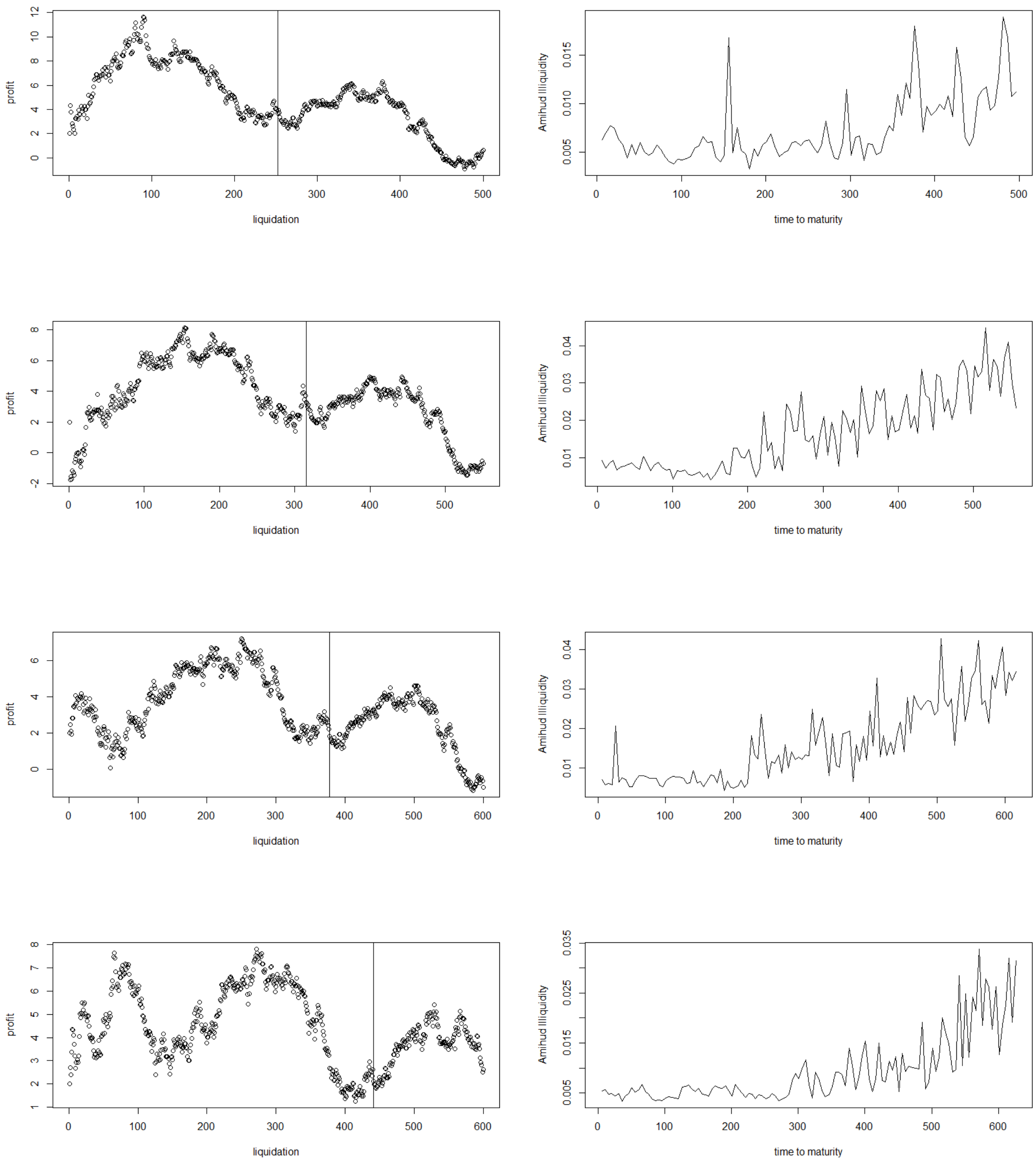

Figure 8: **Time to delivery dependent amihud illiquidity ratio and profitability of the buy and hold strategy for the Q1 contracts (first row), Q2 contracts (second row), Q3 contracts (third row), and Q4 contracts (fourth row)**

number of paths that show a positive drift (as determined by Mann-Kendall) should be roughly equal to the number of negative paths. Anything else indicates that parameters are more likely to grow or decline as a function of time to maturity. To sum up, the steps are the following:

i) Calculate time to maturity dependent time series for the first four centralised moments and the autocorrelation for every contract based on a 50 data point rolling window.

ii) Apply a modified Mann-Kendall test in order to detect possible drifts in those time series. In table 1 #contracts +,-,0 denote the number of contracts that exhibit a significant (positive, negative) or insignificant drift, respectively. Here signs are with respect to an increasing time to maturity, i.e., e.g., variance is decreasing with increasing time to maturity.

iii) Count the number of positive and negative drifts and test whether it is reasonable to assume that they are $B_{0.5}$-distributed. The p-value of this test is shown in the last column of table 1.

The following table contains the results for this test procedure applied on the return time series for the electricity forward data introduced before. [6] [7]

[6]For returns I use classical percent changes of the futures price. If one interprets futures returns as 'return on investment' this implies that I assume a 100 percent margin requirement (cf Lau (2015b)).

[7]For the analysis all quarterly contracts from 2006-2012, i.e. overall 7*4 = 28 contracts are used. The contracts have between 501 (Q1 2008) and 692 (Q4 2011) data points). For the yearly contracts, 'FWYR-$\{99, 00, 01, 02\}$ have between around 300 and 700 data points while 'FWYR-$\{03, 04, 05\}$' and 'ENOYR-$\{06, 07, 08, 09\}$' were tradable for 3 years (i.e. they have around 750 data points each). The time series of the remaining contracts are longer and top out at 1256 data points (5 trading years) for ENOYR-12 due to several changes in contract specifications (cf Lau (2015b)). The length of the available time series is increasing over time due to increasing liquidity in the market and demand for longer running contracts, e.g. in order to hedge financing exposure. Currently, contracts at NASDAQ are tradable up to 10 years in advance. All data points for all the contracts mentioned in this footnote are used.

Table 1: Drift Test Results

Parameter	Contracts	#contracts +	#contracts -	#contracts 0	pvalue
Mean	quarterly	18	8	2	0.076
	yearly	3	7	3	0.344
Variance	quarterly	1	26	1	< 0.001
	yearly	1	11	1	0.031
Skewness	quarterly	7	20	1	0.019
	yearly	3	7	3	0.344
Kurtosis	quarterly	19	4	5	0.003
	yearly	9	3	1	0.146
Autocorrelation	quarterly	8	18	2	0.076
	yearly	7	3	3	0.344
trading volume	quarterly	0	28	0	< 0.001
	yearly	0	7	0	0.016
bid-ask spread	quarterly	28	0	0	< 0.001
	yearly	7	0	0	0.016
Amihud illiquidity	quarterly	28	0	0	< 0.001
	yearly	7	0	0	0.016

Results

From the data compiled above the following conclusions can be drawn

i) Except for the liquidity metrics and variance yearly contracts do not show significant results. Thus, the takeaways in the following focus on results for the quarterly contracts.

ii) A decreasing autocorrelation with increasing time to maturity.

iii) Although the signal is not particularly significant, one finds a drift in the mean of the returns of quarterly futures. The mean return becomes more positive for higher time to maturities. This result is compatible with the idea that for higher time to maturities the market for electricity is a buyers market due to hedging pressure. Note that the information set of the quarterly and the yearly contracts is highly different due to the different time frames (the data series for the yearly contracts starts in 1996 while for the quarterly it starts in 2004).

iv) One finds that the variance of returns (volatility) is decreasing with increasing time to maturity, i.e. the method is able to detect the so-called Samuelson Effect.

v) A decreasing skewness with increasing time to maturity.

vi) An increasing kurtosis with increasing time to maturity. This result is particular interesting since kurtosis can be seen as a measure for extreme events (see (Westfall, 2014)).

vii) All the (il)liquidity metrics point to a gradual worsening of the liquidity situation for higher maturities. However, this assessment is not yet plausible for the autocorrelation since only negative autocorrelations can be associated with a bad state of liquidity. In order to support the claim one has to check whether the autocorrelation is negative for high time to maturities. This is indeed the case: (Lau, 2015b) reports a negative first autocorrelation coefficient for the same data set. The result above points out that autocorrelation becomes more negative for higher time to maturities. Therefore it can be concluded that autocorrelation is negative for higher time to maturities. See also Figure 9.

All the reported results are compatible with the assertion that electricity futures exhibit severe non-stationary features especially for the quarterly contracts where even the returns show an informative change of behavior. So far the qualitative assessment of this fact has been at the heart of the considerations. In a next step characteristic returns of a buy and hold strategy are presented. Note that the version of the Amihud illiquidity metrics used here is based on the traded log-volume since the volume grows by several orders of magnitudes throughout the life span of the contracts, i.e. the results above would be more drastic when using the unmodified volume data.

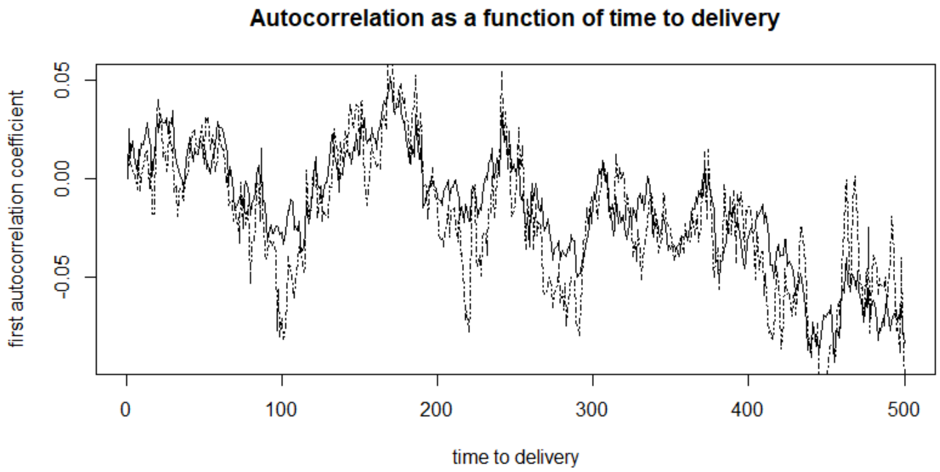

Figure 9: **Mean (solid line) and Median (dashed line) first order autocorrelation coefficient**

4.2.4 Price Dynamics and Cumulative Trading Volume

In the presence of hedging pressure and imbalanced initial endowments it is reasonable to assume that the cumulative trading volume since inception is a price-relevant quantity. The objective of this section is to show that this intuition is justified. However, since the market consists not only of producers and consumer but also contains liquidity suppliers it is reasonable to assume that trading behavior at inception is only from producers to either consumers or liquidity suppliers who both benefit from the risk management considerations of producers. Therefore, it can be expected that earlier trading volume is highly informative for the returns of the hedging pressure strategy while later trading can also be between liquidity suppliers and consumers. As such, it is reasonable to assume that earlier trading volume is more relevant for the hedging pressure argument than later trading volume. Hence, I use a concave transformation $f(\cdot)$ of cumulative trading volume in the regression, i.e. $f(\int_0^t V_s ds)$, where $s = 0$ corresponds to the inception date of the contract and V_s is the trading volume at time s. One can now regress the levels of the future prices on cumulative trading colume while controlling for time-to-maturity dependent behavior. The results are depicted in table 2 and support the assertion that cumulative trading volume is a price-relevant quantity for the average of all the contracts in the four quarters.

Table 2: Regression Results for the Quarterly Contracts.

Contracts	intercept	time to maturity	cumulative log trading volume
ENOQ1-(07-13)	37.367150(***)	$-2.515 \cdot 10^{-3}$(*)	1.296177(***)
ENOQ2-(07-13)	21.557004(***)	$6.455 \cdot 10^{-3}$(***)	1.895912(***)
ENOQ3-(07-13)	22.90(***)	$5.419 \cdot 10^{-3}$(***)	1.615(***)
ENOQ4-(07-13)	29.16(***)	$4.336 \cdot 10^{-3}$(***)	1.541(***)

4.2.5 Liquidity for small time to maturities

The conceptual difference to the setup one usually encounters in a market microstructure environment comes into play for short time to maturities: In equity or foreign exchange markets the number of assets one wants to sell or buy is typically known whereas for electricity markets the demand is random and expected future production is a stochastic process changing over time. With the data set used above it is not possible to isolate the corresponding signal since it would involve building a model for expected future production and a corresponding liquidation strategy. However, it has to be noted that liquidity plays a role for large and small time to maturities and the model below introduces both effects with only one tool.

4.3 A limited liquidity model for stylized facts of electricity futures

The objective of this section is to set up a mathematical model that is able to capture the results of the empirical analysis and to provide a strategy how this model can be used to answer questions on trading behavior of market participants.

4.3.1 A simple liquidity based model

I assume that the price of a future with maturity at T, denoted by F_t^T, is given by

$$F_t^T = S_t^T + L_t^T \ , \tag{83}$$

where S_t^T denotes a current expectation of the future spot price under a risk neutral measure, i.e.

$$S_t^T = E^Q \left[S(T) \mid I^T(t) \right] \ . \tag{84}$$

For a non-storable commodity like electricity it is *fundamentally wrong* to use only the information filtration generated by the asset (Benth and Meyer-Brandis, 2009). One way to circumvent this problem is to model the information set on a per contract basis, (here expressed by the superscript T in I^T), i.e. I assume that every future has its own price determination process. Since this paper is solely concerned with the effects of liquidity on the qualitative properties of the futures price I focus my attention solely on the latter part, i.e. L_T^t and the spot price component will be ignored in the following. However, it can be noted that it is very elegant to think of the pricing equation (83) as a representation dividing the futures price in a component that can be analysed using classical mathematical finance techniques (i.e. concepts that are based on the notion of storability) in the spot price (e.g. through structural models and hedging through the fuel markets Aïd *et al.* (2009)) and a distortion that captures the effects that stem from the non-storable nature of electricity.

Non-storability implies that the energy that is produced at one point in time has to be consumed at the same time. Since electricity demand is random this gives rise to a challenging trading problem where one has to trade towards a moving target in a notoriously illiquid environment. The objective of this paper is to use this problemspecific information to set up a model that produces the stylized facts of electricity futures.

In the following I take a hedging pressure perspective where the position imbalance of the market participants and their corresponding trading activity in order to achieve goals for risk management or supply-demand balancing has a direct influence on the price. I assume that the price of the liquidity component can be modelled as

$$L_t^T := -\frac{\alpha}{T-t} P_t^T, \tag{85}$$

where P_t^T is the *difference to the desired position size* of electricity producing companies or, alternatively, the difference to the equilibrium of the market. In this context α can be interpreted as an illiquidity parameter since $\alpha = 0$ implies that $L_t^T = 0$ and that the futures price is not influenced by any market impact.

In order to proceed it is necessary to specify the process P_t^T. I assume that this process is a brownian process with an initial condition $P_0^T = \eta$ (the total amount of the current estimate of future demand at the start of trading). In addition, I assume that the market has to clear at maturity, i.e. that the difference to the desired position is zero, since supply has to match demand due to non-storability, i.e. $P_T^T = 0$. Augmenting this information into a brownian motion leads to a brownian bridge dynamics, i.e.

$$dP_t^T = \frac{-P_t^T}{T-t}dt + dB_t^T \ . \tag{86}$$

The solution of this process can be derived by a standard variations of parameters argument and is given by

$$P_t^T = \frac{T-t}{T}P_0^T + \int_0^t \frac{T-t}{T-s}dB_s^T . \tag{87}$$

Note that this is an implicit assumption on the trading behavior of the market participants, i.e. the drift in this process can be interpreted as the solution to some optimal liquidation problem that balances between risk management needs and market impact. Consequently, it is also an assumption on limited liquidity since in a perfectly liquid market the optimal liquidation speed is infinite and there is no tradeoff between market impact and risk management considerations (see, e.g., Almgren and Chriss (2001)).

For the market impact component one obtains

$$L_t^T := -\frac{\alpha}{T-t}P_t^T = -\frac{\alpha}{T-t}\left(\frac{T-t}{T}P_0^T + \int_0^t \frac{T-t}{T-s}dB_s^T\right) \tag{88}$$

$$= -\frac{\alpha}{T}P_0^T - \int_0^t \frac{\alpha}{T-s}dB_s^T \ . \tag{89}$$

The process L_t^T displays all the stylized facts that we are interested in, i.e.

i) The stochastic integral is added and consequently the sign of L_t^T is not determined which implies that the overall prices can be negative since the futures price was modelled by adding up a spot price process and the liquidity premium.

ii) The volatility of the process is increasing througout the life span due to the functional form of the integrand in the stochastic integral which exhibits a singularity at T.

iii) The process L_t^T exhibits a drift stemming from the initial condition. As such this process captures the phenomenon of hedging pressure.

This model can easily be simulated (see Figure 10). As before the trading behavior is defined by the superposition of two forces: updates of the expected future production, i.e. dB_t^T and the trading intensity towards the target, i.e. $-\frac{Y_t^T - Z_t^T}{T-t}$, where Y_t represents the current position while Z_t^T represents the desired position, i.e. expected future production. In this instance

$$Z_t^T = \eta + \int_0^t dB_s^T.$$

Since only the difference to the desired position size is modeled here, i.e.

$$d(Y_t^T - Z_t^T) := -\frac{Y_t^T - Z_t^T}{T-t}dt + dB_t^T$$

it is sufficient to specify only the initial value of $Y_0^T = 0$ in order to have a determined dynamics. With $P_t^T := Y_t^T - Z_t^T$ this observation yields the equations above. This observation implies that the only necessary process that has to be simulated in order to know the dynamics of L_t^T is the brownian motion representing expected future production. The upper panel shows the expected future production (solid line) and the corresponding filled position $Z_t^T = Y_t^T - P_t^T$ (dashed line) with an initial expectation of $\eta = 100$. The middle panel shows the corresponding market impact and the bottom panel shows the price of a future assuming a driftless geometric brownian motion that is independent of B^T for the spot model. For the seed of the simulation I picked the first seed that lead to a negative futures price close to maturity in order to demonstrate that this feature can be achieved.

4.3.2 Empirical Strategy

While the primary objective of this paper is to introduce stylized facts, i.e. qualitative aspects, of electricity futures the model introduced here can be used for empirical, i.e. quantitative, considerations as well: Starting from the setup, i.e. $F_t^T = S_t^T + L_t^T$, one can specify appropriate models for all the individual parts, i.e. a spot price model for S_t^T, a futures model for F_t^T and a model for the open position and the corresponding trading behavior impact, L_t^T. In this case the specific model parameters of F_t^T and S_t^T can be estimated from individual time series data of spot and futures prices. After subtraction

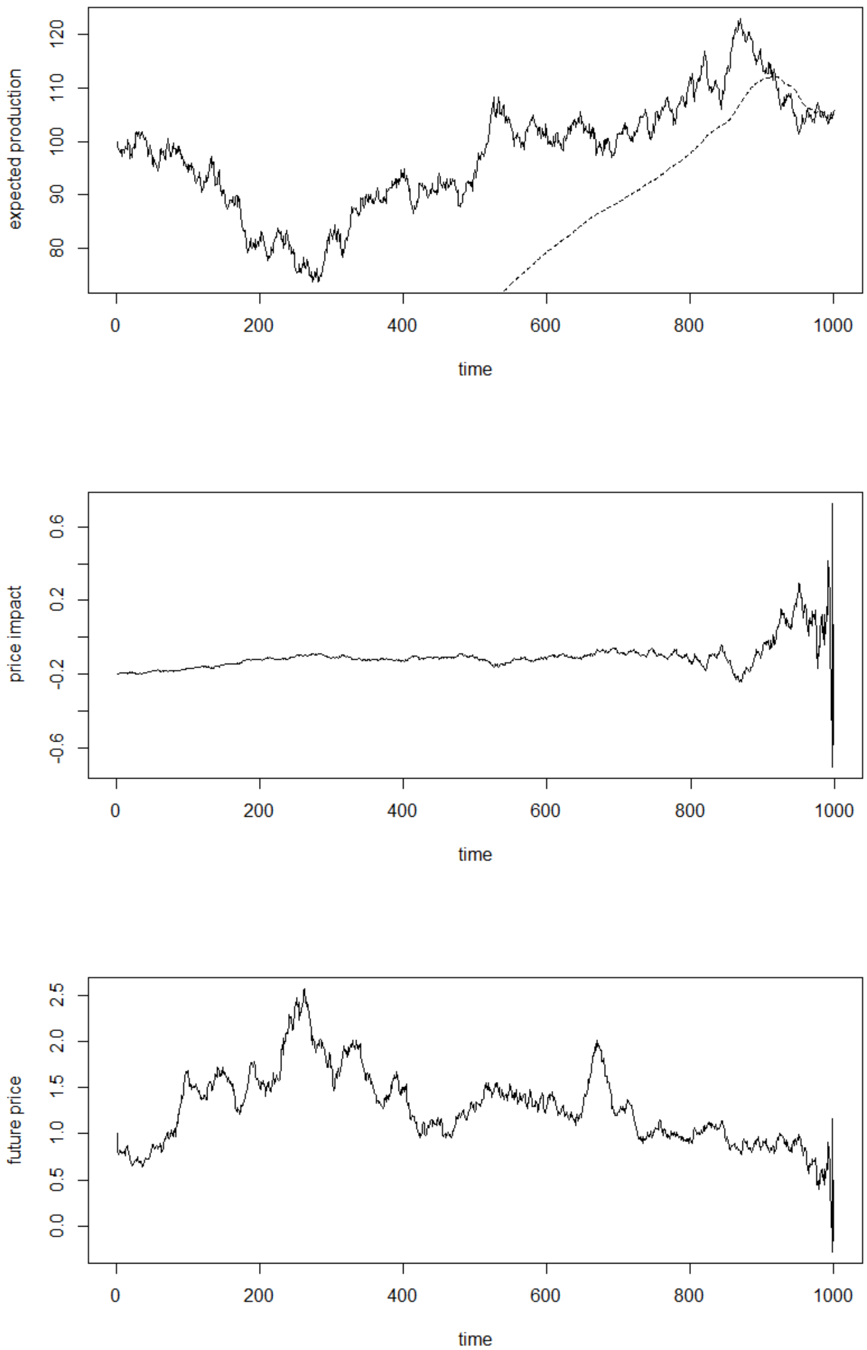

Figure 10: Simulation of the market impact model.

one is in a situation where one has access to the process L_t^T and can estimate its properties. Since L_t^T is a combination of trading behavior and the difference to the desired position size, knowledge on either can be used to infer knowledge on the other, i.e. if one would be able to correctly model future supply and demand shocks, e.g. through weather forecasts or power plant shutdown schedules, one could use the framework introduced above in order to infer aspects trading behavior or vice versa.

For the qualitative argument presented here, I chose a brownian model for all the components due to its simplicity and because this choice leads to an analytical solution for the dynamics from which one can immediately obtain the stylized facts. Conceptually, however, this argument is not limited to diffusion processes and one can simply use, e.g., a jump-diffusion for the spot price model or - since the spot is not a traded asset and cannot interfere with no-arbitrage principles - even non-semi martingale processes like fractional processes. A similar breadth of choices is available for all the other components. Since this makes the space of possible models very large the corresponding model selection problem is an intricate one and is beyond the scope of the current paper.

4.4 Conclusion

The objective of this paper was to provide an alternative way to introduce some of the stylized facts of electricity futures. The disadvantage of most models used in electricity finance is that they are build on the premise of having access to the underlying which is an assumption that is currently not valid in electricity markets. Since mathematical models have to be judged by the credibility of their assumptions it is useful to provide a mechanic that is able to introduce the stylized facts using market specific information. Due to the interplay between price risk management and quantity risk the liquidation problem in electricity markets is an intricate one and the large trading departments that can be found in electricity producing companies are a testimony to this statement. Consequently, it is useful to have a model that incorporates those market specific properties and translates them into the stylized facts of the products traded in those markets. This paper also uses the natural synergy between models in high frequency finance and electricity markets since both fields use additive models. Further exploration of this synergy potential yields a host of interesting questions. In addition, the mechanic provided here is a natural extension to any spot price model due to its additive structure and the appearance of a liquidity parameter which can be tuned to the properties of the market. In addition, in contrast to the setup in high frequency finance using a model that can yield negative prices is a positive property in the context of electricity markets. Throughout the argument I used an ad-hoc assumption regarding the liquidation speed that market participants use. From a theoretical perspective it would be interesting to find the corresponding optimal liquidation problem that yields the assumed behavior as its solution.

References

Aïd, R., Campi, L., Huu, A.N., Touzi, N. (2009). A structural risk-neutral model of electricity prices. *International Journal of Theoretical and Applied Finance*, **12 (07)**, 926–947. doi: 10.1142/s021902490900552x

Aït-Sahalia,Y., Jacod, J. (2009). Testing for jumps in a discretely observed process. *Ann. Stat.*, **37 (1)**, 184–222. doi:10.1214/07-AOS568/1232115932

Aït-Sahalia,Y., Jacod, J. (2014). *High-Frequency Financial Econometrics*, Princeton University Press

Alexander, C. (2009). *Market Risk Analysis, Value at Risk Models*. Volume 4. John Wiley & Sons

Almgren, R., Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, **3 (2)**, 5–39, doi: 10.21314/jor.2001.041

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, **5 (1)**, 31–56, doi: 10.1016/s1386-4181(01)00024-6

Bailey, D. H., Borwein, J. M., da Prado, M. L., Zhu, Q. J.. (2014) Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance. *Notices of the American Mathematical Society* **61 (05)**

Basel-Committee (2016). *Minimum Capital Requirements for Market Risk*. Consultative Document

Benth, F. E., Meyer-Brandis, T. (2009). The information premium for non-storable commodities. *The Journal of Energy Markets*, **2 (3)**, 111–140, doi: 10.21314/jem.2009.021

Benth, F.E., Benth, J.S., Koekebakker, S.. (2008). *Stochastic Modelling of electricity and related markets*. World Scientific - Advanced Series on Statistical Science & Applied Probability (Vol. 11)

Benth, F. E., Cartea, A., Kiesel, R. (2008). Pricing forward contracts in power markets by the certainty equivalence principle: Explaining the sign of the market risk premium. *Journal of Banking & Finance*, **32 (10)**, 2006–2021, doi: 10.1016/j.jbankfin.2007.12.022

Bernstein, P. L. (1998). *Against the Gods: The Remarkable Story of Risk*. Wiley.

Boltzmann, L. (1884). Über die Eigenschaften monozyklischer und damit verwandter Systeme. *Wissenschaftliche Abhandlungen. III. Band (1882 - 1905).* Ed. by Fritz Hasenöhrl. Ch. 73. 122-152.

Bondi, H.; Gold, T. (1948). The Steady-State Theory of the Expanding Universe. *Monthly Notices of the Royal Astronomical Society*, **108 (3)**, 252–270. doi: 10.1093/mnras/108.3.252

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*. **16 (3)** (Aug., 2001), 199–215, stable URL: http://www.jstor.org/stable/2676681.

Casini, A. and Perron, p. (2019). Structural Breaks in Time Series *Oxford Research Encyclopedia of Economics and Finance*, Oxford University Press.

Christensen, G., Miguel, E. (2018).Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature*, **56 (3)**, 920 – 980, doi: 10.1257/jel.20171350

Christoffersen, P. (2011). *Elements of financial risk management*. Academic Press

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

Danielsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O. Shin, H. S. (2001). An Academic Response to Basel II *Special Paper No. 130, LSE Financial Markets Group*

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, **1 (2)**, 223–236, doi: 10.1088/1469-7688/1/2/304

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1–37.

Dahlhaus, R. (2012). Locally Stationary Processes. *available on arXiv: 1109.4174v2.*

Dahlhaus, R., Giraitis, L. (1998). On the Optimal Segment Length for Parameter Estimates for Locally Stationary Time Series *Journal of Time Series Analysis*, **19 (6)** 629–655, doi: https://doi.org/10.1111/1467-9892.00114, Blackwell

Colander, D., Föllmer, H., Haas, A., Goldberg, M., Juselius, K., Kirman, A., Lux, T., Sloth, B. (2010). The Financial Crisis and the Systemic Failure of Academic Economics. *Discussion Papers - Department of Economics - University of Copenhagen* No.09-03

Danielsson, J., Valenzuela, M., Zer, I. (2018). Learning from History: Volatility and Financial Crises *The Review of Financial Studies*, **31 (7)** 2774–2805, doi: https://doi.org/10.1093/rfs/hhy049

Davidson, P. (2012). Is economics a science? Should economics be rigorous? *Real World Economic Review.* **59**, 58–66

Dirac, P.A.M. (1937). The Cosmological Constants. *Nature*, **139 (323)**, doi: https://doi.org/10.1038/139323a0

Dirac, P.A.M. (1938). A New Basis for Cosmology. *Proceedings of the Royal Society A*, **165 (921)**, doi: https://doi.org/10.1098/rspa.1938.0053

Doyle, J. R., Chen, C. H. (2009). The wandering weekday effect in major stock markets. *Journal of Banking & Finance*, **33 (8)**, 1388–1399, doi: 10.1016/j.jbankfin.2009.02.002

Ellsberg, D.. (1961). Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics*. **75 (4)** : 643–669. doi:10.2307/1884324. JSTOR 1884324.

Foucault, T., Pagano, M., Röell, A. (2013). *Market Liquidity.*, Oxford University Press, doi: 10.1093/acprof:oso/9780199936243.001.0001

Geman, S., Bienenstock, E., Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, **4**, 1–58. doi: 10.1162/neco.1992.4.1.1.

Giacomini, R. and Rossi, B. (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies*, **76 (2)**, 669–705.

Giacomini, R. and Rossi, B. (2015). Forecasting in Nonstationary Environments: What Works and What Doesnt in Reduced-Form and Structural Models. *Annual Review of Economics*, **7**, 207–229.

Gigerenzer, G. (2015). Simply rational: Decision making in the real world. Evolution and Cognition.

Gigerenzer, G., Selten, R. (Eds.). (2002). Bounded rationality: The adaptive toolbox. MIT press.

Gigerenzer, G., Todd, P. M. (1999). Simple heuristics that make us smart. Oxford University Press, USA.

Giraitis, L., G. Kapetanios, and S. Price (2013). Adaptive forecasting in the presence of recent and ongoing structural change. *Journal of Econometrics*, **177 (2)**, 153–170.

Giraitis, L., Kapetanios, G., Yates, T. (2014). Inference on stochastic time-varying coefficient models *Journal of Econometrics*, **179 (1)** 46–65, doi: https://doi.org/10.1016/j.jeconom.2013.10.009

Greene, W (2012). *Econometric Analysis, 7th ed.*, Pearson

Greenfield, P. M. (2017). Cultural Change Over Time: Why Replicability Should Not Be the Gold Standard in Psychological Science.*Perspect Psychol Sci.* 2017 Sep;**12(5)**:762-771. doi: 10.1177/1745691617707314.

Grenier, Y. (1983). Time dependent ARMA modelling of nonstationary signals. *IEEE Trans. Acoust. Speech Signal Process.* **31**, 899–911

Hallin, M. (1986). Nonstationary q-dependent processes and time-varying moving average models: invertibility properties and the forecasting problem. *Adv. Appl. Probab.* **18**, 170–210

Hamed, K. H., Rao, A. R. (2009). A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*, **204 (1-4)**, 182–196, doi: 10.1016/s0022-1694(97)00125-x

Heath, D., Jarrow, R., Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, **60 (1)**, 77–105, doi: 10.2307/2951677

Hommes. C. (2005). Heterogeneous Agent Models in Economics and Finance. *TI 2005-056/1 Tinbergen Institute Discussion Paper*

Inoue, A., L. Jin, and B. Rossi (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, **196 (1)**, 55–67.

Ioannidis, J. P. A.. (2005). Why most published research findings are false. *PLoS Med*, **2 (08)** doi: https://doi.org/10.1371/journal.pmed.0020124

Jacod, J., Shiryaev, A. N.. (2003). *Limit Theorems for Stochastic Processes.* Springer - Grundlehren der mathematischen Wissenschaften.

Jeanblanc, M., Yor, M, Chesney, M. (2009). *Mathematical Methods for Financial Markets.*, Springer Finance Textbooks

Karatzas, I., Shreve, S. (1998). *Brownian Motion and Stochastic Calculus.*, Springer - Graduate Texts in Mathematics

Kayhan, A., El-Jaroudi, A. and Chaparro, L. (1994). Evolutionary periodogram for non-stationary signals. *IEEE Trans. Signal Process.* **42**, 1527–1536.

Keynes, J.M. (1921). *A Treatise on Probability.*, Macmillan & Co.

Keynes, J.M. (1930). *Treatise on money.*, Harcourt, Brace and Co.

Keynes, J. M. (1936). *The general theory of employment, interest, and money.* Palgrave Macmillan

Kirstein, M. (2015). From the Ergodic Hypothesis in Physics to the Ergodic Axiom in Economics. *7. Wintertagung des ICAE Linz.*

Knight, F. H.. (1921). *Risk, Uncertainty, and Profit* Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Co.

Koekebakker, S., Ollmar, F. (2005). Forward curve dynamics in the Nordic electricity market. *Managerial Finance*, **31 (6)**, 73–94, doi: 10.1108/03074350510769703

Krause, J. (2019). Representativeness vs Convergence: Optimal Data Identification in Non-Stationary Systems *Working Paper*.

Krause, J. (2019). A Convergence Speed Dependent Data Quantity Definition and Its Effect on Risk Estimation. *Journal of Asset Management*, (forthcoming), doi: 10.1057/s41260-019-00137-1.

Krause, J. (2020). Introducing Stylized Facts of Electricity Futures Through a Market Impact Model. *Journal of Energy Markets*, **13 (1)**, 31–46, doi:10.21314/JEM.2020.201.

Ku, H. H. (1966). Notes on the use of propagation of error formulas.*Journal of Research of the National Bureau of Standards*, **70 (4)**.

Kuhn, T.S. (1957). *The Copernican Revolution: Planetary Astronomy in the Development of Western Thought*, Harvard University Press

Lau, C. (2015). A simple normal inverse gaussian-type approach to calculate value-at-risk based on realized moments.*Journal of Risk*, **70 (4)**.

Lau, C. (2015). *Non-normality in financial markets and the measurement of risk*. PhD Thesis. MLU Halle-Wittenberg. doi: 10.25673/1585

Lebowitz, J.L., Penrose, O. (1973). Modern ergodic theory. *Physics Today*, **26 (2)**, 23–29, doi: 10.1063/1.3127948

Lindsay, B. G. (1995). Mixture models: theory, geometry and applications. *In NSF-CBMS regional conference series in probability and statistics*, pp. i–163. JSTOR.

Lo, A. (2017). *Adaptive Markets*. Princeton University Press.

Lowenstein, R. (2000). *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. Random House.

Lucas, R. E. (1976). Econometric Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy*, **1**, S. 41

Mallaby, S. (2017). *More Money Than God: Hedge Funds and the Making of a New Elite*. Penguin Press.

Meucci, A.. (2012). Effective Number of Scenarios in Fully Flexible Probabilities. *GARP Risk Professional*, pp. 32-35, February 2012. Available at SSRN: https://ssrn.com/abstract=1971808

Mikosch, T., Stărică, C., (2004). Nonstationarities in financial time series, the long range dependence, and the IGARCH effects. *The Review of Economics & Statistics*, **86 (2)**, 378–390.

Mirowski, P. (2013). *Never Let a Serious Crisis Go to Waste: How Neoliberalism Survived the Financial Meltdown*. Verso

Niman, B. N. (1991). Biological Analogies in Marshall's Work. *Journal of the History of Economic Thought*, **13**, 19–36

North, D.C. (1999). Dealing with a non-ergodic world: institutional Economics, property rights and the global environment. *Duke Environmental Law & Policy Forum*, **X (1)**

North, D.C. (2005). *Understanding the Process of Economic Change*. Princeton University Press.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science *Science*, **349 (6251)**, doi: 10.1126/science.aac4716

Pennacchi, G. G. (2005). Risk-based capital standards, deposit insurance, and procyclicality. *Journal of Financial Intermediation*, **14 (4)**, 432–465, doi: https://doi.org/10.1016/j.jfi.2004.09.001

Pérez-González, F., Yun, H. (2013). Risk Management and Firm Value: Evidence from Weather Derivatives. *The Journal of Finance*, **68 (5)**, 2143–2176, doi: 10.1111/jofi.12061

Pesaran, M. H., Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, **137 (1)**, 134–161.

Phillips, P.C.B., Li, D., Gao, J. (2017). Estimating smooth structural change in cointegration models *Journal of Econometrics*, **196 (1)** 180–195, doi: https://doi.org/10.1016/j.jeconom.2016.09.013

Priestley, M.B.. (1965). Evolutionary Spectra and Non-Stationary Processes. *Journal of the Royal Statistical Society: Series B (Methodological)*. **27 (2)**, 204–229, https://doi.org/10.1111/j.2517-6161.1965.tb01488.x

Protter. P.. (2005). *Stochastic Integration and Differential Equations*. Springer - Stochastic Modelling and Applied Probability

Romer, P.M.. (1990). Endogenous technological change. *Journal of Political Economy*. **98 (5)**, 71–102, doi:https://doi.org/10.1086/261725

Rudin, W. (2013). *Principles of Mathematical Analysis*. Example Product Manufacturer

Samuelson, P.A. (1968). What Classical and Neoclassical Monetary Theory Really Was. *Canadian Journal of Economics*, **1 (1)**, 1–15, JSTOR: 133458

Samuelson, P.A. (1965). Proof that Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, **6 (2)**

Samuelson, P.A. (1974). Challenge to judgement. *Journal of Portfolio Management*, **1 (1)**, 17–19, doi:10.3905/jpm.1974.408496

Schumpeter, J.. (1949). *Capitalism, Socialism and Democracy*. Routledge.

Schwartz, E, Smith, J.E. (2000). Short-Term Variations and Long-Term Dynamics in Commodity Prices. *Management Science*, **46 (7)**, 893–911, doi: 10.1287/mnsc.46.7.893.12034

Shao, J. (1998). *Mathematical Statistics*, New York, Springer

Stărică, C., Granger, C. (2005). Nonstationarities in stock returns. *The Review of Economics & Statistics*, Aug 2005, **87 (3)**, 503–522.

Stock, J. H., Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations.*Journal of Business & Economic Statistics*,**14 (1)**, 11–30.

Subba Rao, T. (1970). The fitting of non-stationary time series models with time-dependent parameters..*J. Roy Stat Soc B*,**32**, 312–322.

Uzan, J.-P. (2003). The fundamental constants and their variation: observational and theoretical status. *REVIEWS OF MODERN PHYSICS*, **75**, doi: https://doi.org/10.1103/RevModPhys.75.403

Webb, J., J. King, M. Murphy, V. Flambaum, R. Carswell, and M. Bainbridge (2011). Indications of a spatial variation of the
ne structure constant. *Physical Review Let- ters*, **107 (19)**, 191101.

Webb, J. K., M. Murphy, V. Flambaum, V. Dzuba, J. Barrow, C. Churchill, J. Prochaska, and A. Wolfe (2001). Further evidence for cosmological evolution of the
ne structure constant. *Physical Review Letters*,**87 (9)**, 091301.

Westfall, P. H. (2014). Kurtosis as Peakedness, 1905–2014.R.I.P. *The American Statistician*, **68 (3)**, 191–195, doi: 10.1080/00031305.2014.917055

Working, H. (1949). The theory of price of storage. *The American Economic Review*, **39 (6)**, 1254–1262, stable url: https://www.jstor.org/stable/1816601

Zhang, A. Mykland, P. A., Aït-Sahalia, Y. (2005). A Tale of Two Time Scales. *Journal of the American Statistical Association*, **100 (472)**, 1394–1411, doi: 10.1198/016214505000000169