Advances In Interest Rate and Risk Modeling

Dissertation an der Fakultät für Mathematik, Informatik und Statistik der Ludwig-Maximillians-Universität München

---

vorgelegt von

Henry Alvorado Port

am 22ten Mai, 2020

1. Gutachter: Prof. Stefan Mittnik, PhD
2. Gutachter: Prof. Dr. Christian Pigorsch
3. Gutachter: Prof. Dr. Svetlozar Rachev

Tag der mündlichen Prüfung: 2ter September, 2020

Summary

This dissertation consists of two projects. The first one that is presented is titled "Free CIR Processes" (Fink et al. (2019)). The Cox-Ingersoll-Ross (CIR) process is the solution to the stochastic differential equation (SDE)

$$\mathrm{d}x(t) = a - bx(t)\mathrm{d}t + \sigma\sqrt{x(t)}\mathrm{d}B(t), \quad x(0) = x_0 > 0,$$

for $t \in [0,\infty[$, where $a,b,\sigma > 0$ and $(B(t))_{t\geq 0}$ is a (classical) Brownian motion. Feller showed in Feller (1951) that as long as the Feller condition $2a \geq \sigma^2$ is satisfied the SDE admits a unique positive solution. For this reason the CIR process initially was used as a model for interest rates by Cox et al. (1985). We introduce a CIR equation in the context of Voicelescu's free probability theory, which is a framework for non-commutative random variables that allows for the computation of joint moments. Although, as mentioned, the CIR model initially was introduced as an interest rate model, the property of positivity of the solution shows that it is especially suited for a volatility model as well, which is why it was implemented in the Heston framework to model stochastic volatility (see Heston (1993)). Presently its validity as an interest rate model is not given due to real-world interest rates not being bound below by zero anymore. That is also why we introduced the free CIR equation as a model for volatility. Were it not for negative interest rates existing in reality, the free CIR equation could be seen as an interest rate model as well (at least historically). In Fink et al. (2019) the existence of the positive solution to such an equation is proven, wherein a big challenge lies in ensuring the positivity of the solution in the framework of free probability theory. The proof starts by showing the existence of a global positive solution for a vector-valued version of the CIR equation. From there on we transform the equation step-by-step into an operator-valued SDE driven by a free Brownian motion, the free counterpart to the classical Brownian motion. The transition from an operator-valued SDE driven by a classical Brownian motion to one driven by a free Brownian motion is done by establishing a isometry between both solutions and by expanding the solution to the latter to a global one, constituting the most complex

part in the proof.

The second project published in "Risks", titled "The Impact of Sovereign Yield Curve Differentials on Value-at-Risk Forecasts for Foreign Exchange Rates" (see Fink et al. (2018)) introduces an ARMA-GARCH type time-series model for predicting the value-at-risk (VaR) of foreign exchange (FX) rates r_t by including the information of the corresponding countries' yield curve information. The motivation and speciality of the ARMAFunX-GARCHFunX is the implementation of Sovereign rate yields as exogenous variables via the use of functional principal component analysis. These yield curves map for each point in time t maturities $m \in M$ to the corresponding yields $y_t \in Y$, resulting in a function $y_t : M \to Y$ for each t. Functional principal component analysis allows the implementation of basically the whole yield curve in a parsimonious way, since this data can be transformed such that only a few data vectors, the principal components, are needed to explain most of the variance in the data. We model the one-day ahead forecasts for the risk measure VaR, given by

$$\mathrm{VaR}_{t|t-1}(p) = \inf_x \{x \in \mathbb{R} | \mathbb{P}(r_t < x | \mathcal{F}_{t-1}) \geq p\}, \text{ for } 0 < p < 1,$$

where $\mathcal{F}_{t-1}$ can be seen as the information available up to time $t-1$ (see e.g. Kuester et al. (2006). We estimate the VaR via

$$\widehat{\mathrm{VaR}}_{t|t-1}(p) = \hat{\mu}_t + \Phi^{-1}(p)\hat{\sigma}_t,$$

where Φ^{-1} is the inverse of the cumulative probability function of the standard normal distribution and $\hat{\mu}_t$ and $\hat{\sigma}_t$ are estimates for the conditional mean and conditional standard deviation, respectively, at time t, obtained via the ARMA and GARCH parts. In Fink et al. (2018) we test and compare the efficiency of the model against competitors by evaluating their respective forecasts via likelihood ratio tests as introduced e.g. in Christoffersen (1998). Furthermore, we discard the myth of the 2-year yield curve difference's significance for modeling FX rates.

Zusammenfassung

Die vorliegende Dissertation besteht aus zwei Projekten. Das zuerst präsentierte trägt den Titel „Free CIR Process“(Fink et al. (2019)). Der sogenannte Cox-Ingersoll-Ross (CIR) Prozess ist die Lösung der stochastischen Differentialgleichung (auf Englisch als „SDE“ abgekürzt)

$$\mathrm{d}x(t) = a - bx(t)\mathrm{d}t + \sigma\sqrt{x(t)}\mathrm{d}B(t), \quad x(0) = x_0 > 0,$$

für $t \in [0,\infty[$, wobei $a,b,\sigma > 0$ und $(B(t))_{t\geq 0}$ eine (klassische) Brownsche Bewegung bezeichnet. Feller hat in Feller (1951) gezeigt, dass diese SDE eine eindeutige positive Lösung besitzt, solange die Feller Bedingung $2a \geq \sigma^2$ erfüllt ist. Aus diesem Grund wurde der CIR Prozess ursprünglich von Cox et al. (1985) zur Modellierung von Zinsen verwendet.

Wir präsentieren eine CIR Gleichung im Rahmen von Voicelescus freier Wahrscheinlichkeitstheorie, welche einen Rahmen für nicht-kommutative Zufallsvariablen darstellt, der erlaubt gemeinsame Momente dieser zu berechnen. Obwohl das CIR Modell, wie bereits erwähnt, ursprünglich als Zinsmodell eingeführt worden ist, zeigt die Eigenschaft der Existenz einer positiven Lösung, dass es auch speziell als Volatilitätsmodell geeignet ist, weswegen es auch im Heston Framework implementiert wurde, um stochastische Volatilität zu modellieren (see Heston (1993)). In der heutigen Zeit ist dessen Alleinstellungsmerkmal als Volatilitätsmodell ohnehin nicht mehr gegeben, da Zinsen in der echten Welt nicht mehr nach unten durch Null beschränkt sind. Daher haben wir die freie CIR Gleichung als Volatilitätsmodell eingeführt. Ohne Negativzinsen in der Realität könnte man die freie CIR Gleichung auch als Zinsmodell in Betracht ziehen (zumindest historisch). In Fink et al. (2019) wird die Existenz der positiven Lösung solch einer Gleichung bewiesen, wobei ein große Herausforderung darin besteht, die Positivität solch einer Lösung im Rahmen der freien Wahrscheinlichkeitstheorie zu erhalten. Der Beweis beginnt damit, die Existenz für eine Vektor-wertige Version der CIR Gleichung zu zeigen. Von diesem Punkt ausgehend formen wir die Gleichung Stück für Stück in eine Operator-wertige SDE, die von einer freien Brownschen Bewe-

gung, dem freien Gegenstück zur klassischen Brownschen Bewegung, getrieben wird, um. Der Übergang von einer Operator-wertigen SDE, getrieben von einer klassischen Brownschen Bewegung, zu einer Operator-wertigen SDE, getrieben von einer freien Brownschen Bewegung, wird geschaffen indem eine Isometrie beider Lösungen gezeigt und die der letzteren zu einer globalen Lösung erweitert wird. Dies stellt den aufwendigsten Abschnitt im Beweis dar.

Das zweite Projekt, welches in „Risks“ unter dem Titel „The Impact of Sovereign Yield Curve Differentials on Value-at-Risk Forecasts for Foreign Exchange Rates“ (siehe Fink et al. (2018)) veröffentlicht wurde, führt ein Zeitreihenmodell des ARMA-GARCH Typs zur Vorhersage des value-at-risk (VaR) von Wechselkursen (auf Englisch als „FX rates“ abgekürzt) r_t ein, indem Informationen aus den Zinskurven der entsprechenden Ländern einbezogen wird. Die Motivation und das Alleinstellungsmerkmal des sogenannten ARMAFunX-GARCHFunX Modells ist das Miteinbeziehen der Staatsanleihen als exogene Variablen mithilfe der funktionalen Hauptkomponentenanalyse. Diese sogennanten yield curves (die englische Bezeichnung ist auch in der deutschen Sprache geläufig.) bilden für jeden Zeitpunkt t Laufzeiten $m \in M$ (für den englischen Begriff „Maturity“) auf yields $y_t \in Y$ ab und sind dadurch als Funktionen $y_t : M \to Y$ für jedes t darstellbar. Mithilfe der funktionalen Hauptkomponentenanalyse können wir quasi die gesamte yield curve auf eine sparsame Art und Weise miteinbeziehen, da Daten so transformiert werden können, dass nur ein paar Daten-Vektoren, die sogenannten Hauptkomponenten, nötig sind, um den Großteil der Varianz in den Daten zu erklären. Wir modellieren Vorhersagen für einen Tag in der Zukunft für das Risiko-Maß $\mathrm{VaR}(p)$, welches durch

$$\mathrm{VaR}_{t|t-1}(p) = \inf_x \{x \in \mathbb{R} | \mathbb{P}(r_t < x | \mathcal{F}_{t-1}) \geq p\}, \text{ for } 0 < p < 1,$$

gegeben ist, wobei $\mathcal{F}_{t-1}$ als die Information angesehen werden kann, welche bis zum Zeitpunkt $t-1$ verfügbar ist (siehe z.B. Kuester et al. (2006)). Wir schätzen den VaR via

$$\widehat{\mathrm{VaR}}_{t|t-1}(p) = \hat{\mu}_t + \Phi^{-1}(p)\hat{\sigma}_t,$$

wobei Φ^{-1} die Inverse der kummulativen Wahrscheinlichkeitsfunktion der Standardnormalverteilung ist und $\hat{\mu}_t$ und $\hat{\sigma}_t$ die Schätzer für den bedingten Erwartungswert und die bedingte Standardabweichung zum Zeitpunkt t darstellen, die mit dem ARMA- und dem GARCH-Teil erhalten werden. In Fink et al. (2018) testen wir die Effizienz

unseres Modells und vergleichen sie mit der anderer Wettstreiter, indem wir die dementsprechenden Vorhersagen mittels Likelihood ratios tests, welche zum Beispiel in Christoffersen (1998) präsentiert werden, gegenüberstellen. Außerdem räumen wir mit dem Mythos des signifikanten Einflusses der 2-Jahres Yield-Differenz für die Modellierung von Wechselkursen auf.

Acknowledgements

There are a few people without whom I would have never made it to the end and a few who influenced me to take this path to begin with.

First of all I thank my parents for always providing me with the support and security I needed to make it through my studies and in particular through the last years. Knowing they were always there for me no matter what gave me the strength I needed to finish this thesis.

At school I struggled with mathematics and physics quite a lot. I did not have a huge problem with English, Latin or French but nature's native language did not open up to me. I did not even understand what it was that I did not understand, which is, as I learned later, the biggest problem for most. As a consequence I hired a tutor to help me with my problem. He helped me seeing things from a different angle and managed to strike my interest in this language by his caring guidance. From there on I was hooked and followed the mathematical lectures at school with enthusiasm.

At the end of my time at school I asked my guitar teacher whether I should study guitar or mathematics. Surprisingly he did not hesitate even for a second and strongly recommended the latter. I followed through with this decision and have never regretted it since.

I want to thank all my colleagues and friends at the institute, in particular Christoph who had to endure all my worries and listen to my doubts.

I want to thank Holger for introducing me to the Institute, for his guidance and especially for the unwavering support throughout my PhD studies.

Of course I have to mention the heart and soul of our chair. Martina has always been taking care of all of us. Whenever I needed someone to talk to she always had an open ear for me and encouraged me whenever I needed it.

Furthermore I want to thank Stefan for taking me in in the first place. He was always there for me, providing unconditional support and was always able to brighten my outlook into the future and to dissolve all my worries no matter what. He is taking care of all of us. I know that I am speaking for all of us. Thank you so much for your

support and your loving care.

And finally I want to thank Christina for her support and for always bringing me back on track whenever I was struggling. Whenever I lost focus and my motivation seemed to fade she always managed to straighten me up and to keep me going. I would not have made it without her, at least not within this time frame and not like this.

Contents

I. Introduction And Methods

I.1. Free Probability Theory and free SDE

In the following section the basic notions and ideas of free probability theory are presented. In the first subsection I.1.1 the basic structure and setup of non-commutative probability spaces is developed. In particular von Neumann algebras are presented as the structure of choice for the current work. Free independence, also referred to as "freeness", a rule to compute joint moments of non-commutative random variables is introduced in I.1.3. In addition a few selected connections between classical and free probability theory will be worked out in I.1.4. Consequently free stochastic differential equations are introduced in I.1.7 and finally the existence proof of the free CIR equation in Fink et al. (2019) is discussed in I.1.8.

I.1.1. Non-commutative Probability Theory

Free probability can be seen as a non-commutative analogue to classical probability theory with a law to calculate joint moments, given by the "freeness" property, an analogous notion to (classical) independence. The most prominent examples of such non-commutative random objects are operators and random matrices. Initially it was "discovered" by Dan Voicelescu while dealing with an isomorphism problem in the context of a special class of operator algebras, the von Neumann algebras. The beginnings of free probability were purely operator-algebraic until it was discovered that large limits of certain random matrix ensembles behave like these free objects in distribution, a connection called "asymptotic freeness" (cf. Mingo and Speicher (2017)). The foreword "Background and outlook" in Voiculescu et al. (2016) by the aforementioned Dan Voicelescu gives a good overview of the subject. Analogous to classical probability theory there are counterparts to the concepts of convolution (cf. Mingo and Speicher (2017)) and the central limit theorem (cf. Nourdin and Taqqu (2014)), where for the latter notion the corresponding limiting distribution is given by the semicircular distribution.

Before proceeding we remind the reader of the following definition (see e.g. Murphy (2014)). Note that we draw from Murphy (2014) for several definitions and notions in this section.

I.1.1 Definition. *An algebra $\mathcal{A}$ is a vector space over a field $\mathbb{F}$ together with an associative $\mathbb{F}$-bilinear map*

$$\mathcal{A} \times \mathcal{A} \to \mathcal{A}, \quad (A,B) \mapsto A \cdot B.$$

We call an element id of $\mathcal{A}$ (the) unit (of $\mathcal{A}$) if it satisfies

$$id \cdot A = A \cdot id, \quad \forall A \in \mathcal{A}.$$

An algebra containing a unit is referred to as unital algebra.

In general a non-commutative probability space is given by the tupel $(\mathcal{A},\varphi)$, where $\mathcal{A}$[1] is a unital algebra and the latter an element of its dual space[2] s.t. the unit in the algebra is mapped to 1, where the last part can be seen as an equivalent to the postulation $\mathbb{P}(\Omega) = 1$ in classical probability theory for the probability measure $\mathbb{P}$ and the whole event space Ω[3] (see I.1.2). Note that this definition of a non-commutative probability space does not have any analytical properties yet and is therefore sometimes referred to as an algebraic non-commutative probability space (cf. Voiculescu et al. (1992)).

Non-commutative probability spaces can be formulated on algebras, algebras with an involution $* : \mathcal{A} \to \mathcal{A}$ (such as e.g. complex conjugation in the context of $\mathbb{C}$), C^*-algebras and von Neumann algebras, which are unital subalgebras of the bounded linear operators on a Hilbert space $\mathcal{H}$ that are closed with respect to the weak operator topology (weak continuity is discussed later), which is only one of a few equivalent definitions (e.g. cf. Murphy (2014)). The choice depends on the context and its applications. To give an example and motivation for such a particular choice we shortly mention that in order to implement a notion of positivity for φ we need an involution on the algebra. We remind the reader of the definition of an involution (see e.g. Murphy (2014)).

[1] We will use this letter for different structures but it will always be clear from the context what is meant.

[2] For a vector space V and a field $\mathbb{F}$, in most cases chosen to be $\mathbb{R}$ or $\mathbb{C}$, the dual space of V denoted by V^* is the set of all linear maps $V \to \mathbb{F}$, referred to as functionals.

[3] See Tao (2012).

I.1.2 Definition. *An involution on an algebra $\mathcal{A}$ over a field $\mathbb{F}$ is a map*

$$^* : \mathcal{A} \to \mathcal{A}, \quad A \mapsto A^*$$

such that

1. $A^{**} = A, \quad \forall A \in \mathcal{A}$
2. $(A+B)^* = A^* + B^*, \quad \forall A,B \in \mathcal{A}$
3. $(\lambda A)^* = \overline{\lambda} A^*, \quad \forall A \in \mathcal{A}$ *and* $\lambda \in \mathbb{F}$
4. $(AB)^* = B^* A^*$.

We refer to algebras with an involution as $*$-algebras or involutive algebras.

I.1.3 Definition. *We call an homomorphism*

$$f : \mathcal{A} \to \mathcal{B}$$

between $*$-algebras $\mathcal{A}$ and $\mathcal{B}$ a $*$-homomorphism (or involutive homomorphism) if

$$f(A^*) = (f(A))^*$$

for all $A \in \mathcal{A}$.

I.1.4 Definition. *The elements $A \in \mathcal{A}$ that satisfy $A^* = A$ are called self-adjoint.*

Note that self-adjoint (continuous) operators A have a real spectrum (cf. Werner (2018)). This is one of the reasons we will restrict our framework to self-adjoint random variables.

We recall the definition of the spectrum of an operator in a normed vector space as it is defined e.g. in Blackadar (2006):

I.1.5 Definition. *Let $A \in \mathcal{A}$ for $\mathcal{A}$ a normed vector space. We define the spectrum of A via*

$$\sigma(A) = \{\lambda \in \mathbb{C} | (A - \lambda id)^{-1} \text{ does not exist}\}^4.$$

As pointed out in Blackadar (2006) this can be seen as a generalization of the concept of eigenvalues.

[4] id denotes the identity operator and is the unit in the corresponding algebra.

I.1.6 Definition. *We call the functional φ on a $*$-algebra positive iff $\varphi(A^*A) \geq 0$ for all $A \in \mathcal{A}$, whereas an operator is called positive if it is of the form B^*B for operators B. These have a non-negative spectrum . We will denote positive operators by $A \geq 0$ and strictly positive operators by $A > 0$. We remark that these are two different notions of positivity. Positivity of a functional means that it maps positive elements (in this context elements of the form A^*A) of the algebra to positive elements in $\mathbb{R}$, where the definition of positivity is known. Further note that $|A| := \sqrt{A^*A}$.*

If the spectrum of an operator is strictly positive, meaning not just non-negative but in $(0,\infty)$, it is invertible. This property will play a crucial role in Fink et al. (2019).

Since von Neumann algebras, the structure of choice for our framework, are a special class of C^*-algebras, we define the latter in the following before proceeding to von Neumann algebras (see cf. Murphy (2014)).

I.1.7 Definition. *A Banach $*$-algebra is a $*$-algebra $\mathcal{A}$ with a complete submultiplicative norm s.t. $\|A\| = \|A^*\|$ for $A \in \mathcal{A}$. If the algebra is unital with $\|id\| = 1$ it is called a unital Banach $*$-algebra.*

I.1.8 Definition. *A C^*-algebra is a Banach $*$-algebra s.t. $\|A^*A\| = \|A\|^2$ for all $A \in \mathcal{A}$.*

Furthermore von Neumann algebras are a special class of C^*-algebras which we will define in the following.

I.1.9 Definition. *Von Neumann algebras $\mathcal{A}$ are the unital $*$-subalgebras of the bounded linear operators on a Hilbert space $\mathcal{H}$, denoted by $\mathcal{B}(\mathcal{H})$[5], that are closed w.r.t. the weak operator topology (and w.r.t. the strong operator topology). We remind the reader that weakly closed in this context means that for $A \in \mathcal{B}(\mathcal{H})$ and a net[6] A_α in $\mathcal{A}$ s.t. $\langle A_\alpha u,v\rangle$ [7] converges to $\langle Au,v\rangle$ for $u,v \in \mathcal{H}$ we always have $A \in \mathcal{A}$. Strongly closed is defined accordingly w.r.t. the strong operator topology.*

Recall that for such an A in $\mathcal{B}(\mathcal{H})$ that $A_i \to A$ in the weak (operator) topology on $\mathcal{B}(\mathcal{H})$ if $\langle A_i\xi,\eta\rangle \to \langle A\xi,\eta\rangle$ for all $\xi,\eta \in \mathcal{H}$. The strong (operator) topology is given by the operator norm $\|A\| = \sup_{\|v\|\leq 1}\|Av\|$ for $v \in \mathcal{H}$.

[5] In this case the algebra's multiplication is given by the composition of operators.

[6] A net in $\mathcal{T}$ is a map $\mathcal{I} \to \mathcal{T}$, where $\mathcal{I}$ is a directed set (see e.g. Werner (2018)). A net with $\mathcal{I} = \mathbb{N}$ is just a sequence.

[7] $\langle\cdot,\cdot\rangle$ denotes the inner product the Hilbert space is endowed with.

Another defining property of von Neumann algebras $\mathcal{A}$ is the fact that it is equal to its bicommutant $\mathcal{A}''$. The commutant $\mathcal{A}'$ is defined by

$$\mathcal{A}' = \{B \in \mathcal{B}(\mathcal{H}) | \text{ for all } A \in \mathcal{A} : BA = AB\}$$

and with $\mathcal{A}'' = (\mathcal{A}')'$, the definition of the bicommutant is clear.

The von Neumann bicommutant theorem (see Neumann (1930)) states that the following properties are equivalent:

1. $\mathcal{A}'' = \mathcal{A}$
2. $\mathcal{A}$ is closed w.r.t. to the weak operator topology
3. $\mathcal{A}$ is closed w.r.t. to the strong operator topology

The third way to define a von Neumann algebra $\mathcal{A}$ among the C^*-algebras is the existence of a unique predual $\mathcal{A}_*$, that is a Banach space s.t. its dual $(\mathcal{A}_*)^*$ is equal $\mathcal{A}$ (see e.g. Sakai (2012)).

For our purposes we choose a von Neumann algebra with a faithful, normal and unital trace. Before explaining all of these notions, we first mention that the von Neumann algebras are classified into different types relating to their projections (see e.g. Blackadar (2006)).

I.1.10 Definition. *We call a $p \in \mathcal{B}(\mathcal{H})$ an (orthogonal) projection if it satisfies*

$$p = p^* = p^2.$$

I.1.11 Remark. *Von Neumann algebras, in contrast to general C^*-algebras, always have projections (see e.g. Werner (2018)). See Remark I.1.25.*

I.1.12 Definition. *We call a (bounded) linear functional φ on a von Neumann algebra $\mathcal{A}$ normal if for a bounded increasing net of positive A_i in $\mathcal{A}$ with $A = sup_i A_i$ we have that $\varphi(A) = \lim_{i\to\infty} \varphi(A_i)$[8] and faithful if $\varphi(A^*A) = 0$ implies $A = 0$.*

I.1.13 Definition. *We call the linear functional φ a tracial state if we have $\varphi(id) = 1$ (unital), $\varphi(A^*A) \geq 0$ (positive) and $\varphi(AB) = \varphi(BA)$[9] for $A,B \in \mathcal{A}$. We will refer to such a φ simply as (unital) trace.*

[8] See e.g. Blackadar (2006).

[9] This property sometimes referred to as traciality can be seen as giving the objects a bit of commutativity back, as is pointed out in Tao (2012).

The following two examples of non-commutative probability spaces can be found in Nica and Speicher (2006) and give a good introduction and feeling for this way of non-commutative thinking.

I.1.14 Example. *A simple and intuitive example of a non-commutative probability space is given by choosing for $\mathcal{A}$ the algebra of $N \times N$-matrices over $\mathbb{R}$ or $\mathbb{C}$ and for φ one chooses the normalized matrix-trace[10] (normality is used to ensure that the unit in the algebra is mapped to 1). That is the trace φ is defined via*

$$\varphi(A) = \frac{1}{N} \sum_{k=1}^{N} a_{kk},$$

where $A \in \mathcal{A}$ and $A = (a_{ij})_{1 \leq i,j \leq N}$. This example will play a particular role when talking about asymptotic freeness later on.

I.1.15 Example. *Quite unintuitively we can represent a classical probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a set, with $\mathcal{F}$ an appropriate σ-algebra and $\mathbb{P}$ a probability measure as a non-commutative probability space. Now consider the algebra $L^{\infty}(\Omega, \mathbb{P})$ of bounded random variables and the expectation, being an element from its dual space, defined by*

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) \, d\,\mathbb{P}(\omega)$$

for $X \in L^{\infty}(\Omega, \mathbb{P})$. Then $(\mathcal{A}, \varphi)$ with $\mathcal{A} = L^{\infty}(\Omega, \mathbb{P})$ and $\varphi = \mathbb{E}$ is a non-commutative probability space. But we see the level on which free probability is formulated. It basically represents the same thing as the classical formulation does, but the central object of study is an algebra of objects rather than the particular object, in this case the classical random variables.

As mentioned before there are different possible nuances for the definition of a non-commutative probability space but for our purposes we will define a non-commutative probability space the following way.

I.1.16 Definition. *We call $(\mathcal{A}, \varphi)$ a non-commutative probability space, where $\mathcal{A}$ is a von Neumann algebra and φ a faithful, normal and unital trace on $\mathcal{A}$.*

I.1.17 Remark. *As mentioned before we will only consider self-adjoint elements.*

[10] We will later denote the normalized matrix-trace, the trace divided by the dimension of the corresponding matrix, by tr.

Note that we did not not use the term "free" yet. As mentioned before, freeness is essentially a rule to compute joint moments of non-commutative objects. Before introducing the notion of freeness we first show the construction of such a functional φ that can serve as the free counterpart[11] to expectation in this setting.

I.1.2. Functional Calculus and Spectral Theorem

We will present the spectral theorem for self-adjoint operators of $\mathcal{B}(\mathcal{H})$ and introduce the continuous and the (measurable) Borel functional calculus. A functional calculus can be seen as a way to give meaning to the application of certain functions to operators. We will follow and heavily draw from Werner (2018)[12] for this introduction.[13] In this process the free counterpart to classical expectation is derived by introducing the spectral measure. Note that this theory can be formulated in a more general way for normal operators[14] of $\mathcal{B}(\mathcal{H})$ (cf. Murphy (2014)) but we will introduce this slightly easier theory and access to support the main understanding of the core idea. That is why we introduce these concepts mainly for the self-adjoint bounded operators on a Hilbert space since the generality suffices to transport the idea and to give a first introduction to these concepts. We cite (Werner, 2018, Korollar VII.1.2.):

I.1.18 Theorem. *For a self-adjoint $A \in \mathcal{B}(\mathcal{H})$ we have that $\sigma(A) \subset \mathbb{R}$. We have in particular $\sigma(A) \subset [m(A), M(A)]$, where $m(A) = \inf\{\langle Ax, x\rangle \,|\, \|x\| = 1\}$ and where $M(A) = \sup\{\langle Ax, x\rangle \,|\, \|x\| = 1\}$. For a positive A we further have $\sigma(A) \subset [0, \infty)$.*

I.1.19 Remark. *Note that the numerical range*

$$W(A) = \{\langle Ax, x\rangle \,|\, \|x\| = 1\}$$

is bounded.

We proceed by defining the continuous functional calculus as in (Werner, 2018, Satz VII.1.3):

[11] Technically it should be referred to as the non-commutative counterpart but we use the term "free" here to contrast the free with the classical setting.
[12] Mostly from Chapters VII and IX in Werner (2018).
[13] For another prominent source see Blackadar (2006).
[14] An operator is referred to as normal if it commutes with its adjoint.

I.1.20 Theorem. *Let $A \in \mathcal{B}(\mathcal{H})$ be self-adjoint. Then there is a unique map*

$$\Phi : C(\sigma(A))^{15} \to \mathcal{B}(\mathcal{H})$$

s.t.

1. $\Phi(\boldsymbol{id}) = A$ *and* $\Phi(\boldsymbol{1}) = id$, *where* $\boldsymbol{id}$ *is the identity function and* $\boldsymbol{1}$ *is the constant function that maps to 1.* [16]
2. Φ *is an involutive homomorphism of algebras, where we have in particular that* $\Phi(f \cdot g) = \Phi(f) \circ \Phi(g)$, *with* $\circ$ *denoting composition of operators.*
3. Φ *is continuous*

I.1.21 Definition. *We call the map Φ the continuous functional calculus of A and we write $f(A) := \Phi(f)$ for $f \in C(\sigma(A))$.*

We state Satz VII.1.4. in Werner (2018) that describes several important properties of the functional calculus:

I.1.22 Theorem. *Let A be a self-adjoint element of $B(\mathcal{H})$ and the map $f \mapsto f(A)$ defines the continuous functional calculus on $C(\sigma(A))$. Then it holds that*

1. $\|f(A)\| = \|f\|_\infty = \sup_{\lambda \in \sigma(A)} |f(\lambda)|$
2. *For $f \geq 0$ we have that $f(A) \geq 0$*
3. *$Ax = \lambda x$ implies $f(A)x = f(\lambda)x$*
4. $\sigma(f(A)) = f(\sigma(A)) = \{f(\lambda) | \lambda \in \sigma(A)\}$
5. $\{f(A) | f \in C(\sigma(A))\}$ *is a commutative algebra of operators. All $f(A)$ are normal in the sense that they commute with their adjoint. Furthermore $f(A)$ is real-valued iff f is self-adjoint.*

The continuous functional calculus can be extended to elements in $B(\sigma(A))$, the bounded measurable functions on $\sigma(A)$, which is then referred to as the measurable (Borel) functional calculus. We formulate the measurable functional calculus by the following theorem (see (Werner, 2018, Satz VII.1.6.)).

[15] $C(\sigma(A))$ denotes the continuous functions on $\sigma(A)$.
[16] id denotes, as before, the unit in the corresponding algebra.

I.1.23 Theorem. *Let $A \in \mathcal{B}(\mathcal{H})$ be self-adjoint. Then there is a unique map*

$$\hat{\Phi} : B(\sigma(A)) \to \mathcal{B}(\mathcal{H})$$

s.t.

1. $\hat{\Phi}(\boldsymbol{id}) = A$ *and* $\hat{\Phi}(\boldsymbol{1}) = id$
2. $\hat{\Phi}$ *is an involutive homomorphism of algebras*
3. $\hat{\Phi}$ *is continuous*
4. $f_n \in B(\sigma(A))$ *with* $\sup_n \|f_n\|_\infty < \infty$ *and* $f_n(t) \to f(t)$ *for all* $t \in \sigma(A)$ *implies that* $\left\langle \hat{\Phi}(f_n)x, y \right\rangle \to \left\langle \hat{\Phi}(f)x, y \right\rangle$ *for all* $x, y \in \mathcal{H}$.

We denote the characteristic function (or indicator function) on a set S by χ_S. It is defined by

$$\chi_S(t) = \begin{cases} 1, & \text{for } t \in S \\ 0, & \text{else} \end{cases} .$$

We further have by (Werner, 2018, Lemma VII.1.7)):

I.1.24 Lemma. *We have for Borel sets $S \subset \sigma(A)$ that $E_S := \chi_S(A)$ is an orthogonal projection.*

I.1.25 Remark. *As pointed out in Werner (2018), for a self-adjoint operator A in a von Neumann algebra $\mathcal{A}$ we always have for $S \subset \sigma(A)$ that $E_S \in \mathcal{A}$, a property that is not guaranteed for general C^*-algebras.*

I.1.26 Remark. *Note that we apply the functional calculus to give meaning to the expression $\chi_S(A)$.*

We proceed by developing the notion of a projection-valued measure, the spectral measure.

See (Werner, 2018, Lemma VII.1.8)):

I.1.27 Lemma. *Let $A \in \mathcal{B}(\mathcal{H})$ be self-adjoint. Then we have that*

1. $\chi_\varnothing(A) = 0$, *and* $\chi_{\sigma(A)}(A) = id$

2. *For pairwise disjoint Borel sets* $S_1, S_2, \cdots \subset \sigma(A)$ *and* $x \in \mathcal{H}$ *we have that*

$$\sum_{i=1}^{\infty} \chi_{S_1}(A)x = \chi_{\cup_{i=1}^{\infty} S_i}(A)x.$$

3. $\chi_S(A)\chi_T(A) = \chi_{S\cap T}(A)$ *for Borel sets* $S,T \subset \sigma(A)$.

Before applying both lemmata we give the definition of a spectral measure (see e.g. (Werner, 2018, Definition VII.1.9)).

I.1.28 Definition. *Let* Σ *be the Borel* σ*-algebra on* $\mathbb{R}$*. We call a map*

$$E : \Sigma \to \mathcal{B}(\mathcal{H}), \quad S \mapsto E_S$$

spectral measure if the E_S *are orthogonal projections and we have that*

1. $E_\emptyset = 0$ *and* $E_{\mathbb{R}} = id$

2. *For pairwise disjoint* $S_1, S_2, \cdots \in \Sigma$ *we have*

$$\sum_{i=1}^{\infty} E_{S_i}(x) = E_{\cup_{i=1}^{\infty} S_i}(x) \text{ for all } x \in \mathcal{H}.$$

In particular E has compact support if there is a compact set K s.t. $E_K = \mathrm{id}$.

I.1.29 Remark. *We have in particular that* $E_S E_T = E_T E_S = E_{S\cap T}$.

By the two lemmata above this gives us that

$$E : S \mapsto \chi_{S\cap\sigma(A)}(A), \text{ for } S \in \Sigma$$

is a spectral measure.

This allows the definition of the integral

$$\int f \mathrm{d}E$$

for a measurable function f with respect to a spectral measure E. This is done via step functions (see Werner (2018)).

I.1.30 Remark. *If* E *has compact support* K*, we may also denote this integral by* $\int_K f(\lambda) dE_\lambda$*, where we define* $E_\lambda = E((-\infty,\lambda])$ *for each* λ*. We may switch between these notations depending on the context.*

We finally present the spectral theorem for self-adjoint bounded operators (see (Werner, 2018, Theorem VII.1.13)):

I.1.31 Theorem. *Let $A \in \mathcal{B}(\mathcal{H})$ be self-adjoint. Then there is a unique spectral measure with compact support on $\mathbb{R}$, denoted by E s.t.*

$$A = \int_{\sigma(A)} \lambda dE_\lambda.$$

The map $f \mapsto f(A) = \int f(\lambda) dE_\lambda$ defines the measurable functional calculus. We have that $f(A)$ is defined by

$$\langle f(A)x, y\rangle = \int_{\sigma(A)} f(\lambda)\, d\langle E_\lambda x, y\rangle .$$

I.1.32 Remark. *The above notation needs a bit of clarification. Since E is a spectral measure, that is it maps Borel sets to projections, the notation $d\langle E_\lambda x, y\rangle$ means integration with respect to the unique spectral measure for which the existence is ensured by the above theorem, that is for the corresponding Borel sets S the map $S \mapsto \langle E_S x, y\rangle$.*

I.1.33 Remark. *Note that we have for a measurable $S \subset \sigma(A)$ in particular that*

$$\int \chi_{\sigma(A)}\, dE = E_{\sigma(A)} = id.$$

These results can be generalized to normal operators, that is operators that commute with their adjoints, which is trivially given for self-adjoint operators. We state the following generalization given by (Werner, 2018, Korollar IX.3.8):

I.1.34 Theorem. *Let $A \in \mathcal{B}(\mathcal{H})$ be a normal element. Then there is an isometric*[17] *$*$-homomorphism*

$$\Phi : C(\sigma(A)) \to \mathcal{B}(\mathcal{H})$$

s.t.

1. $\Phi(\boldsymbol{id}) = A$
2. $\Phi(\overline{\boldsymbol{id}}) = A^*$
3. $\Phi(\boldsymbol{1}) = id.$

[17] We call a map $\iota : \mathcal{A} \to \mathcal{B}$ between normed vector spaces $(\mathcal{A}, \|\cdot\|_{\mathcal{A}})$ and $(\mathcal{B}, \|\cdot\|_{\mathcal{B}})$ an isometry or "isometric" if $\|\iota(\cdot)\|_{\mathcal{B}} = \|\cdot\|_{\mathcal{A}}$.

Having been introduced to the concepts of the functional calculus and the spectral theorem for self-adjoint operators in $\mathcal{B}(\mathcal{H})$ we emphasize the following construction.

I.1.35 Remark. *Note that we have in particular an isometric homomorphism*

$$B(\sigma(A)) \to \langle A, id \rangle \text{ [18]}$$

for a self-adjoint $A > 0$. This construction is of special importance for the main proof in Fink et al. (2019).

Now we can define our functional φ via

$$\begin{aligned} \varphi(A) &= < A\xi, \xi > \\ &= < \int_{\sigma(A)} \lambda \, \mathrm{d}E_\lambda \xi, \xi > \\ &= \int_{\sigma(A)} \lambda \, \mathrm{d} \langle E_\lambda \xi, \xi \rangle , \end{aligned}$$

where $\xi \in \mathcal{H}$ is a unit vector, that is $\|\xi\| = 1$, and $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathcal{H}$ (cf. Voiculescu (2002)). Defining the scalar-valued measure

$$\mathrm{d}\mu(\cdot) = \mathrm{d} \langle E(\cdot)\xi, \xi \rangle$$

gives a scalar-valued "expectation", where the real-valued measure μ is referred to as the spectral (probability) distribution of A. This "expectation" can naturally be generalized for certain functions of A via the functional calculus.

For our case (the case of a self-adjoint operator) μ is a compactly supported measure on $\mathbb{R}$ (see e.g. Voiculescu (2002)). We therefore can unambiguously write

$$\varphi(A^k) = \int_{\sigma(A)} \lambda^k \mathrm{d}\mu.$$

I.1.3. Free Independence

We proceed by finally introducing the notion of "freeness". Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space in the sense of Definition I.1.16.

[18] This denotes the subalgebra generated by A and the identity.

I.1.36 Definition. *Let $\mathcal{A}_1, \dots, \mathcal{A}_n$ be subalgebras of $\mathcal{A}$. We call $\mathcal{A}_1, \dots, \mathcal{A}_n$ free if*

$$\varphi(A_{i(1)} \cdots A_{i(n)}) = 0,$$

when

$$\varphi(A_{i(s)}) = 0 \text{ and } i(s+1) \neq i(s) \text{ for all } s.$$

Random variables are called free if the algebras they generate are free.

I.1.37 Definition. *We call a non-commutative probability space together with the notion of freeness a free probability space.*

In order to motivate this, at first glance, rather unintuitive definition, we draw from the introduction to the concept of free probability presented in Tao (2012) since this section helps understanding the philosophy behind this concept and this way of non-commutative thinking. Note that in this philosophical discussion we start with explaining the philosophy behind non-commutative probability theory before discussing the concept of freeness. We point this out since although non-commutative probability theory is a framework for possibly non-commutative random variables, commuting random variables can be considered in a non-commutative probability space as well, as we have seen in Example I.1.15, which makes the notion of a non-commutative probability space more general than a classical probability space. But freeness and therefore free probability theory is a concept particularly for non-commutative objects, since commuting random variables are free only in very specific situations, as we will see with an example later on. Therefore, free probability theory is not to be seen as a generalization of classical probability theory but rather as a parallel field of study. We heavily draw from (Tao, 2012, Chapter 2.5., p. 155ff) in the following:

First off we try to explain how the transition from classical to non-commutative probability is to be understood philosophically by comparing it with the transition from measure theory to (classical) probability theory. The objects of study in the former are the measurable sets and the measurable functions. As the name suggests, these notions depend on a measure space which is presented as having a central and defining property for those objects (see Tao (2012)). In (classical) probability theory however the level on which we observe things (we can just think of this as an analogy to working on the micro- and the macroscopic level) is the level of events and their probabilities. We know of course that the events belong to a space but this is rather

obscured. The notions "random variable" and "expectation" are then concepts evolving from there or objects that can be defined on that level of abstraction.

The step from classical probability to non-commutative probability is similar. Now we can abstract matters even more and view the structure of random variables and the corresponding expectations, where the chosen structure is taken to be a specific algebra (of random variables), as the central objects or the building blocks of the theory. On this level the initial sample space, the algebra of the corresponding events, as well as the initial probability measure are not central anymore. As before this can be understood as viewing something on a microscope but not zoomed in enough to see these underlying concepts.

As pointed out in Tao (2012) one reason for taking this further step of abstraction is that constructs built on the levels of algebras and their expectation are rather stable when taking particular limits. The example given in Tao (2012) is the limit of normalized matrix moments, a construction that we will encounter later in the context of approximate freeness.

The second reason is given by the simple fact that a free probability space does not need to specify a sample space or an event space in the classical sense and allows, as mentioned before, for a more general formulation by allowing non-commutative random variables. This generality allows for the implementation of the spectral theory of random matrices and operators into the context of probability theory in such a way that the matrices, or operators, respectively, take over the role of random variables with a corresponding expectation.

A corresponding notion of "independence" on this level obviously needs to be formulated on the level of the central objects, which are the algebras of the random variables, which are in turn random matrices or operators, in the context and abstraction level of free probability. "Freeness" or "independence" can roughly be seen as the lack of commutativity or dependence. There are structures that are more free than others. To see how these notions can differ depending on the structure in question we present the following demonstration given in (Tao, 2012, p. 157) that gives further insight in the development of such a notion as "free independence".

I.1.38 Example. *Let us first consider the case of abelian groups. Trying to put as few restrictions on the corresponding structure arising by two generators A and B but of course obeying the axioms or properties of an abelian group we get the free abelian group of words A^nB^m for $n,m \in \mathbb{Z}$. Note that for an abelian group these words commute, since*

A and B commute by definition. This structure is isomorphic to the group $\mathbb{Z}^2 = \mathbb{Z} \times \mathbb{Z}$. So loosely speaking $\mathbb{Z}^2$ is as free as it gets for such an abelian group.

Let's now relax the requirement of the structure in question and allow for general groups. In this case the word $A^{n_1}B^{m_1}\cdots A^{n_k}B^{m_k}$ for $n_i, m_i \in \mathbb{Z}$ and $1 \leq i \leq k$ can not be simplified to the form A^nB^m, since A and B no longer commute. Therefore this structure, the free group $\mathbb{F}_2$, has a lot more elements than $\mathbb{Z}^2$.

Consider now two classical bounded and scalar-valued, in particular $\mathbb{R}$-valued random variables X and Y on a classical probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In this setting we have that X and Y are independent if

$$\mathbb{E}(f(X)g(Y)) = 0 \tag{I.1}$$

for all well-behaved[19] functions $f,g : \mathbb{R} \to \mathbb{R}$ s.t. $\mathbb{E}(f(X)) = \mathbb{E}(g(Y)) = 0$.[20] In this case we have formulated classical independence "in the same form" as free independence. The generalization to commuting bounded self-adjoint X and Y is straightforward. As in the previous group-theoretic demonstration, the expectation of a product of multiple such factors of corresponding functions depending on X or Y can be brought in such a form as long as they commute; since multiplication is well-defined these can always be ordered in the form in (I.1). Therefore, the definition of independence given above suffices. But for non-commuting variables this is not the case anymore and therefore such a product can not be arranged in such a way. Therefore we introduce another form of independence for these objects, namely the following. We call X and Y free if

$$\mathbb{E}(f_1(X)g_1(Y)\cdots f_k(X)g_k(Y)) = 0$$

for all well-behaved functions $f_i, g_i : \mathbb{R} \to \mathbb{R}$ s.t. $\mathbb{E}(f_i(X)) = \mathbb{E}(g_i(Y)) = 0$ for all $1 \leq i \leq k$.

This example shows that the notion of "independence" or "freeness" changes depending on the corresponding structure and further illustrate how such an at first glance unintuitive definition can be derived.

I.1.39 Definition. *We refer to the self-adjoint elements of the free probability space*

[19] An object is said to be well-behaved if it does not show extreme or unnatural behaviour. What it means to be unnatural or extreme, of course, depends on the object in question. It more or less just means to exclude certain edge-cases.

[20] See e.g. Whittle (2012).

as (free) random variables.

We give an example to show how "freeness" behaves.[21]

I.1.40 Example. *Let $\{A,C\}$ and $\{B,D\}$ be free.*

1. *We consider*

$$\varphi((A-\varphi(A))(B-\varphi(B)) = \varphi(AB) - \varphi(A)\varphi(B)$$

and since $\varphi(A-\varphi(A)) = \varphi(B-\varphi(B)) = 0$ we get by the definition of freeness

$$\varphi(AB) = \varphi(A)\varphi(B),$$

which is what is to be expected coming from the classical situation. However, the results do differ for higher moments.

2. *In the same way we get*

$$\varphi(ABCD) = \varphi(AC)\varphi(B)\varphi(D) + \varphi(A)\varphi(C)\varphi(BD) - \varphi(A)\varphi(C)\varphi(B)\varphi(D),$$

which differs from what one would expect from classical probability.

3. *A special case of the above example, namely the case $A = C$, $B = D$ with the additional property $\varphi(A) = \varphi(B) = 0$, shows the different natures of the commutative and non-commutative world especially well. In the commutative case we would get*

$$\varphi(ABAB) = \varphi(A^2)\varphi(B^2),$$

which in general does not vanish. For the non-commutative case we have for free A and B that

$$\varphi(ABAB) = 0$$

by the definition of freeness.

We mentioned before that commuting random variables are free only in very specific situations. We illustrate this by the following example (see e.g. Nica and Speicher (2006) or Mingo and Speicher (2017)):

[21] Such basic examples can be found in most of the standard texts to free probability

I.1.41 Example. *Let A,B be commuting random variables and let A and B be free. Consider*

$$\varphi(AABB) = \varphi(ABAB),$$

where the left side is equal to $\varphi(A^2)\varphi(B^2)$ and we get for the right-hand side that it simplifies to $\varphi(A^2)\varphi(B)^2 + \varphi(A)^2\varphi(B^2) - \varphi(A)^2\varphi(B)^2$. This yields

$$\varphi((A-\varphi(A))^2)\varphi((B-\varphi(B))^2) = 0$$

and by the properties of φ this implies that A or B must be scalar.

These examples hint towards the importance of non-crossing partitions and free cumulants for free probability theory. We will shortly describe the relevance of set partitions for classical probability and then establish the connection to the free case. This will lead us to the free version of the central limit theorem (cf. Nourdin and Taqqu (2014)) and we will see that the semicircular distribution takes over the role of the Gaussian distribution in the free case.

I.1.4. Free Partitions and the Free Central Limit Theorem

As presented in Mingo and Speicher (2017) we present Wick's formula (cf. Wick (1950)) by recalling the notion of a Gaussian random vector. This formula gives an interesting way of calculating joint moments via its covariances and has been known for even longer (cf. Isserlis (1918)). We will present the free analogue to the Wick formula to show the similarities between the free and the classical "version". Finally we formulate the free counterpart of the central limit theorem. We will see that the semicircle law plays a similar role in free probability theory as the Gaussian distribution does in the classical case. This has crucial implications for the theory of free stochastic differential equations and their connection to the classical case. The similar properties of these two distributions are necessary for the isometry argument in the main proof of Fink et al. (2019).

For this, note that higher cumulants than the second of Gaussian random variables vanish. Although the theory of cumulants and in particular their free counterparts, is very interesting and gives deep insight into the structure of free random variables and random variables in general, we will not go into detail but will only take away the implications for calculating joint moments of (free) random variables. We heavily draw

from Mingo and Speicher (2017)[22] in the following. We start by defining a Gaussian random vector (see e.g. Mingo and Speicher (2017)).

I.1.42 Definition. *We call a random vector $X = (X_1, \dots, X_n)$ Gaussian if there is a positive definite $n \times n$ real symmetric matrix M s.t.*

$$\mathbb{E}(X_{i_1} \cdots X_{i_k}) = \int_{\mathbb{R}^n} t_{i_1} \cdots t_{i_k} \frac{\exp(-\langle Mt, t\rangle/2)}{(2\pi)^{n/2} \det(M)^{-1/2}} dt,$$

where $\langle \cdot, \cdot \rangle$ denotes the corresponding inner product.

Let $i_i, \dots, i_k \in [n] := \{1,2,\dots,n\}$ and $\mathcal{P}_2(k)$ be the pairings of $[k]$, that is the groupings of this set in distinct subsets containing exactly two elements. For such a pairing $\pi \in \mathcal{P}_2(k)$ we define, for an even k (see Mingo and Speicher (2017))

$$\mathbb{E}_\pi(X_1, \dots, X_k) = \prod_{(a,b)\in\pi} \mathbb{E}(X_a X_b). \tag{I.2}$$

Wick's formula gives

$$\mathbb{E}(X_{i_1} \cdots X_{i_k}) = \sum_{\pi \in \mathcal{P}_2(k)} \mathbb{E}_\pi(X_{i_1}, \dots, X_{i_k}) \text{ where } i_1, \dots, i_k \in [n].$$

Before proceeding to the pendant in free probability we give the following definition.

I.1.43 Definition. *The number C_n defined by*

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{n!(n+1)!}$$

is called the n-th Catalan number.

These numbers play a particular role in free probability theory. As mentioned before, the semicircular distribution takes over the role of the Gaussian distribution and we therefore use the following definition from Mingo and Speicher (2017).

I.1.44 Definition. *A self-adjoint random variable S with vanishing odd moments and even moments $\varphi(S^{2n}) = \sigma^{2n} C_n$ with $\sigma > 0$ constant is called a semicircular element of variance σ^2. For $\sigma = 1$ we call it a standard semicircular element.*

[22] In this thesis we mostly draw from Chapters 1,2 and 4 in Mingo and Speicher (2017).

In the following we give the probability density function of the semicircular probability distribution.

I.1.45 Definition. *For $r > 0$ the probability density function, supported on the interval $[-r,r]$ given by*

$$f(\lambda) = \begin{cases} \frac{2}{\pi r^2}\sqrt{(r^2-\lambda^2)}, & \text{for } \lambda \in [-r,r] \\ 0, & \text{else} \end{cases},$$

is referred to as semicircular distribution or sometimes as Wigner semicircle distribution to the honour of Eugene Wigner.

I.1.46 Remark. *We have in particular the connection*

$$C_n = \frac{1}{2\pi}\int_{-2}^{2} \lambda^{2n}\sqrt{4-\lambda^2}\,d\lambda,$$

which can e.g. be seen in Mingo and Speicher (2017). That means that the $2n$-th moment of the semicircular distribution with $r = 2$ is equal to the n-th Catalan number.

The free counterpart to Wick's formula (with φ taking over the role of $\mathbb{E}$ in (I.2)) for semicircular elements S_i[23] is given by

$$\varphi(S_{i_1}\cdots S_{i_k}) = \sum_{\pi\in \mathrm{NC}_2(k)} \varphi_\pi(S_{i_1},\ldots,S_{i_k}), \tag{I.3}$$

where $\mathrm{NC}_2(k)$ denotes the non-crossing pairings of $[k]$, where non-crossing is quite intuitively explained by the following example.

I.1.47 Example. *If we partition the set $\{1,2,3,4\}$ in pairings we have that $\{(1,2),(3,4)\}$ is non-crossing, whereas $\{(1,3),(2,4)\}$ is crossing. To visualize the process one can just note the numbers side by side and drawing connecting lines between the pairings. When connecting lines cross the pairing is crossing.*

We furthermore state the free central limit theorem:

I.1.48 Theorem. *If $(A_i)_{i\in\mathbb{N}}$ are self-adjoint, freely independent and identically distributed with zero mean and variance σ^2, we have that*

$$\frac{\sum_{k=1}^{l} A_k}{\sqrt{l}} \to S \text{ for } l \to \infty$$

[23] For more details we refer the reader to Mingo and Speicher (2017).

in distribution, where S is a centred semicircular element of variance σ^2.

As pointed out in Mingo and Speicher (2017) this result was proven first in Voiculescu (1985) and later on by Speicher (1990).

In the context of Wick's formula we encountered the set of pairings $\mathcal{P}_2(n)$ of $[n]$. Pairings are a special case of partitions, a notion we will shortly define (see Mingo and Speicher (2017) and Speicher (2019)).

I.1.49 Definition. *We call $\pi = \{V_1, \ldots, V_r\}$ a partition of $[n]$ if the following properties hold:*

1. *The V_i are subsets of $[n]$ and non-empty for $i = 1, \ldots, r$*
2. *The V_i are pairwise disjoint for $i = 1, \ldots, r$*
3. $\bigcup_{i=1}^{r} V_i = [n]$

We call the V_i blocks and denote the set of all partitions of $[n]$ by $\mathcal{P}([n])$. We define $\mathcal{P}(n) = \mathcal{P}([n])$.

Although the idea of generalizing pairings to general set partitions is straightforward we recall the formal definition of non-crossing set partitions, as it is done in Mingo and Speicher (2017).

I.1.50 Definition. *We call a partition $\pi \in \mathcal{P}(n)$ non-crossing if there are no numbers $i,j,k,l \in [n]$ with $i < j < k < l$ such that i and k are in the same block of π and j and l are in the same block of π but i and j are not. The set of all non-crossing partitions of $[n]$ is denoted by $NC(n)$.*

To visualize this definition we refer the reader to Example I.1.47. It is the same core idea but allowing general partitions, as long as the connecting lines do not cross.

The formula connecting moments and free cumulants can be seen as a generalization of the free counterpart to Wick's formula.

The connection between the Wick formula and its free counterpart hints towards the special role of non-crossing partitions of a set for free probability theory. We close this subsection with the introduction of the free cumulants and a formula connecting moments and cumulants, that is a very useful tool for explicit computations of joint moments of free random variables.

The free cumulants are inductively (and implicitly) defined by the following definition (cf. Mingo and Speicher (2017)[24]).

I.1.51 Definition. *Let $(\mathcal{A},\varphi)$ be a non-commutative probability space. The corresponding free cumulants*

$$\kappa_n : \mathcal{A}^n \to \mathbb{C}$$

are defined by

$$\varphi(A_1 \cdots A_n) = \sum_{\pi \in NC(n)} \kappa_\pi(A_1, \ldots, A_n), \tag{I.4}$$

where for a partition $\pi = \{V_1, \ldots, V_r\}$ we define

$$\kappa_\pi(A_1, \ldots, A_n) = \prod_{\substack{V \in \pi \\ V=(i_1,\ldots,i_l)}} \kappa_l(A_{i_1}, \ldots, A_{i_l}).$$

Note that as mentioned in Mingo and Speicher (2017), the $i_1, \ldots, i_l$ are ordered increasingly.

We again analyze the corresponding situation in the classical counterpart in order to help seeing the parallels and analogies between these different "probability theories".

For classical random variables $X_1, \ldots, X_n$ in a classical probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we obtain a similar relation between the classical cumulants k_l[25] by the following relation

$$\mathbb{E}(X_1 \cdots X_n) = \sum_{\pi \in \mathcal{P}(n)} k_\pi(X_i, \ldots, X_n),$$

where

$$k_\pi(X_1, \ldots, X_n) = \prod_{\substack{V \in \pi \\ V=(i_1,\ldots,i_l)}} k_l(X_{i_1}, \ldots X_{i_l}).$$

We again see that in the classical setting we use all partitions in the set of partitions of $[n]$, denoted by $\mathcal{P}(n)$, whereas in the non-commutative setting only the non-crossing are contributing.

I.1.52 Remark. *Knowing that the k-th cumulants of a Gaussian and the k-th cumulants of a semicircular random variable both vanish for $k > 2$ (cf. Mingo and Speicher*

[24] For further reading we refer to Nica and Speicher (2006). Many of these concepts are introduced there and discussed in more detail.

[25] These (classical) cumulants can be obtained via the series expansion of the corresponding logarithm of the moment-generating function. We will only use them here for the sake of comparison with regard to free cumulants and their corresponding roles.

(2017)) gives further insight into the correspondence between Wick's formula and its free counterpart and stresses furthermore the correspondence and roles between those distributions and therefore can serve as bridge between the classical and the free world.

We illustrate the procedure of computing free cumulants by the following examples[26]:

I.1.53 Example. *We consider the most simple but nevertheless quite interesting cases* $n = 1$ *and* $n = 2$ *first. Applying the formula given by* (I.4) *we get for the first case*

$$\varphi(A_1) = \kappa_1(A_1)$$

and for the second

$$\begin{aligned} \varphi(A_1A_2) &= \kappa_{\{(1,2)\}}(A_1, A_2) + \kappa_{\{(1),(2)\}}(A_1, A_2) \\ &= \kappa_2(A_1, A_2) + \kappa_1(A_1)\kappa_1(A_2). \end{aligned}$$

This gives us by the above

$$\kappa_2(A_1, A_2) = \varphi(A_1A_2) - \varphi(A_1)\varphi(A_2).$$

Finally we present the final example from Mingo and Speicher (2017) to show the recursive procedure.

I.1.54 Example. *For the case* $n = 3$ *we get*

$$\begin{aligned} \varphi(A_1A_2A_3) &= \kappa_{\{(1,2,3)\}}(A_1, A_2, A_3) + \kappa_{\{(1,2),(3)\}}(A_1, A_2, A_3) + \kappa_{\{(1),(2,3)\}}(A_1, A_2, A_3) \\ &+ \kappa_{\{(1,3),(2)\}}(A_1, A_2, A_3) + \kappa_{\{(1),(2),(3)\}}(A_1, A_2, A_3) \\ &= \kappa_3(A_1, A_2, A_3) + \kappa_2(A_1, A_2)\kappa_1(A_3) + \kappa_2(A_2, A_3)\kappa_1(A_1) \\ &+ \kappa_2(A_1, A_3)\kappa_1(A_2) + \kappa_1(A_1)\kappa_1(A_2)\kappa_1(A_3). \end{aligned}$$

Inserting the second and first cumulants we can solve the above equation by $\kappa_3(A_1, A_2, A_3)$ *and get*

$$\begin{aligned} \kappa_3(A_1, A_2, A_3) &= \varphi(A_1A_2A_3) - \varphi(A_1)\varphi(A_2A_3) \\ &- \varphi(A_2)\varphi(A_1A_3) - \varphi(A_3)\varphi(A_1A_2) + 2\varphi(A_1)\varphi(A_2)\varphi(A_3). \end{aligned}$$

[26] Such basic examples can be found in most of the standard texts to free probability

We conclude this section with another characterization of freeness by free cumulants. We quote (Nica and Speicher, 2006, p. 173): "Free independence can be characterized by the vanishing of mixed cumulants." Thus we state accordingly Theorem 11.16 in Nica and Speicher (2006):

I.1.55 Theorem. *Let $(\mathcal{A},\varphi)$ be a non-commutative probability space and let $(\kappa_n)_{n\in\mathbb{N}}$ be the corresponding free cumulants. Consider unital subalgebras $(\mathcal{A}_i)_{i\in I}$ of $\mathcal{A}$. Then the following two statements are equivalent.*

1. *$(\mathcal{A}_{i\in I})$ are freely independent.*
2. *We have for all $n \geq 2$ and for all $A_j \in \mathcal{A}_{i(j)}$ $(j = 1,\ldots,n)$ with $i(1),\ldots,i(n) \in I$ that $\kappa_n(A_1,\ldots,A_n) = 0$ whenever there exist $1 \leq l,k \leq n$ with $i(j) \neq i(k)$.*

As pointed out in Nica and Speicher (2006) this simplifies things, since in contrast to the aforementioned characterization neither the condition $i(1) \neq i(2) \neq \cdots \neq i(n)$ nor the assumption that the A_i are centred are necessary.

I.1.5. Asymptotic Freeness

As mentioned before, free probability was developed as a purely theoretic concept and initially was not very appealing for data applications until connections between the eigenvalue distribution of certain random matrices and free probability theory had been discovered. "Asymptotic freeness", as described for example in Mingo and Speicher (2017), is the fact that some random matrices (with independent entries) are free in the limit. That is why we can treat very large random matrices (asymptotically) as free random random variables and employ the techniques of the free world.

We state the definition for asymptotic freeness from (Mingo and Speicher, 2017, Definition 1., Chapter 4). The authors differentiate between "asymptotic freeness" and "almost sure asymptotic freeness". For the sake of completeness and interest we state both of them as well.

I.1.56 Definition. *Let $(A_N)_{N\in\mathbb{N}}$ and $(B_N)_{N\in\mathbb{N}}$ be sequences of random matrices s.t. for each N the matrices A_N an B_N are defined on the probability space $(\Omega_N,\mathbb{P}_N)$, with the corresponding expectation $\mathbb{E}_N$.*

1. *A_N and B_N are asymptotically free if $A_n, B_N \in (\mathcal{A}_N, \mathbb{E}(tr(\cdot)))$, with $\mathcal{A}_n$ denoting the algebra generated by A_N and B_N, converge in distribution to elements A and B in a free probability space $(\mathcal{A},\varphi)$ s.t. A and B are free.*

2. *Consider the product space* $\Omega = \prod_{N\in\mathbb{N}} \Omega_N$ *and we have* $\mathbb{P} = \prod_{N\in\mathbb{N}} \mathbb{P}_N$ *the product measure of the* P_N *on the product space* Ω. *We call* A_N *and* B_N *almost surely asymptotically free, if there are free* A,B *in a non-commutative probability space* $(\mathcal{A},\varphi)$ *s.t. for almost all* $\omega \in \Omega$ *we have that* $A_N(\omega), B_N(\omega) \in (M_N(\mathbb{C}), tr(\cdot))$ *converge to the elements* $A,B \in (\mathcal{A},\varphi)$ *in distribution.*

We mention the specific limiting distributions for two prominent matrix ensembles, the Gaussian unitary ensemble (GUE) ensemble and the Wishart matrices.[27]

I.1.57 Definition. *A GUE matrix is* $N \times N$*-matrix* $A_N = (A_{ij})_{1\leq i,j\leq n}$ *with complex random variables* A_{ij} *such that* $A_{ji} = \overline{A_{ij}}$ *for* $i \neq j$ *and* $A_{ii} = \overline{A_i}$ *and the additional property that for all* A_{ij} *with* $i < j$ *we have*

$$A_{ij} = x_{ij} + \sqrt{-1}y_{ij},$$

with x_{ij}, y_{ij} *for* $1 \leq i < j \leq N$ *are independent real Gaussian with mean 0 and variance* $1/(2N)$ *each.*

I.1.58 Definition. *We refer to random matrices of the form* $\frac{1}{N}X^*X$ *where* X *is a* $N \times M$*-random matrix with independent Gaussian entries as Wishart matrices.*

I.1.59 Definition. *The law of the Marchenko-Pastur distribution (also referred to as free Poisson law) with support on* $[a,b]$ *for* $0 < c < \infty$ *is given by*

$$d\nu_c(x) = (1-c)\delta_0 + \frac{\sqrt{(b-x)(x-a)}}{2\pi x}dx,$$

where δ_0 *is the Dirac-Delta function on 0.*

It turns out that the limiting eigenvalue distribution of the GUE ensemble is given by the semicircular distribution and of the Wishart ensemble by the Marchenko-Pastur distribution, which can be seen as a free version of the Poisson distribution (cf. Mingo and Speicher (2017)). Since independent matrices from the GUE and the Wishart ensemble are asymptotically free we can derive the asymptotic eigenvalue distribution of their sums via a free convolution (see Mingo and Speicher (2017)), a free counterpart to the classical convolution of random variables. Due to their properties Wishart

[27] Note that the following definitions are standard but can be found e.g. in Mingo and Speicher (2017).

matrices are a suitable choice for the modeling of covariance matrices, a property that can be exploited via asymptotic freeness. For an interesting application we further mention Ryan and Debbah (2007) who use asymptotic freeness to denoise data via free deconvolution. Applying this procedure for covariance matrices of portfolios (as mentioned in e.g. Ryan (2008)) and modeling and forecasting e.g. eigenvalue distributions of explicit covariance (data) matrices by using asymptotic freeness might be an interesting and promising start for future research to apply the results in Fink et al. (2019) on real-world data. Consider in particular the asymptotic eigenvalue distributions (cf. Mingo and Speicher (2017)): We have in particular for $N \times N$-GUE matrices A_N that

$$\lim_{N\to\infty} \mathbb{E}(\mathrm{tr}(A_N^m)) = \begin{cases} \frac{1}{n+1}\binom{2n}{n}, & m = 2n \\ 0, & m \text{ odd} \end{cases}$$

and for $N \times M$-Wishart matrices A s.t. $\lim_{N,M\to\infty} \frac{M}{N} = c \in (0,\infty)$ we have that

$$\lim_{\substack{N,M\to\infty \\ M/N\to c}} \mathbb{E}(\mathrm{tr}(A^k)) = \sum_{\pi\in \mathrm{NC}(k)} c^{\#(\pi)},$$

where $\#(\pi)$ denotes the number of blocks of the partition π. As it is done in Mingo and Speicher (2017) we illustrate the averaged and the almost sure approximations for the GUE ensemble and the Wishart ensemble, or their empirical eigenvalue distribution, respectively, via Figures I.1 and I.2 and Figures I.3 and I.4, respectively. For the averaged version, the asymptotic freeness, we simulate 10k GUE matrices and plot the histograms of the eigenvalues for all 10k matrices for $N\times N$-matrices with $N = 3$, $N = 15$ and 40, together with its limiting distribution (see Figure I.1). Figure I.2 shows the approximation for the almost sure version of asymptotic freeness. For this we used only one realization for each plot for $N = 15$, $N = 450$ and $N = 4500$. The same illustration is done in Figures I.3 and I.4 for $N = 3$ and $M = 5$, for the first plot, $N = 6$ and $M = 10$ for the second, and $N = 60$ and $M = 100$ for the third. The almost sure version is shown for $N = 36$ and $M = 60$, and $N = 360$ and $M = 600$, and finally $N = 3600$ and $M = 6000$.

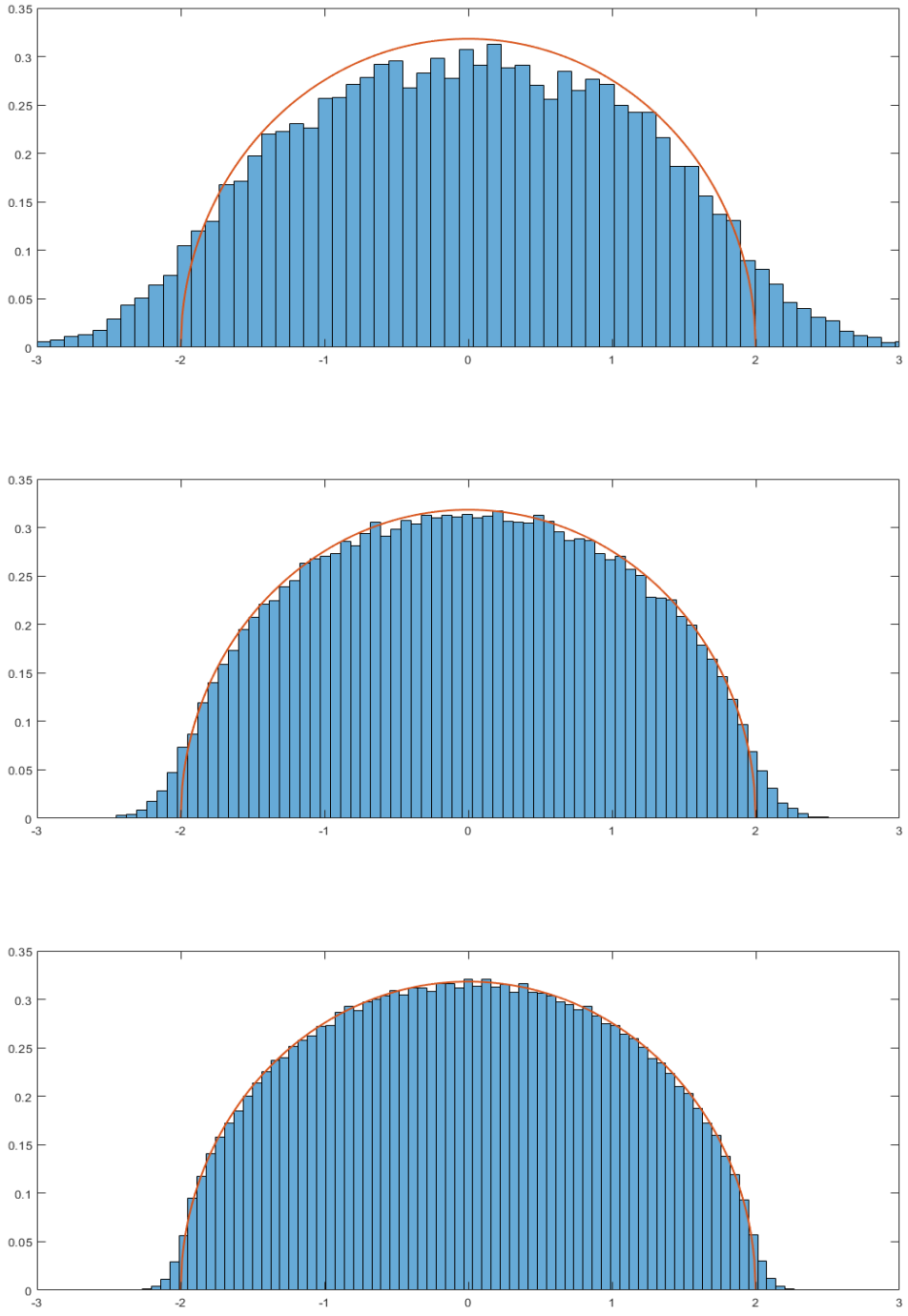


Figure I.1.: *Histograms for 10k simulations of $N \times N$-GUE matrices with $N = 3$, $N = 15$ and $N = 40$, respectively (from above to below) together with its limiting distribution*

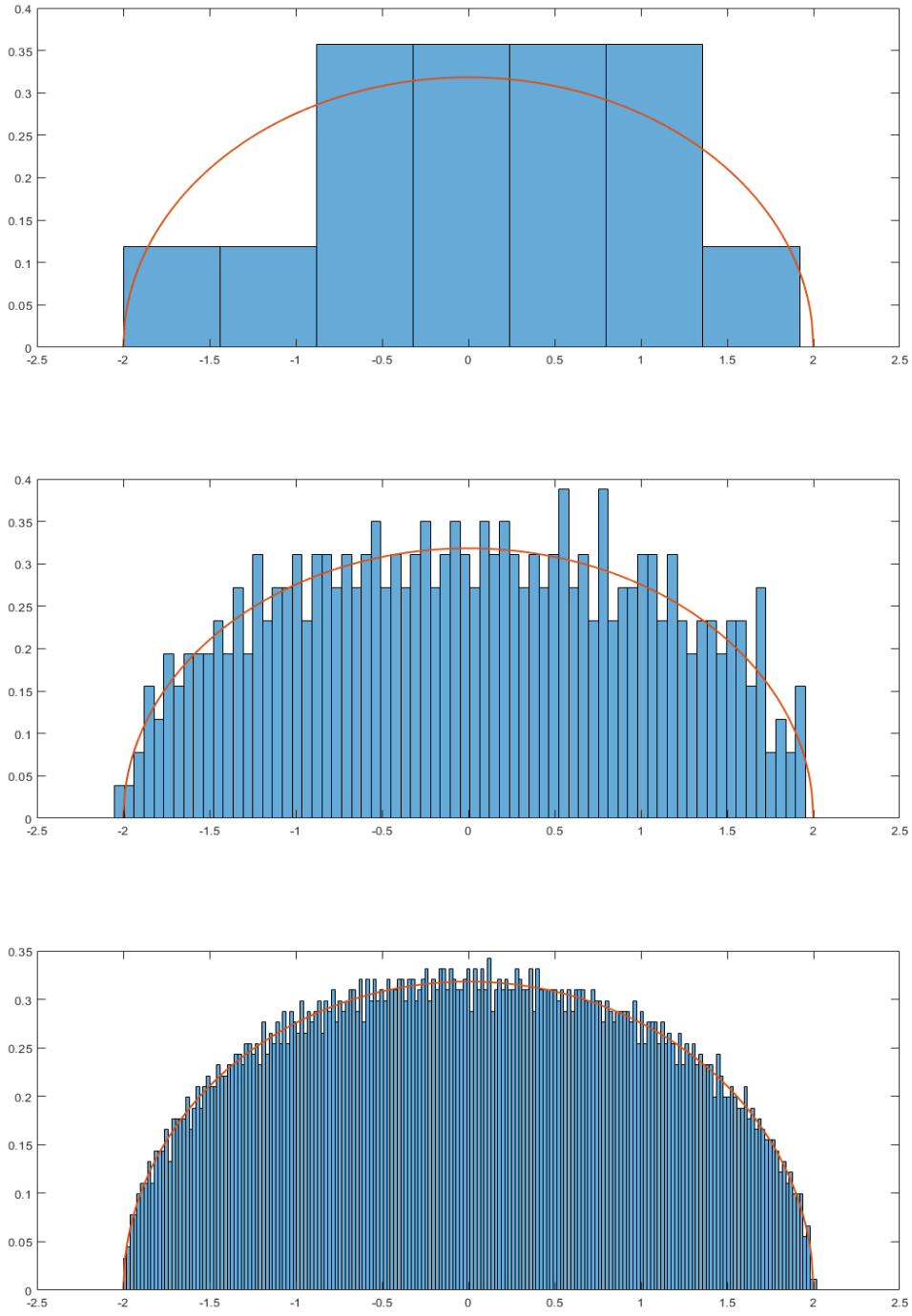


Figure I.2.: *Histograms of $N \times N$-GUE matrices with $N = 15$, $N = 450$ and $N = 4500$, (from above to below) together with its limiting distribution*

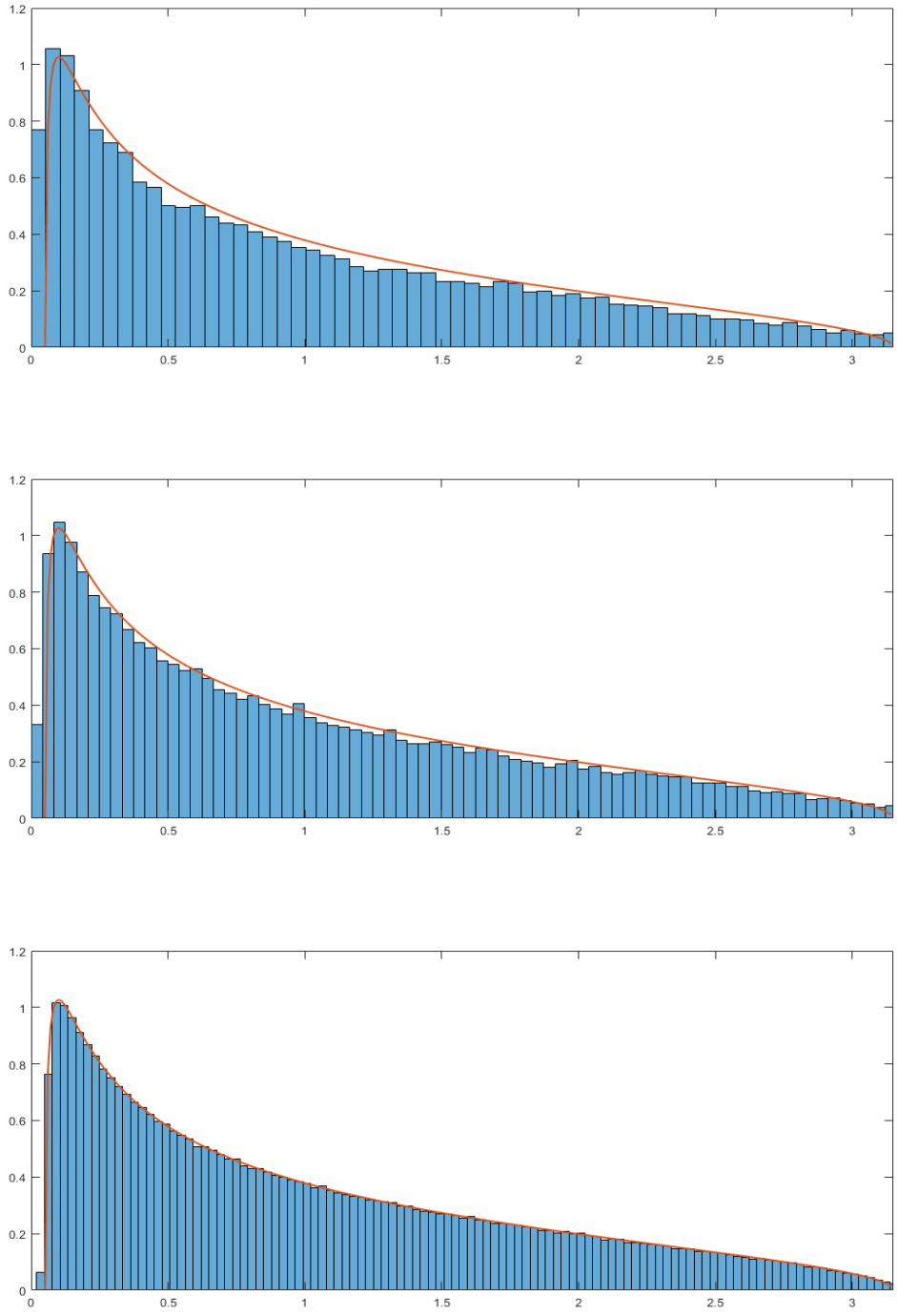


Figure I.3.: *Histograms for 10k simulations of $N \times M$-Wishart matrices with $N = 3$ and $M = 5$, and $N = 6$ and $M = 10$, and $N = 60$ and $M = 100$, respectively (from above to below) together with its limiting distribution*

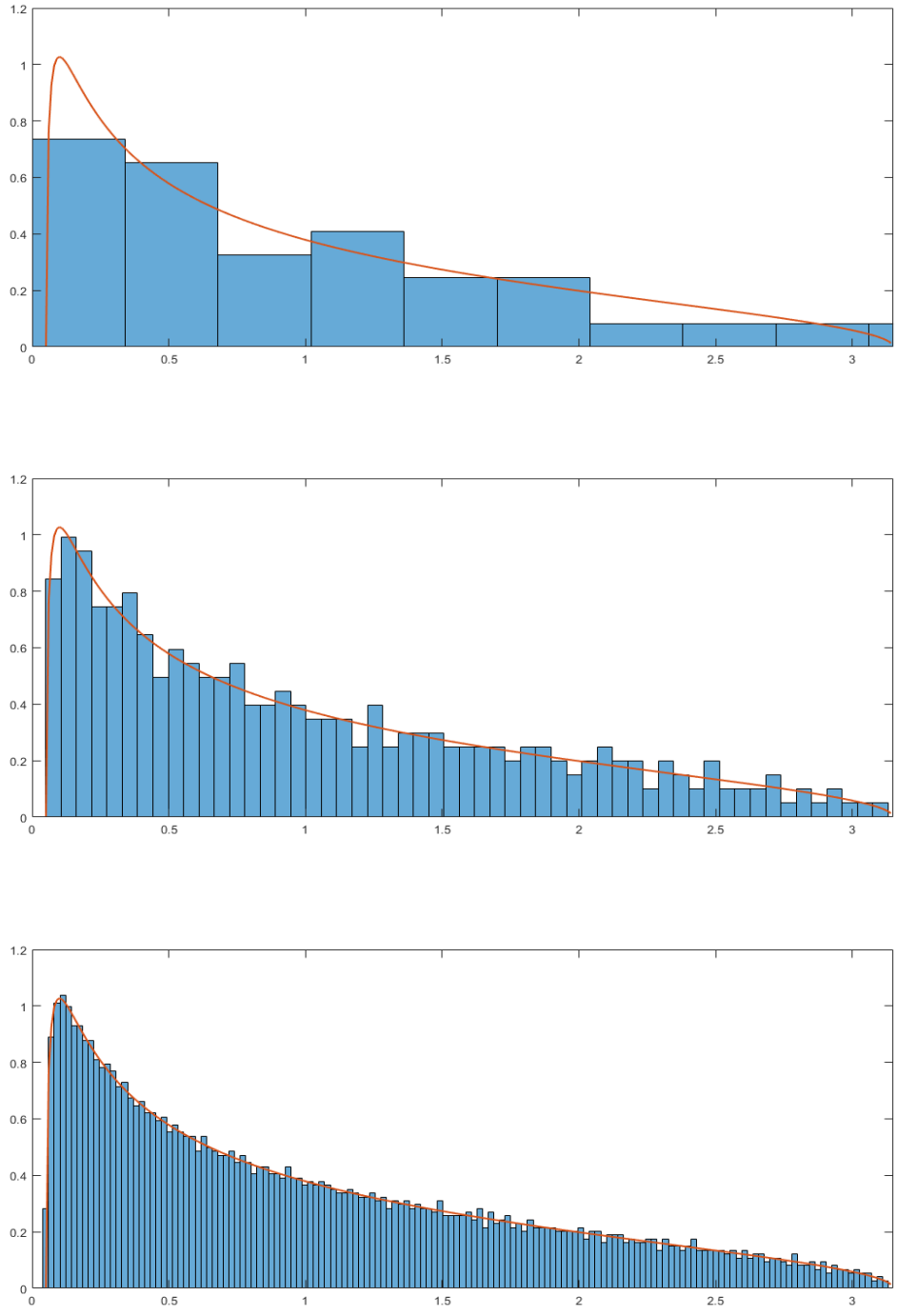


Figure I.4.: *Histograms of $N \times M$-Wishart matrices with $N = 36$ and $M = 60$, and $N = 360$ and $M = 600$, and $N = 3600$ and $M = 6000$, respectively (from above to below) together with its limiting distribution*

I.1.6. The Cauchy Transform

For the sake of completeness we will shortly mention the Cauchy transform, a valuable tool in free probability that can be used to extract the information of a free random variable's distribution by its moments. This can be found e.g. in Kargin (2011) or more detailed in Nica and Speicher (2006).

I.1.60 Definition. *The Cauchy transform of a real-valued probability measure μ of a self-adjoint $A \in \mathcal{A}$ is defined by*

$$g_A(z) = \int_{\mathbb{R}} (a - z)^{-1} d\mu(a).$$

Defining the resolvent of A by $G_A(z) = (A - z)^{-1}$ shows that

$$g_A = \varphi(G_A).$$

There is a way to recover the measure μ given g_A. This is done via the Stieltjes inversion formula, given by the following connection.

For a Borel set B s.t. for its boundary ∂B we have that $\mu(\partial B) = 0$ it holds that

$$\mu(B) = \pi^{-1} \lim_{\epsilon \searrow 0} \int_B \mathcal{I}m\, g_A(a + i\epsilon) \mathrm{d}a,$$

where $\mathcal{I}m$ denotes the imaginary part of a complex number and i denotes the imaginary unit $i^2 = -1$.

This allows the extraction of the information about the distribution of A via only knowing the expectation of its resolvent.

As we have seen there are many free counterparts to classical tools and theorems, like the central limit theorem, independence or convolution (see e.g. Mingo and Speicher (2017)). Therefore it should come as no surprise that there is also a free stochastic calculus with its own counterparts to stochastic differential equations and Brownian motion.

I.1.7. Free Stochastic Differential Equations

In this section we explain a few basic notions of the theory of free stochastic differential equations that are needed and partly are introduced in Fink et al. (2019) as well. The notion of stochastic differential equations necessitates free stochastic calculus,

which was first mentioned in Speicher (1990). Before proceeding to free stochastic differential equations we need a corresponding free counterpart to the driving process in the classical sense, namely the free Brownian motion.

As defined e.g. in Kargin (2011)[28] we introduce the free Brownian motion in the following.

I.1.61 Definition. *A free Brownian motion is a stochastic process* $(W(t))_{t\geq 0}$ *of elements in a von Neumann algebra fulfilling the following three properties*

1. $W(0) = 0$
2. *The increments* $W(t) - W(s)$ *are free from the algebra generated by* $W(\tau)$ *with* $\tau \leq s$ *for* $t > s$
3. *The increments* $W(t) - W(s)$ *follow a semicircular distribution with expectation* 0 *and variance* $t - s$

This definition shows that the semicircular distribution plays the same role the Gaussian distribution does in the classical case. The fact that the increments of the classical and the free Brownian motion have equal first and second moments respectively will play a crucial role in the existence proof in Fink et al. (2019). In particular, when the isometry between a stochastic process driven by a classical and and by a free Brownian motion is established, this property is essential.

We introduce a simplified version of free stochastic differential equations in the following to show their peculiarity. These are of the form

$$\mathrm{d}X(t) = a(t,X(t))\mathrm{d}t + b(t,X(t))\mathrm{d}W(t)c(t,X(t)),$$

where $(W(t))_{t\geq 0}$ denotes a free Brownian motion and a,b,c are operator-valued functions. Due to the non-commutative nature of these random variables we have that in general $b(t,X(t))\mathrm{d}W(t)c(t,X(t))$ and $b(t,X(t))c(t,X(t))\mathrm{d}W(t)$ are different. Therefore it is important from which side we multiply the Brownian motion. For the construction of the free stochastic integral we refer the reader to Biane and Speicher (1998).

We further mention the free Itô formula (cf. Biane and Speicher (1998)) in differential notation, as it is presented in Kargin (2011) and stated in Fink et al. (2019):

[28] This is not the first mentioning of the free Brownian motion but it is compactly presented there.

I.1.62 Lemma. *For operator-valued functions a_t, b_t, c_t, d_t and a free Brownian motion $(W(t))_{t\geq 0}$ we have*

$$\begin{aligned} a_t\,dt \cdot b_t\,dt &= a_t\,dt \cdot b_t\,dW(t)c_t = a_t\,dW(t)b_t \cdot c_t\,dt = 0 \\ a_t\,dW(t)b_t \cdot c_t\,dW(t)d_t &= \varphi(b_t c_t)a_t d_t\,dt. \end{aligned}$$

I.1.63 Remark. *Note that as Kargin (2011) does we will only consider functions that do not explicitly depend on time.*

Kargin (2011) showed the existence and uniqueness of a local solution to a certain class of free stochastic differential equations with locally operator Lipschitz coefficients, namely

$$\mathrm{d}X(t) = a(X(t))\mathrm{d}t + \sum_{k=1}^{m} b_k(X(t))\mathrm{d}W(t)c_k(X(t)),$$

or

$$X(t) = X_0 + \int_0^t a(X(s))\mathrm{d}s + \sum_{k=1}^{m}\int_0^t b_k(X(s))\mathrm{d}W(s)c_k(X(s)), \quad X(0) = X_0, \tag{I.5}$$

respectively, where $(W(s))_{s\geq 0}$ is a free Brownian motion.

We remind the reader of the locally operator Lipschitz property as it is presented e.g. in Kargin (2011):

I.1.64 Definition. *A function $f : \mathbb{R} \to \mathbb{C}$ is called locally operator Lipschitz if it is locally bounded, measurable and that for all $C > 0$ there exists a $K(C) > 0$ s.t. $\|f(X)-f(Y)\| \leq K(C)\|X-Y\|$, where X,Y are self-adjoint operators and $\|X\|, \|Y\| < C$.*

Note that we will sometimes refer to this as "Lipschitz" but it will be clear what is meant from the context.

We now state the local existence result in (Kargin, 2011, Theorem 3.1):

I.1.65 Theorem. *Suppose that a_i, b_i and c_i are locally operator Lipschitz functions and $\overline{X}$ is bounded in operator norm. Then, there exists a $T > 0$ and a family of operators $X(t)$ defined for all $t \in [0,T)$ and bounded in operator norm, such that $X_0 = \overline{X}$, and $X(t)$ is a unique solution of (I.5) for $t < T$.*

In Fink et al. (2019) we show the global existence of a positive (in the sense of having

a strictly positive spectrum) solution to the free stochastic differential equation

$$\mathrm{d}X(t) = (a - bX(t))\mathrm{d}t + \frac{\sigma}{2}\sqrt{X(t)}\mathrm{d}W(t) + \frac{\sigma}{2}\mathrm{d}W(t)\sqrt{X(t)},$$

with a,b,σ and $X(0) = X_0$ being elements in the von Neumann algebra with a strictly positive spectrum. Note that in the non-commutative world

$$\mathrm{d}X(t) = (a - bX(t))\mathrm{d}t + \sigma\sqrt{X(t)}\mathrm{d}W(t)$$

differs from the equation above. Nevertheless, the existence of a solution to the latter can be derived from the existence of a solution to the former via an argument involving self-adjointness and the property of the trace.

Apart from the usual operator norm $\|\cdot\|$ defined by

$$\|A\| = \sup_{\|v\|\leq 1} \|Av\|$$

for $A \in \mathcal{A}$ and $v \in \mathcal{H}$ we can define the p-norms $\|A\|_p$ via the trace φ by

$$\|A\| = \varphi(|A|^p)^{1/p} \text{ for } 1 \leq p < \infty.$$

In Fink et al. (2019) we will use an isometry argument for the 2-norm given for a self-adjoint A by $\|A\|_2 = \varphi(A^2)^{1/2}$. In particular the operator norm $\|\cdot\|$ and $\|\cdot\|_2$ will be of importance for the main proof in Fink et al. (2019).

Before giving a short overview of the existence proof in Fink et al. (2019) we mention that the first section in Fink et al. (2019) provides an introduction to the CIR equation and its connection to free probability theory.

I.1.8. Short Overview of the Existence Proof in "Free CIR Processes"

In this section we give a short guideline and overview to the existence proof in Fink et al. (2019).

As mentioned before, the CIR (stochastic differential) equation is of the form

$$\mathrm{d}x(t) = a - bx(t)\mathrm{d}t + \sigma\sqrt{x(t)}\mathrm{d}B(t)^{29},$$

[29] In some literature the drift term is of the form $a(b - x(t))$ resulting in the corresponding Feller

where $a,b,\sigma > 0$. As long as the Feller condition

$$2a \geq \sigma^2 \tag{I.6}$$

is satisfied we have for $x_0 > 0$ a strictly positive solution by Feller (1951). For a historical overview we refer to the introduction in Fink et al. (2019).

We start by introducing a supplementary stochastic differential equation.

I.1.66 Definition. *We call a stochastic differential equation of the form*

$$dV(t) = \left(a - \frac{\sigma^2}{2}\right)\frac{1}{2}V^{-1}(t) - \frac{1}{2}bV(t)\,dt + \frac{\sigma}{2}dD(t), \quad V(0) = V_0 > 0$$

"square-root process", where $(D(s))_{s\geq 0}$ is a classical or free Brownian motion and the process $(V(s))_{s\geq 0}$ can be scalar-valued, vector-valued or von Neumann algebra-valued.

We will work with this SDE and show that the existence of its (positive) solution gives us the existence of the (positive) solution to the CIR equation in the free setting due to the free Itô formula (cf. Biane and Speicher (1998)). Note that although $V^{-1}(t)$ does appear in the drift we get a "Brownian motion part" that does not depend on $V(t)$. This trade-off is preferred for the purposes in this proof.

The first part of the proof is to show the existence of a positive solution to the square-root process under Feller for a vector-valued process $(V(t))_{t\geq 0}$. This is done by employing point mass measures and the Feller condition.

The next step is to transform the vector-valued solution into the the general (non-commutative) von Neumann algebra-valued case by employing the functional calculus and the spectral theory which basically allows the applications of transformations of functions defined on their respective spectra.

Via the functional calculus we have an isometric homomorphism

$$T : B(\sigma(V_0)) \to \langle V_0, \mathrm{id}\rangle$$

between the function space of bounded, measurable functions on the spectrum of an operator $V_0 > 0$ denoted by $B(\sigma(V_0))$ and the von Neumann algebra generated by the identity operator and V_0, denoted by $\langle V_0, \mathrm{id}\rangle$. In particular this induces a mapping

condition $2ab \geq \sigma^2$. However, this formulation is equivalent to the one used here.

between the corresponding dual spaces s.t. we can define $\mathbb{E}_{\mathbb{P}_\varphi} = T^*(\varphi)$ given by

$$T^*\varphi(g) = \mathbb{E}_\varphi(g) = \int_{\sigma(V_0)} g \mathrm{d}\mathbb{P}_\varphi = \varphi(T(g)) \text{ for all } g \in B(\sigma(V_0)).$$

This allows us to conclude the existence of a global positive solution for the general (non-commutative) von Neumann algebra-valued case (driven by a classical Brownian motion). Note that the corresponding Feller condition in the operator-valued case is to be understood as the operator $2a - \sigma^2$ having a non-negative spectrum. When we refer to the Feller condition (I.6) in the operator-valued context that is what is meant.

The main part of the existence proof lies in changing the driving process from a classical Brownian motion to a free Brownian motion. This is done by showing that the solutions to the von Neumann algebra-valued SDE with the classical Brownian motion as the driving process (denoted by $\overline{V}$) and the one with the free Brownian motion as the driving process (denoted by V) are L_2-isometric (with respect to the corresponding expectations) on a common existence interval $[0,T]$ for a $T > 0$. A crucial part in this proof are the similar properties of the classical and the free Brownian motions. We shortly mention the following property from (Kargin, 2011, p. 829):

I.1.67 Remark. *Let A be free from C and from D s.t. $\varphi(A) = 0$. We have that*

$$\varphi(ACAD) = \varphi(A^2)\varphi(C)\varphi(D).$$

Choosing A as an increment of a free or classical Brownian motion, respectively is an essential part in establishing the isometry in the main proof in Fink et al. (2019).

The final part of the global existence proof is to extend the solution beyond T. This is done by showing the invertability of $V(T)$.

Having established the global existence of a positive solution to the free square-root process (driven by a free Brownian motion), applying the free pendant to the Itô formula (cf. Biane and Speicher (1998) or Anshelevich (2002)) gives us the existence of a unique positive solution to the free SDE

$$\mathrm{d}X(t) = (a - bX(t))\mathrm{d}t + \frac{\sigma}{2}\sqrt{X(t)}\mathrm{d}W(t) + \frac{\sigma}{2}\mathrm{d}W(t)\sqrt{X(t)},$$

for $t \in [0,\infty[$. We point out that in the non-commutative (free) world the equation

$$\mathrm{d}X(t) = (a - bX(t))\mathrm{d}t + \sigma\sqrt{X(t)}\mathrm{d}W(t),$$

is different to the one above but its solutions are isometric with respect to the trace φ and therefore we have proven its existence as well. We will define these two free stochastic processes:

I.1.68 Definition. *Let (I.6), we then call the unique positive solution to the free SDE*

$$dX(t) = (a - bX(t))\,dt + \frac{\sigma}{2}\sqrt{X(t)}\,dW(t) + \frac{\sigma}{2}\,dW(t)\sqrt{X(t)},$$

for $t \in [0,\infty[$, a non-classical free CIR process.

I.1.69 Definition. *Let (I.6), we then call the unique positive solution to the free SDE*

$$dX(t) = (a - bX(t))\,dt + \sigma\sqrt{X(t)}\,dW(t),$$

for $t \in [0,\infty[$, a (classical) free CIR process.

I.2. The Impact of Yield Curves on FX rates

In the next sections we will introduce the necessary theory for the ARMAFunX-GARCHFunX model. As mentioned in the introduction the basic idea is to increase the accuracy of risk predictions for the EURUSD exchange rate via the implementation of corresponding sovereign rate curves. These curves can be estimated via the use of functional principal component analysis. This makes it possible to dramatically decrease the size of data by extracting its main factors, which are still able to explain most of the data's variance. We will discuss the basics of time-series models allowing for the modeling of a conditional mean (ARMA) and a conditional variance (GARCH) and how the data is processed via the functional principal component analysis. In Fink et al. (2018) we compare the risk prediction performance of three models with respect to likelihood ratio tests.

I.2.1. Yield Curves, Data and Motivation

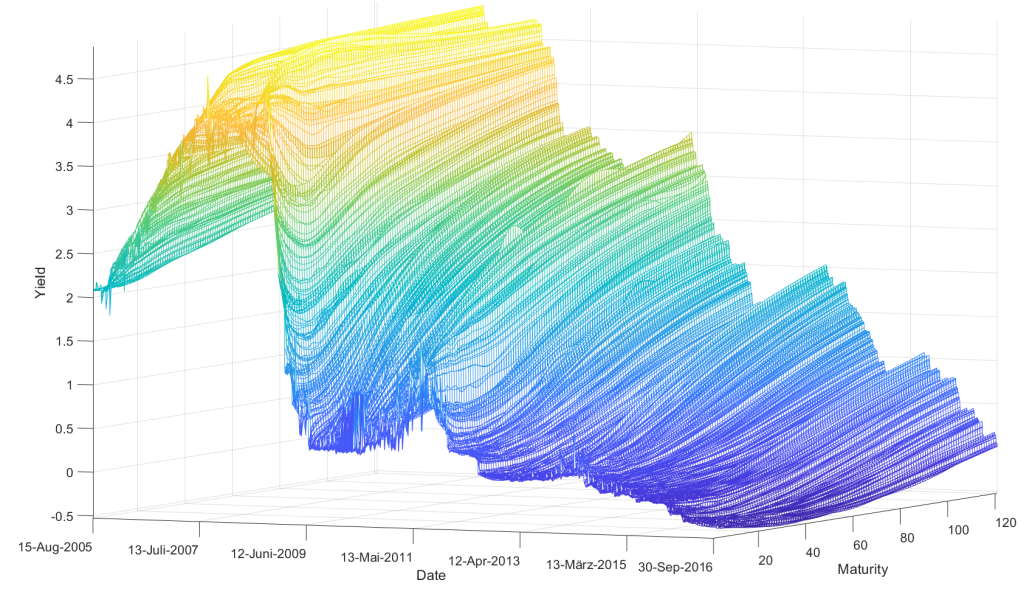
The data in question is the EURUSD foreign exchange (FX) rate y_t. To ensure stationarity of our data we work, as it is custom in the literature, with returns, in particular logarithmic returns. With that our time-series of returns is given by

$$r_t = \log\left(\frac{y_t}{y_{t-1}}\right).$$

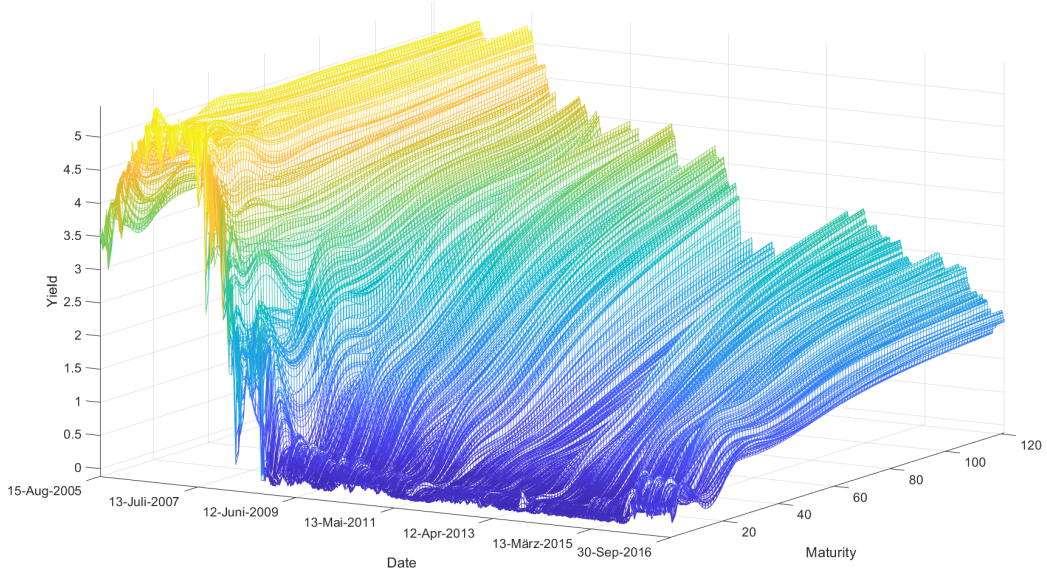
The chosen time-frame of daily data is given by 15 August 2005 to 20 September 2016. The exogenous data is given by the corresponding areas' yield curves, or the difference thereof, respectively. Yield curves can be viewed as mappings from a set of maturities $\{0,\dots,M\}$ to yields for each observation time t. Note that we allow for a maturity of zero. This is due to the fact that we incorporate overnight indexes, which can be interpreted as having a maturity of zero. The shortest maturity will be denoted by 0 and the longest by M. For each $t \in \mathcal{T} = [0,T]$ we have for the yield curve process $(x_t)_{t\in\mathcal{T}}$ that it is of the form

$$x_t : \{0,\dots,M\} \to \mathbb{R}, \quad m \mapsto x_t(m).$$

As a proxy for the risk-free rate for the US, we use government bonds and the federal funds rate for maturity zero. The corresponding proxy used for EUR in "The Impact of Sovereign Yield Curve Differentials on Value-at-Risk Forecasts for Foreign Exchange Rates" is given by the Overnight index Swaps on the EONIA (European Overnight Index Average) and the EONIA itself for maturity zero. Figure I.5 shows the surface areas of the corresponding yield curves. Note that we used cubic B-splines to interpolate the missing maturities s.t. we ended up with 1 (for the overnight rate) + 120 = 121 maturities in months. This is necessary for taking the differences and to ensure comparability. The nature of yield curve processes being maps for each point in time t suggests the usage of functional data analysis. The idea is to include the information of the relation of both yield curves, given by their difference, to use as exogenous data for the modeling of the corresponding exchange rate EURUSD. Apart from our ARMAFunX-GARCHFunX model, a model for the returns of the EURUSD FX rate, incorporating yield curve data, we will analyse a classical ARMAX-GARCHX model, that is an ARMA-GARCH model, with an exogenous term, to analyse the forecasting power of the 2-year yield difference of EUR and the US, which is assumed by practitioners to have significant forecasting power.



a) Yield curve surface for EUR



b) Yield curve surface for the US

Figure I.5.: *Yield curve surfaces for EUR and the US in the time frame 15 August 2005 to 20 September 2016 (daily) and maturities 0 to 120 months*

I.2.2. ARMAFunX-GARCHFunX

ARMA-GARCH models are the bread and butter of time-series analysis in econometrics. It can be seen as a combination of, the autoregressive moving average (ARMA) and the generalized autoregressive conditional heteroscedasticity model (GARCH) (cf. Bollerslev (1986))[30]. The ARMA model is used to model a time-series with a conditional mean μ_t, whereas the GARCH model is used to model a time-series with a conditional variance σ_t^2. The GARCH model is for example able to capture the phenomenon of volatility clustering persistent in financial data, which is one of the stylized facts of financial time-series. We will recall the corresponding definition shortly[31]:

I.2.1 Definition. *We call a process $(x_t)_{t\geq 0}$ an ARMA(p,q)-GARCH(r,s) process if*

$$\begin{aligned} x_t &= \alpha_0 + \sum_{k=1}^{p} \alpha_k x_{t-k} + \sum_{l=1}^{q} \beta_l \epsilon_{t-l} + \epsilon_t \\ \epsilon_t &= \sigma_t e_t \\ \sigma_t^2 &= \omega + \sum_{m=1}^{r} \delta_m \sigma_{t-m}^2 + \sum_{n=1}^{s} \gamma_n \epsilon_{t-n}^2, \end{aligned}$$

where $e_t \sim WN(0,1)$, a white noise process.

This is the basic setup of every ARMA-GARCH model. There exist several variations to this framework in the literature, among which the ARMA-logGARCH is of particular interest for our purposes. The extension to such models including exogenous variables, commonly referred to as ARMAX-GARCHX is straightforward. The exogenous information in question is the information given by the differences of the two yield curve processes for EUR and the US introduced earlier. The information of such a possibly infinite-dimensional process[32] x_t will be encoded by $\int_0^M \lambda(m) x_t dm$, where λ serves as a weight function[33]. This particular exogenous ARMA-GARCH model is defined by the following definition (see Fink et al. (2018)).

I.2.2 Definition. *Let x_t be drawn from a curve-valued exogenous process. Then r_t*

[30] This was anticipated by the ARCH model proposed in Engle (1982).
[31] For the sake of a simpler notation we will later work with $p = q = r = s = 1$.
[32] We can consider all maturities in the interval $[0,M] \subset \mathbb{R}$.
[33] We will denote another weight function in the GARCH part by ρ.

follows an ARMA(1,1)FunX-logGARCH(1,1)FunX process if

$$\begin{aligned} r_t &= \alpha_0 + \alpha r_{t-1} + \beta\epsilon_{t-1} + \epsilon_t + \int_0^M \lambda(m)x_{t-1}\,dm \\ \epsilon_t &= \sigma_t e_t \\ \log\sigma_t^2 &= \omega + \gamma\log\epsilon_{t-1}^2 + \delta\log\sigma_{t-1}^2 + \int_0^M \rho(m)x_{t-1}\,dm, \end{aligned}$$

with $e_t \sim WN(0,1)$, *a white noise process.*

The challenge herein lies in making this infinite-dimensional exogenous part usable in practice. The motivation is to lose as little information as possible, while decreasing the data size. Functional principal component analysis, or principal component analysis in general, allows us to extract the most important information present in the data, meaning the information explaining most of the variance, by a basis transformation s.t. these new "vectors of data" represented via the new basis, share no information due to the orthogonality of the basis vectors. This means that there is no redundant information present in the transformed data. The more principal components or the representation of the data by those, respectively, are added, the closer this representation is to the original data. Of course, the "approximation" by all principal components is exact. However, in many cases as few as 3 principal components suffice to explain most of the data's variance.

In the following we explain the basic idea of functional principal component analysis as it is done in Ramsay and Silverman (2005).

I.2.3. Functional Principal Component Analysis

We consider the exogenous process as an element of the Hilbert space $\mathcal{L}^2([0,M])$, that is the Hilbert space of on $[0,M]$ square-integrable real-valued functions. Naturally a Hilbert space is endowed with an inner product, which is in our setting given by the mapping

$$\langle\cdot,\cdot\rangle : \mathcal{L}^2([0,M]) \times \mathcal{L}^2([0,M]) \to \mathbb{R}, \quad (a,b) \mapsto \int_0^M a(s)b(s)\mathrm{d}s.$$

Assuming stationarity we have a mean function $\mu(s) = \mathbb{E}(x_t(s))$ and a covariance

operator given by

$$\mathcal{C} : \mathcal{L}^2([0,M]) \to \mathcal{L}^2([0,M]), \quad f \mapsto \int_0^M \mathrm{Cov}(r,s) f(r) \mathrm{d}r,$$

where $\mathrm{Cov}(r,s) = \mathrm{Covariance}(x_t(r), x_t(s))$, which both are independent of time.

This covariance can be written via the inner product as $\mathcal{C}(z) = \mathbb{E}[\langle x_t - \mu, z\rangle (x_t - \mu)]$, for a $z \in \mathcal{L}^2([0,M])$. We want to express our centered data $x_t - \mu$ via a new basis s.t. the variance of this projection has maximal (unconditional) variance. That is we need $\gamma_k \in \mathcal{L}^2([0,M])$ s.t. the variance of the scores

$$\xi_k = \langle \gamma_k, x_t - \mu \rangle = \int_0^m \gamma_k(s)(x_t(s) - \mu(s))$$

is maximal. These γ_k are exactly the eigenfunctions of $\mathcal{C}$.

The key to representing the data via this new basis of orthonormal eigenvectors $(\gamma_k)_k$ is the Karhunen-Loève theorem (see Karhunen (1947) and Loeve (1948)). This allows us to represent the centred data in the eigenbasis $(\gamma_k)_k$ via

$$x_t - \mu = \sum_{k=1}^{\infty} \xi_{k,t} \gamma_k.$$

We refer to the scores as the principal components since the vector $(\xi_{k,t})_k$ is the data-vector at time t represented in the new basis $(\gamma_k)_k$[34].

Note that we have

$$\mathbb{E}(\xi_{k,t}) = 0, \quad \mathbb{E}(\xi_{k,t}\xi_{l,t}) = \begin{cases} 0, & k \neq l \\ \lambda_k, & \text{else} \end{cases}.$$

This allows us to order the γ_k accordingly to their contribution in explaining the process' variation. Employing the properties of the orthonormal basis $(\gamma_k)_k$ we get the following representation of our exogenous part:

$$\int_0^M \lambda(m) x_t \mathrm{d}m \approx \sum_{k=1}^{K} \xi_{k,t} b_k + \text{const.},$$

where the b_k are real-valued coefficients that can be estimated by classical means. During the estimation process the constant const. will be "captured" by the models'

[34] Note that some literature refers to the eigenfunctions as principal components.

intercept. Note that the approximation is exact if we sum over all k. But as mentioned in most cases only a small number of the $\xi_{k,t}$ (ordered decreasingly by their explained variation) will suffice to explain most of the variance.

In the case of Fink et al. (2018) the exogenous part is given by the difference of the sovereign rate curve for EUR and for the US is given by

$$x_t(\cdot) = z_t^{\text{EUR}}(\cdot) - z_t^{\text{US}}(\cdot),$$

where $z_t^{\text{US}}(\cdot)$ is the sovereign rate curve for the US and $z_t^{\text{EUR}}(\cdot)$ is the sovereign rate curve for EUR.

The $\xi_{k,t}$ can be estimated via standard software for PCA (see Ramsay (2014)). This allows us to estimate the model defined in I.2.2 by an ARMAX-GARCHX type model:

$$\begin{aligned} r_t &= \alpha_0 + \alpha r_{t-1} + \beta\epsilon_{t-1} + \epsilon_t + \sum_{k=1}^{K} b_k \hat{\xi}_{k,t-1} \\ \epsilon_t &= \sigma_t e_t \\ \log\sigma_t^2 &= \omega + \gamma \log\epsilon_{t-1}^2 + \delta \log\sigma_{t-1}^2 + \sum_{l=1}^{L} c_l \hat{\xi}_{l,t-1}, \end{aligned}$$

where $\hat{\xi}_{k,t-1}$ are estimates of the scores obtained by numerical integration. Instead of estimating an infinite-dimensional weight function we now only have to estimate the b_k and c_l, respectively.

I.2.4. Value-at-risk and Backtesting

This setup allows us to forecast (1-day ahead) the value-at-risk (VaR) of the return process r_t, defined by

$$\text{VaR}_{t|t-1}(p) = \inf_x \{x \in \mathbb{R} | \mathbb{P}(r_t < x | \mathcal{F}_{t-1}) \geq p\}, \text{ for } 0 < p < 1,$$

where $\mathcal{F}_{t-1}$ denotes the information set up to $t-1$. We estimate this quantity via the forecasts for the conditional mean $\hat{\mu}_{t|t-1}$ and conditional standard deviation $\hat{\sigma}_{t|t-1}$ obtained by the ARMAFunX-GARCHFunX model:

$$\widehat{\text{VaR}}_{t|t-1}(p) = \hat{\mu}_{t|t-1} + \hat{\sigma}_{t|t-1}\Phi^{-1}(p),$$

where $\Phi^{-1}(p)$ is the p-quantile of the standard normal distribution. We employ a moving window approach for the 1-day ahead forecasts. The moving window approach for our case can shortly be explained by the following.

Let $[t_0, T]$ be the time-interval on which the return data is present. In order to forecast the VaR we take a subset of the above interval, namely $[\alpha + t_0, T - \beta]$ for $\alpha, \beta \geq 0$ and use the data in this interval to estimate the VaR at $T - \beta + 1$. For the estimate $\widehat{\text{VaR}}_{T-\beta+2}$ the interval $[\alpha + t_0 + 1, T - \beta + 1]$ is used and so forth, such that the length of the interval used for each forecast is constant. The choice of the interval's length depends among other factors on the length of the data set. The case for a k-day ahead forecast is analogous[35].

We want to point out that although we conducted model selection procedures on the whole data set (see Section 4.1. in Fink et al. (2018)), we use the full models, that is the models containing all parameters, for the VaR predictions. However, using information outside your current window implies knowledge of the future. In order to circumvent this issue one would need to select the model for each window separately before the next forecast. We restrained from doing that to keep computations simple and to allow for a better comparability of the models.

In Fink et al. (2018) we let three models compete against each other: the ARMAFunX-GARCHFunX, a classical ARMA-GARCH and an ARMAX-GARCHX employing the 2-year yield difference for the exogenous part.

The measure of choice for the evaluation of the models' performance is given by the likelihood ratio tests presented in Christoffersen (1998), the unconditional coverage test, the independence test and the conditional coverage test, where the latter can be seen as a combination of the former two. In the following we will draw heavily from Christoffersen (1998).

We define an indicator variable I_t by

$$I_{t|t-1} = \begin{cases} 1, & r_t < \widehat{\text{VaR}}_{t|t-1} \\ 0, & \text{else} \end{cases} .$$

An output of 1 is referred to as violation.

Consequently likelihood ratio tests evaluate the empirical number of violations against the expected number. The unconditional coverage test checks how close the ratio of the empirical and expected values is to 1, assuming independent violations. The inde-

[35] For the k-day ahead forecast it is assumed that $\beta \leq k$.

pendence test checks for dependence of the violations and the conditional coverage test is a combination of the former two tests. We define Ψ_T to be the information set (of the violations) up until T. The unconditional coverage test tests the null-hypothesis $\mathbb{E}(I_t) = p$ against $\mathbb{E}(I_t) \neq p$, assuming an i.i.d. and Bernoulli distributed sequence of violations I_t. That is the likelihood of a violation under the null-hypothesis is given by $\log(L(p;\Psi_T))$[36] and under the alternative to the null-hypothesis by $\log(L(\hat{\pi};\Psi_T))$ for a $T < t$. Let further N_1 be the number of violations and N_0 the number of times $I_t = 0$. Then $\hat{\pi}$, the empirical probability of a violation, is given by

$$\hat{\pi} = \frac{N_1}{N_1 + N_0}.$$

I.2.3 Definition. *We then define the unconditional coverage statistics via*

$$LR_{uc} = -2\log(L(p;\Psi_T)) + 2\log(L(\hat{\pi};\Psi_T)),$$

and have that

$$LR_{uc} \sim \chi^2_1.$$

The independence test accounts e.g. for volatility clusters by checking for the independence of the violations. For that we consider the transition probability matrix

$$\Pi_1 = \begin{pmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{pmatrix},$$

where π_{ij} denotes the transition probability of I_t from the state i in $t-1$ to the state j in t. We therefore get

$$L(\Pi_1;\Psi_T) = (1-\pi_{01})^{N_{00}}\pi_{01}^{N_{01}}(1-\pi_{11})^{N_{10}}\pi_{11}^{N_{11}},$$

with N_{ij} denoting the number of observations where the state i is followed by j.

Further we define

$$\hat{\Pi}_1 = \begin{pmatrix} \frac{N_{00}}{N_{00}+N_{01}} & \frac{N_{01}}{N_{00}+N_{01}} \\ \frac{N_{10}}{N_{10}+N_{11}} & \frac{N_{11}}{N_{10}+N_{11}} \end{pmatrix}.$$

The null-hypothesis for the independence test corresponds to the violations being independent, that is in our case, that the transition matrix is given by

[36] $L(p;\Psi_T)$ is given by $(1-p)^{N_0}p^{N_1}$.

$$\hat{\Pi}_1 = \begin{pmatrix} 1-\pi_2 & \pi_2 \\ 1-\pi_2 & \pi_2 \end{pmatrix}.$$

Under the null-hypothesis we have the likelihood

$$L(\Pi_2; \Psi_T) = (1-\pi_2)^{(N_{00}+N_{10})} \pi_2^{(N_{01}+N_{11})}$$

and its maximum likelihood estimator

$$\hat{\Pi}_2 = \hat{\pi}_2 = \frac{N_{01}+N_{11}}{(N_{00}+N_{10}+N_{01}+N_{11})}.$$

I.2.4 Definition. *We define the independence statistics via*

$$LR_{ind} = -2\log(L(\hat{\Pi}_2; \Psi_T)) + 2\log(L(\hat{\Pi}_1; \Psi_T)),$$

and have that

$$LR_{ind} \sim \chi_1^2.$$

Finally the combination of the former two is given by the joint test for coverage and independence.

I.2.5 Definition. *We define the conditional coverage statistics via*

$$LR_{cc} = -2\log(L(p; \Psi_T)) + 2\log(L(\hat{\Pi}_1; \Psi_T)),$$

and have that

$$LR_{cc} \overset{asymptotically}{\sim} \chi_2^2.$$

In Fink et al. (2018) we derived the above coverage test statistics for our three models as a measure of the forecasting performance. Apart from introducing and testing the ARMAFunX-GARCHFunX model, the second finding is that the 2-year yield difference does not seem to have a significant impact on the forecasting performance of the corresponding FX rate, as is derived by establishing confidence intervals for the corresponding parameter estimates. This is of particular interest since this spread's impact on the FX rate can be classified as a myth that is scattered in particular among practitioners.

Bibliography

Anshelevich, M. (2002). Itô formula for free stochastic integrals. *Journal of Functional Analysis*, 188(1):292–315.

Biane, P. and Speicher, R. (1998). Stochastic calculus with respect to free Brownian motion and analysis on Wigner space. *Probability Theory and Related Fields*, 112(3):373–409.

Blackadar, B. (2006). *Operator algebras: Theory of C^*-algebras and von Neumann algebras*, volume 122. Springer Science & Business Media.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(04):841–862.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics*, pages 173–182.

Fink, H., Fuest, A., and Port, H. (2018). The impact of sovereign yield curve differentials on value-at-risk forecasts for foreign exchange rates. *Risks*, 6(3):84.

Fink, H., Port, H., and Schlüchtermann, G. (2019). Free CIR processes. *arXiv preprint arXiv:1912.11714*.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343.

Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139.

Kargin, V. (2011). On free stochastic differential equations. *Journal of Theoretical Probability*, 24(3):821–848.

Karhunen, K. (1947). *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, volume 37. Sana.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 04(01):53–89.

Loeve, M. (1948). Functions aleatoires du second ordre. *Processus stochastique et mouvement Brownien*, pages 366–420.

Mingo, J. A. and Speicher, R. (2017). *Free probability and random matrices*, volume 4. Springer.

Murphy, G. J. (2014). *C*-algebras and operator theory*. Academic Press.

Neumann, J. v. (1930). Zur Algebra der Funktionaloperationen und Theorie der normalen Operatoren. *Mathematische Annalen*, 102(1):370–427.

Nica, A. and Speicher, R. (2006). *Lectures on the combinatorics of free probability*, volume 13. Cambridge University Press.

Nourdin, I. and Taqqu, M. S. (2014). Central and non-central limit theorems in a free probability setting. *Journal of Theoretical Probability*, 27(1):220–248.

Ramsay, J. (December, 3., 2014). FDA Toolbox. *http://www.psych.mcgill.ca/misc/fda/index.html*. (last accessed on 29 January 2020).

Ramsay, J. and Silverman, B. (2005). *Functional data analysis*. Springer, New York.

Ryan, Ø. (2008). Free probability: Basic concepts, tools, applications, and relations to other fields.

Ryan, Ø. and Debbah, M. (2007). Free deconvolution for signal processing applications. In *International Symposium on Information Theory*, pages 1846–1850. IEEE.

Sakai, S. (2012). *C*-algebras and W*-algebras*. Springer Science & Business Media.

Speicher, R. (1990). A new example of 'independence'and 'white noise'. *Probability theory and related fields*, 84(2):141–159.

Speicher, R. (2019). Lecture notes on "free probability theory". *arXiv preprint arXiv:1908.08125*.

Tao, T. (2012). *Topics in random matrix theory*, volume 132. American Mathematical Society.

Voiculescu, D. (1985). Symmetries of some reduced free product C^*-algebras. In *Operator algebras and their connections with topology and ergodic theory*, pages 556–588. Springer.

Voiculescu, D. (2002). Free entropy. *Bulletin of the London Mathematical Society*, 34(3):257–278.

Voiculescu, D., Dykema, K. J., and Nica, A. (1992). *Free random variables*. Number 1. American Mathematical Society.

Voiculescu, D., Stammeier, N., and Weber, M. (2016). *Free probability and operator algebras*.

Werner, D. (2018). *Funktionalanalysis*. Springer.

Whittle, P. (2012). *Probability via expectation*. Springer Science & Business Media.

Wick, G.-C. (1950). The evaluation of the collision matrix. *Physical review*, 80(2):268.

II. Free CIR Processes

Fink, H., Port, H., & Schlüchtermann, G. (2019). Free CIR Processes. arXiv preprint arXiv:1912.11714.

Author Contributions: Georg Schlüchtermann is responsible for the conceptualization and the main idea. Henry Port and Georg Schlüchtermann worked on the theorems, proofs and the methodology. They are also both responsible for writing and editing of the original draft. Holger Fink developed the story and the embedding in the context. Henry Port was responsible for the introduction. Proofreading was done by Holger Fink, Henry Port and Georg Schlüchtermann.

Free CIR processes

Holger Fink[1,a,b], Henry Port,[*,2,a,c] Georg Schlüchtermann[3,d,e]

30th December 2019

ABSTRACT

For stochastic processes of non-commuting random variables we formulate a Cox-Ingersoll-Ross (CIR) stochastic differential equation in the context of free probability theory which was introduced by Voicelescu. By transforming the classical CIR equation and the Feller condition, which ensures the existence of a positive solution, into the free setting (in the sense of having a strictly positive spectrum), we show the existence of a free CIR equation. The main challenge lies in the transition from a stochastic differential equation driven by a classical Brownian motion to a stochastic differential equation driven by the free analogue to the classical Brownian motion, the so-called free Brownian motion.

AUTHORS INFO

[a] Center for Quantitative Risk Analysis CEQURA, Ludwig-Maximilians-Universität München, Akademiestr. 1/I, 80799 Munich, Germany

[b] Department of Computer Science and Mathematics, Munich University of Applied Sciences, Munich, Germany

[c] Chair of Financial Econometrics, Institute of Statistics, Ludwig-Maximilians-Universität München, Akademiestr. 1/I, 80799 Munich, Germany

[d] Department of Mechanical, Automotive and Aeronautical Engineering, Munich University of Applied Sciences, Munich, Germany

[e] Faculty of Mathematics, Computer Science and Statistics, Ludwig-Maximilians-Universität München, Theresienstrasse 39/I, 80333 Munich, Germany

[1] holger.fink@hm.edu [2] henry.port@stat.uni-muenchen.de [3] georg.schluechtermann@hm.edu

[*] Corresponding author

KEYWORDS

free probability, CIR, volatility, stochastic differential equations, stochastic processes

Mathematics subject classification

60H10, 65C30, 46L54, 97M30

1 Introduction

The Black-Scholes (BS) framework is considered as one of the benchmarks for modeling the price process $(X(t))_{t\geq 0}$ of an underlying asset. Initially mentioned in Black and Scholes (1973), it is based on the stochastic differential equation (SDE)

$$\mathrm{d}X(t) = \mu X(t)\mathrm{d}t + \sigma X(t)\mathrm{d}B_1(t),$$

for $t \in [0,\infty[$, where $\mu \in \mathbb{R}$, $\sigma > 0$ and $(B_1(t))_{t\geq 0}$ denotes a Brownian motion.

One major drawback of this model is the fact that it does not account for certain common properties of financial data, known as stylized facts, in particular, volatility clustering and the so-called leverage effect. The first term refers to the fact that volatility exhibits a highly autocorrelated structure and the second to the negative correlation between volatility and returns frequently found in financial data (cf. Pagan (1996); Mandelbrot (1997); Cont (2001)).

A possible solution is to model the variance separately as a time-dependent stochastic process that accounts for these effects. A desired property for such a model is to ensure that the variance returns to an average value in finite time, which is referred to as mean-reversion. Therefore, the Vasicek model (cf. Vasicek (1977)), which was originally developed to model the evolution of interest rates, appears as a promising candidate. That means the variance process $(s(t)^2)_{t\geq 0}$ of $(X(t))_{t\geq 0}$ is modeled as the solution to the SDE

$$\mathrm{d}s(t)^2 = a(b - s(t)^2)\mathrm{d}t + \sigma \mathrm{d}B_2(t),$$

for $t \in [0,\infty[$, where $b > 0$ is referred to as the long term mean level, $a > 0$ is a measure of the time it takes $s(t)^2$ to return to b and $\sigma > 0$ signifies the influence of the random shocks and $(B_2(t))_{t\geq 0}$ is another Brownian motion, correlating with $(B_1(t))_{t\geq 0}$ via

$$\rho \mathrm{d}t = \mathrm{d}B_1(t)\mathrm{d}B_2(t),$$

for $t \in [0,\infty[$, where $-1 \leq \rho \leq 1$. This setup appears reasonable and sufficient at first glance but unfortunately it can not guarantee positivity of $(s(t)^2)_{t\geq 0}$, which is of course a necessity for a variance model.

However, in 1993 Heston (1993) improved upon the BS framework by addressing the aforementioned issues, allowing the variance to be modeled by a separate SDE, namely the Cox, Ingersoll and Ross (CIR) process as developed by Cox et al. (1985), via the SDE

$$\mathrm{d}s(t)^2 = (a - bs(t)^2)\mathrm{d}t + \sigma\sqrt{s(t)^2}\mathrm{d}B_2(t),$$

for $t \in [0,\infty[$, which allows for mean-reversion but also addresses the issue of negativity: Feller (1951) has shown that as long as the so-called Feller condition

$$2a \geq \sigma^2 \tag{1}$$

holds there exists a positive unique solution.

Since then, many researchers have been developing and generalizing this framework for volatility modeling. Motivated by the limitation of not being able to recreate the term structure of the analyzed asset, or interest rate, in question (cf. Hull and White (1990); Yang (2005); Keller-Ressel and Steiner (2008)), the extended CIR equation was introduced in Maghsoodi (1996) allowing time dependency for all parameters. In order to model potential long memory effects in volatility which were observed by some researchers (cf. Baillie et al. (1996); Bollerslev and Mikkelsen (1996)), Comte and Renault (1998) implemented a fractional Brownian motion as the driving process. This setup was adapted by Comte et al. (2012) who proposed a fractional Heston model to explain the mystery of the steepness present in volatility smiles of long term options. Schlüchtermann and Yang (2016) permitted a dynamic term of the form σx^q, where $q \geq 0$, in the fractional setup and showed the existence of a positive solution of these so called generalized fractional CLKS-type equations by imposing a Feller-like condition on the coefficients, for pathwise forward integrals as well as for integrals in the Wick sense. Fink and Schlüchtermann (2018) expanded the fractional CIR equation even further to the Mandelbrot-Van Ness fractional Lévy process-driven case with time-dependent coefficients. Yet another modification is the implementation of a Hurst index of $H < \frac{1}{2}$. The resulting so called rough Heston models give a good mixture of a decent fit to historical data and implied volatility without touching upon the curse of dimensions (cf. Gatheral et al. (2018); El Euch et al. (2018); Jaber and Euch (2019); El Euch and Rosenbaum (2019); Bayer et al. (2019)).

The task of modeling prices for two or more assets on the other hand poses additional challenges. Taking covariances into account necessitates a joint model. A promising candidate is given by Wishart

autoregression processes, which were introduced in Bru (1991). Since Wishart processes do not need additional constraints to ensure positive definiteness almost surely (cf. Gouriéroux (2006)), they are suited especially well for the role of a matrix-valued CIR process.

Nevertheless, with an increasing number of assets the complexity of volatility models adequately describing these systems increases rapidly, demanding the usage of more variables and therefore the curse of dimension (cf. Gourieroux and Sufana (2010)) threatens a feasible application.

Some matrix ensembles such as the GUE (Gaussian unitary ensemble) and the Wishart random matrices (these are of the form $\frac{1}{N}XX^*$, where X is $N \times M$ random matrix with independent Gaussian entries and X^* denotes its adjoint) behave like so-called free random variables in (eigenvalue) distribution, when their size gets very large, which is referred to as "asymptotic freeness" (cf. Mingo and Speicher (2017)). The limiting eigenvalue distribution of the former is given by the semicircular distribution and for the latter by the Marchenko-Pastur distribution (also known as "free Poisson distribution").

Since these matrix ensembles behave like free random variables in high dimensions we may employ the tools of free probability theory, which allow for concepts like convolution and a pendant to the central-limit theorem (cf. Nourdin and Taqqu (2014)) to adequately work with random matrices and operators in the probabilistic context. Free probability, developed by Dan Voiculescu (around 1986 while trying to solve an isomorphism problem about free groups (cf. "Background and outlook" in Voiculescu et al. (2016)), is a fitting framework to analyse the distribution of non-commutative random objects in the desired generality. It can be considered as a non-commutative analogue to classical probability theory together with the notion of freeness that allows the computation of the joint distribution of non-commutative random variables. In the context of random matrices this distribution can be understood as the spectral distribution. In particular, portfolio theory benefits immensely from properties of free random variables, since it allows for cleaning of empirical correlation matrices, which basically are Wishart matrices, to derive the distribution of a "true" noise-free correlation matrix via so-called "free deconvolution" (cf. Ryan and Debbah (2007); Ryan (2008); Bouchaud and Potters (2009)). Since free probability is a fitting framework for dealing with high- or even infinite-dimensional random-matrices we can employ this setting to adequately describe a volatility-model for high-dimensional portfolios, namely via a potential free CIR model. This idea necessitates a free stochastic calculus.

Free stochastic calculus first appeared in Speicher (1990). This theory was further developed by

Kümmerer and Speicher (1992), Biane (1997) and Biane and Speicher (1998), where among other ground laying definitions the notion of free stochastic processes and a free Brownian motion were introduced. For an in-depth study of the free Itô integral, we refer to Anshelevich (2002). This framework allows for defining the notion of free SDEs. In particular free stochastic processes form a vivid research area (cf. Biane (1998); Biane and Speicher (2001); Barndorff-Nielsen et al. (2002); Fan (2006); Gao (2006); Gao et al. (2008); An and Gao (2015)).

Some processes, such as the item of interest of this paper, the CIR process, arise naturally as the solution of SDEs. A special class of free SDEs are studied in depth in Kargin (2011), where the author provided proofs for existence and uniqueness of a local solution of free SDEs of the form

$$\mathrm{d}X(t) = a(X(t))\mathrm{d}t + \sum_{k=1}^{m} b_k(X(t))\mathrm{d}W(t)c_k(X(t)),$$

for $t \in [0,\infty[$, where $(W(t))_{t\geq 0}$ is a free Brownian motion and a, b_k, c_k are locally operator Lipschitz functions. To recall: We call a function $f : \mathbb{R} \to \mathbb{C}$ locally operator Lipschitz, if it is locally bounded, measurable and we also have that for all $C > 0$, there is a constant $K(C) > 0$, with the property that $\|f(X) - f(Y)\| \leq K(C)\|X - Y\|$, provided that X and Y both are self-adjoint operators such that $\|X\|, \|Y\| < C$, with $\|\cdot\|$ denoting the operator norm. Unfortunately due to the fact that the coefficient functions do not fulfill the necessary locally operator Lipschitz condition, the local existence theorem does not imply the existence of a solution to the classical free CIR equation of the form

$$\mathrm{d}X(t) = (a - bX(t))\,\mathrm{d}t + \sigma\sqrt{X(t)}\mathrm{d}W(t), \quad X(0) = X_0 \in \mathcal{A}_+,$$

for $t \in [0,\infty[$ and $a,b,\sigma \in \mathcal{A}_+$. We denote by $\mathcal{A}_+$ the self-adjoint elements of the von Neumann algebra $\mathcal{A}$ with a strictly positive spectrum. Those are representable by the square of a self-adjoint operator which gives meaning to the expression $\sqrt{A}$. Therefore a solution to this equation needs to be bounded below by zero (in the sense of having a positive spectrum), just as in the scalar-valued case. We introduce and show the existence of the free CIR equation and derive an equivalent condition to the classical Feller condition. Note that for possibly time-dependent operators $a(t), b(t), \sigma(t)^2$ the Feller condition is to be understood in the sense of $2a(t) - \sigma(t)^2$ having a non-negative spectrum. When refering to the Feller condition (1) in the context of operators that is how it is meant to be interpreted.

This paper is structured as follows: In Section 2 we cover the necessary basics of free probability theory. Section 3 proves our main theorems on the existence and uniqueness of positive solutions for

the free CIR setup.

2 Free Probability and Free Stochastic Calculus

In the following we give a short introduction to the theory of free probability and free SDEs. For the theory of operator theory and in particular von Neumann algebras we refer the reader to the multivolume works "Theory of operator algebras" by Takesaki (cf. Takesaki (2002); Takesaki (2013); Takesaki (2003)) and Murphy (2014). For an introduction to free probability we refer to Voiculescu et al. (1992), Nica and Speicher (2006), Voiculescu et al. (2016) and Mingo and Speicher (2017). In the following we will draw heavily on Biane and Speicher (1998) and Kargin (2011).

2.1 Definition. *Let $\mathcal{A}$ be a type I von Neumann algebra and φ a faithful, normal and unital trace on $\mathcal{A}$. We call the tupel $(\mathcal{A}, \varphi)$ a non-commutative probability space and the self-adjoint elements of $X \in \mathcal{A}$ are called (free) random variables. In this paper we only consider self-adjoint random variables.*

The trace φ induces the p-norms $\|\cdot\|_p$ by

$$\|X\|_p = \varphi\left(|X|^p\right)^{\frac{1}{p}}, \quad 1 \leq p \leq \infty,$$

where $\|\cdot\| := \|\cdot\|_\infty$ is the operator norm. For more information on non-commutative integration see e.g. Fack and Kosaki (1986) and Terp (1981).

We proceed by defining the notion of free independence or freeness, as introduced by Voiculescu.

2.2 Definition. *Let $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be subalgebras of $\mathcal{A}$. We call $\mathcal{A}_1, \ldots, \mathcal{A}_n$ free if*

$$\varphi\left(A_{i(1)} \ldots A_{i(m)}\right) = 0$$

whenever

$$\varphi\left(A_{i(s)}\right) = 0 \text{ and } i(s+1) \neq i(s) \quad \text{for each } s.$$

We call random variables free if the algebras they generate are free.

In order to introduce free SDEs, a "free" notion of Brownian motion is necessary.

2.3 Definition. *A free Brownian motion is a stochastic process $(W(t))_{t\geq 0}$ of elements in a von Neumann algebra with the the three properties: $W(0) = 0$, the increments $W(t) - W(s)$ are free from*

$\mathcal{W}_s = \langle W(\tau) \mid \tau \leq s \rangle$ *for* $t > s$, *and* $W(t) - W(s)$ *follows a semicircular distribution with expectation* $\varphi(W(t) - W(s)) = 0$ *and variance* $\varphi\left((W(t) - W(s))^2\right) = t - s$.

We proceed by introducing an integral of the form

$$\mathcal{I} = \int_0^T a(X(s))\,\mathrm{d}W(s)b(X(s)),$$

where $a(X(s))$ and $b(X(s))$ are operator-valued functions of $X(s)$ and $T > 0$. Certain properties on a resp. b will be specified later. For a detailed construction of this free stochastic integral, we refer to Anshelevich (2002) or Biane and Speicher (1998). We will describe the general construction shortly as it is done in Kargin (2011).

Given an interval $[0,T]$ and $s \in [0,T]$, we let $a_s, b_s \in \mathcal{W}_s$ and assume that $s \mapsto a_s$ and $s \mapsto b_s$ are continuous maps w.r.t. to $\|\cdot\|$. Consider further $s_0, \ldots, s_n, \tau_1, \ldots, \tau_n \in \mathbb{R}$ with $0 = s_0 \leq s_1 \leq \cdots \leq s_n = T$ and $0 \leq \tau_k \leq s_{k-1}$. We denote the collection of all s_i and τ_j by $\Delta := \{s_i, \tau_j \mid 0 \leq i \leq n\ ,\ 1 \leq j \leq n\}$. Consider the expression

$$\mathcal{I}(\Delta) = \sum_{i=1}^{n} a_{\tau_i}(W(s_i) - W(s_{i-1}))\, b_{\tau_i}.$$

For $d(\Delta) = \max_{1 \leq k \leq n}(s_k - \tau_k)$ we get that

$$\lim_{d(\Delta) \to 0} \mathcal{I}(\Delta) = \mathcal{I},$$

where the limit is meant w.r.t. $\|\cdot\|$ and is independent of the choice of s_i and τ_i. $\mathcal{I}$ is called the "free stochastic integral". The convergence relies heavily on the so called free Burkholder-Gundy inequality (see (Biane and Speicher, 1998, Theorem 3.2.1)). In the following lemma we state the free analogues of the Itô formula in an abbreviated form as we will use later on. For a detailed discussion we refer to Biane and Speicher (1998).

2.4 Lemma (Free Itô)**.** *Let* a_t, b_t, c_t, d_t *be operator-valued functions and* $(W(t))_{t \geq 0}$ *as above. Then*

$$\begin{aligned} a_t\,dt \cdot b_t\,dt &= a_t\,dt \cdot b_t\,dW(t)c_t = a_t\,dW(t)b_t \cdot c_t\,dt = 0 \\ a_t\,dW(t)b_t \cdot c_t\,dW(t)d_t &= \varphi(b_t c_t)\, a_t d_t\,dt \end{aligned}$$

From here on we will restrict ourselves to coefficients that do not depend on time explicitly (as it is done in the local existence results in Kargin (2011) as well) and will denote them by e.g. $a(X(t))$.

3 Free CIR equations

As mentioned in the introduction it is known in the (commutative) scalar-valued case that the Feller condition (1) ensures for the CIR equation

$$\mathrm{d}x(t) = (a - bx(t))\,\mathrm{d}t + \sigma\sqrt{x(t)}\mathrm{d}B(t), \quad x(0) = x_0 > 0,$$

for $t \in [0,\infty[$, and $a,b,\sigma > 0$ a global positive solution, where $(B(t))_{t\geq 0}$ is a (classical) Brownian motion.

Therefore it is natural to ask, if the above Feller condition guarantees global existence of a positive solution for a free SDE

$$\mathrm{d}X(t) = (a - bX(t))\,\mathrm{d}t + \sigma\sqrt{X(t)}\mathrm{d}W(t), \quad X(0) = X_0 \in \mathcal{A}_+, \tag{2}$$

for $t \in [0,\infty[$, and $a,b,\sigma \in \mathcal{A}_+$, where $(W(t))_{t\geq 0}$ is a free Brownian motion. From here on we will refer to the (classical) Brownian motion by $(B(t))_{t\geq 0}$ and to the free Brownian motion by $(W(t))_{t\geq 0}$.

Due to the non-commutativity of $X(t)$ the free SDE, which we call (non-classical) free CIR equation,

$$\mathrm{d}X(t) = (a - bX(t))\,\mathrm{d}t + \frac{\sigma}{2}\sqrt{X(t)}\mathrm{d}W(t) + \frac{\sigma}{2}\mathrm{d}W(t)\sqrt{X(t)}, \quad X(0) = X_0 \in \mathcal{A}_+, \tag{3}$$

for $t \in [0,\infty[$, and $a,b,\sigma \in \mathcal{A}_+$, may differ from the classical free CIR equation (2). But, since all elements involved are self-adjoint, an easy argument shows that the traces of the solutions to (2) and (3) coincide and therefore it is enough to show the existence of the solution to the latter.

We start showing existence and uniqueness of these free SDEs by introducing a SDE with a simple additive Brownian motion term of the form

$$\mathrm{d}V(t) = \left(a - \frac{\sigma^2}{2}\right)\frac{1}{2}V^{-1}(t) - \frac{1}{2}bV(t)\mathrm{d}t + \frac{\sigma}{2}\mathrm{d}B(t), \quad V(0) = V_0 > 0,$$

for $t \in [0,\infty[$, and $a,b,\sigma > 0$, which we will refer to as square-root process (the reason for that will become clear later), and by step by step transforming it into a free SDE (driven by a free Brownian motion). Since the existence of a positive solution to a scalar-valued SDE of the form above and the classical CIR equation are equivalent by the classical Itô lemma we know that such an equation

has a positive solution as long as the Feller condition is satisfied. We will transform this connection first into the setting of (commutative) function spaces (see Theorem 3.3), followed up by the general (non-commutative) von Neumann algebra-valued case (see Theorem 3.5). Note that the driving process is still a classical Brownian motion. The final and most elaborate part (Theorem 3.9) consists of changing the driving process to a free Brownian motion. We will do this by showing that the solutions to the von Neumann algebra-valued SDE driven by a classical and the one driven by a free Brownian motion are L_2-isometric for $t \geq 0$. Finally the free Itô lemma gives us the existence of a global solution to our free CIR equation (under Feller).

Given a positive V_0 we can select a special probability space to transfer the SDE into a usual vector-valued SDE. Using functional calculus resp. the spectral theorem we have an isometric homomorphism

$$T : B(\sigma(V_0)) \longrightarrow \langle V_0, \mathrm{id} \rangle ,$$

where $\langle V_0,\mathrm{id}\rangle$ is the von Neumann algebra generated by V_0 (and the identity) and $B(\sigma(V_0))$ is the function space of bounded, measurable functions on $\sigma(V_0)$, the spectrum of V_0. If φ is a unital, faithful trace, then consider $\mathbb{E}_{\mathbb{P}_\varphi} = T^*(\varphi)$ with the identity

$$\mathbb{E}_{\mathbb{P}_\varphi}(g) = \int_{\sigma(V_0)} g \mathrm{d}\mathbb{P}_\varphi = \varphi(T(g)) \text{ for all } g \in B(\sigma(V_0)).$$

Next, for dealing with operator-valued coefficients, we introduce the (free) conditional expectation $\mathbb{E}_\mathcal{B}$ by the following lemma.

3.1 Lemma. *Let $\mathcal{B} \subset \mathcal{A}$ be a von Neumann subalgebra. Then there exists a conditional expectation*

$$\mathbb{E}_\mathcal{B} : \mathcal{A} \longrightarrow \mathcal{B}$$

s.t.

$$\varphi(ab) = \varphi(\mathbb{E}_\mathcal{B}(a)b) \quad \textit{for all } a \in \mathcal{A}, b \in \mathcal{B}.$$

(cf. Biane and Speicher (2001))

We fix a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$ and start with a vector-valued version for the existence of a local solution of a classical SDE. For this let $(B(t))_{t\geq 0}$ be a classical Brownian motion on $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$. First, we state a result which seems folklore, but is not cited explicitly.

3.2 Theorem. *Let E be a Banach space. Let a and b_i for $i = 1,\dots,m$ be continuous and locally Lipschitz. For $V_0 \in E$ there exists a $T > 0$ and a unique continuous solution $V : [0,T[\to E$, such that $V(0) = V_0$, $V \in C([0,T[, E)$ and for all $t \in [0,T[$*

$$V(t) = V_0 + \int_0^t a(V(s))ds + \sum_{i=1}^m \int_0^t b_i(V(s))dB(s).$$

Proof. The proof follows a standard argument via Picard approximations and the Banach fixed-point theorem. Applying the Hahn-Banach theorem we use a $x^* \in X^*$ in the dual of X to transform our approximations into the real setting in order to make use of the (classical) Burkholder-Gundy inequality. The details are analogous to the proof of (Kargin, 2011, Theorem 3.1.). □

Having established the existence of a local solution we will show that the classical Feller condition ensures a positive solution in the case of the commutative von Neumann algebra $C(K)$ of continuous functions of a compact Hausdorff space K. We further denote the positive cone by

$$C(K)_+ = \{f \in C(K) \mid f(u) > 0, \text{ for all } u \in K\}.$$

We note that using the Itô formula, under the Feller condition the square-root process $V(t)$ enjoys a global solution provided the initial condition is positive. We state a vector-valued extension.

3.3 Theorem. *Let $K \subset]0,\infty[$ be compact. Let $a,b,\sigma : [0,\infty[\to C(K)_+$ be continuous, such that (1) holds and let $\hat{V}_0 \in C(K)_+$. Then the SDE*

$$d\hat{V}(t) = \left(\left(a(t) - \frac{\sigma^2(t)}{2}\right)\frac{1}{2}\hat{V}^{-1}(t) - \frac{b(t)}{2}\hat{V}(t)\right)dt + \frac{\sigma(t)}{2}\mathbf{1}_K\, dB(t), \quad \hat{V}(0) = \hat{V}_0,$$

for $t \in [0,\infty[$, has a global solution $V \in C([0,\infty[, L_2(\mathbb{P}_\varphi, C(K)_+))$.

Proof. Consider a classical Brownian motion $(B(t))_{t\geq 0}$ on the probability space $(\Omega,\mathcal{F},\mathbb{P}_\varphi)$. Then using point mass measures δ_k for $k \in K$ and the classical Feller condition prove the global existence of

$$\mathrm{d}\hat{V}_k(t) = \left(a_k(t) - \frac{\sigma_k(t)^2}{2}\right)\frac{1}{2}\hat{V}_k^{-1}(t) - \frac{1}{2}b_k(t)\hat{V}_k(t)\mathrm{d}t + \frac{\sigma_k(t)}{2}\mathrm{d}B(t), \quad \hat{V}_k(0) = \hat{V}_{0,k} > 0,$$

for $t \in [0,\infty[$.

Using a countable dense subset $\tilde{K} \subset K$ and the point mass functional we show that the paths keep positive except at a $\mathbb{P}$-zero set $N := \bigcup_{k\in\tilde{K}} N_k$. So for all $\omega \in \Omega \setminus N$ the paths stay positive on $[0,\infty[$

for all k. □

3.4 Corollary. *Let K,a,b,σ be given as in Theorem 3.3. Then the generalized CIR equation*

$$d\overline{X}(t) = \big(a(t) - b(t)\overline{X}(t)\big)\,dt + \sigma(t)\sqrt{\overline{X}(t)}\,dB(t), \quad \overline{X}(0) = \overline{X}_0 \in C(K)_+,$$

for $t \in [0,\infty[$, has a global solution.

Proof. The proof is immediate by Itô's lemma. □

The next step is to transform the square-root process into the setting of a non-commutative von Neumann algebra $\mathcal{A}$ but still with a classical Brownian motion as driving process.

3.5 Theorem. *Let $V_0 \in \mathcal{A}_+$ and let a,σ,$b : [0,\infty[\to \langle V_0, id\rangle_+$ such that (1) holds for all $t \in [0,\infty[$. Then the SDE*

$$d\overline{V}(t) = \left(\left(a(t) - \frac{\sigma^2(t)}{2}\right)\frac{1}{2}\overline{V}^{-1}(t) - \frac{b(t)}{2}\overline{V}(t)\right)dt + \frac{\sigma(t)}{2}\,dB(t), \quad \overline{V}(0) = V_0, \tag{4}$$

for $t \in [0,\infty[$, has a global solution in $V \in C([0,\infty[, L_2(\mathbb{P}_\varphi,\mathcal{A}_+))$. Note that id is the unit in the corresponding von Neumann algebra.

Proof. Let $\hat{V}(t)$ be a global solution in $C(K)_+$ by Theorem 3.3. By the functional calculus we see that $T(\hat{V}(t))$ is a positive solution to (4) under the Feller condition, where $T : C(K) \longrightarrow \langle V_0, id\rangle$ is the functional calculus mapping for the self-adjoint element V_0. In particular we have that $\overline{V}(t) \in \langle V_0, id\rangle_+$ for all $t \in [0,\infty[$.

□

Using the Itô calculus we get:

3.6 Corollary. *Let $a(t)$,$b(t)$,$\sigma(t)$ be given as in Theorem 3.5. Then the generalized CIR equation*

$$d\overline{X}(t) = \big(a(t) - b(t)\overline{X}(t)\big)\,dt + \sigma(t)\sqrt{\overline{X}(t)}\,dB(t), \quad \overline{X}(0) = \overline{X}_0 \in \mathcal{A}_+,$$

for $t \in [0,\infty[$, has a global solution $X \in C([0,\infty[, L_2(\mathbb{P}_\varphi,\mathcal{A}_+))$.

In the above Theorem 3.5 we restricted the coefficient functions to the von Neumann subalgebra $\langle V_0, \mathrm{id}\rangle_+$. We now follow up with our main Theorem 3.9 where we show the existence of our solution

in the case of a free Brownian motion. For this setting we can allow $a : [0,\infty[\to \mathcal{A}_+$ but will have to restrict b this time to the case of a scalar $b > 0$.

3.7 Remark. 1. The last step we need for the existence of a positive solution in the context of free probability, is to change the driving process in the general von Neumann algebra-valued SDE from a classical Brownian motion to a free Brownian motion. Having established that the corresponding SDE driven by a classical Brownian motion stays positive under Feller by Theorem 3.5 we will show that both solutions (the classical and the free Brownian motion case) are L_2-isometric for all $t \in [0,T]$, a common existence interval of both solutions. Finally we will prove, using the upcoming proposition, that $V(T)$, the solution to the free equation evaluated at T, is invertible and therefore the free SDE has a global positive solution as well.

2. For a more convenient notation we may write φ instead of $\mathbb{E}_{\mathbb{P}_\varphi}$ in the sequel for the process $(\overline{V}(t))_{t\geq 0}$. No confusion will arise at any point.

3. For Lipschitz-continuous and $\mathcal{A}$-valued functions a,b,c we call a process

$$V(t) = V_0 + \int_0^t a(V(s))ds + \int_0^t b(V(s))dW(s)c(V(s)),$$

for $t \in [0,T[$, a free Itô process. The existence is guaranteed by the main result due to Kargin (2011).

Before stating the main result on the existence of a global solution, we formulate a result, which may be of independent interest.

3.8 Proposition. *Let $T > 0$ and let $V \in C([0,T[,\mathcal{A}) \cap C([0,T], L_2(\varphi))$ be a free Itô process and let $\overline{V} \in C([0,T],L_2(\mathbb{P}_\varphi,\mathcal{A}))$ be a vector-valued Itô processes. Suppose for all $t \in [0,T[$ and all orthogonal projections p free of $V(t)$ that we have*

$$\|pV(t)^2p\|_1 = \|p\overline{V}(t)^2p\|_1. \tag{5}$$

1. *If $V(t) \in \mathcal{A}$ and $V(T) \in L_2(\varphi)$ then $V(T) \in \mathcal{A}$.*

2. *If $\overline{V}(t)$ is invertible for all $t \in [0,T]$, $V(t)$ is invertible for all $t \in [0,T[$ and $V(T) \in \mathcal{A}$, then $V(T) \in \mathcal{A}$ is also invertible.*

Proof. 1. We suppose that $V(T) \notin \mathcal{A}$. Then, since $V(T) \in L_2(\varphi)$, we find a sequence of projections (p_n), such that

$$\frac{1}{\varphi(p_n)} \|p_n V(T)^2 p_n\|_1 \geq n.$$

Since $\overline{V} \in C([0,T], L_2(\mathbb{P}_\varphi, \mathcal{A}))$, we find a $\tilde{T} < T$ such that

$$\|\overline{V}(T)^2 - \overline{V}(\tilde{T})^2\|_1 < 1.$$

Since $\overline{V}(T) \in \mathcal{A}$, we get

$$\frac{1}{\varphi(p_n)} \|p_n \overline{V}(T)^2 p_n\|_1 \leq \|\overline{V}(T)^2\| < \infty.$$

Since $V(T) = V(\tilde{T}) + (V(T) - V(\tilde{T}))$, the element $V(T)$ is generated by the projections in the von Neumann algebras $\mathcal{A}_{\tilde{T}}$ and $\mathcal{A}_T$, which are generated by $\{W(t) \mid t \leq \tilde{T}\} \cup \{V_0, \mathrm{id}\}$ and $V(T) - V(\tilde{T})$, respectively. Therefore $\mathcal{A}_{\tilde{T}}$ and $\mathcal{A}_T$ are free. Consider two cases:

a) $p_n \in \mathcal{A}_{\tilde{T}}$ for infinitely many $n \in \mathbb{N}$. (For simplicity we assume for all $n \in \mathbb{N}$). In this case we consider a new starting value $V(\tilde{T})$ instead of V_0 and instead of the interval $[0,T]$ the interval $[\tilde{T},T]$. Then for all $n \in \mathbb{N}$ the projection p_n is free of increments of the free Brownian motion $W(t) - W(s)$ for $\tilde{T} \leq s < t \leq T$ and the proof runs as in the case that all $p_n \in \langle V_0, \mathrm{id} \rangle_+$.

b) Only for finitely many $n \in \mathbb{N}$ we have $p_n \in \mathcal{A}_{\tilde{T}}$. Then we choose the infinitely many $p_n \in \mathcal{A}_T$, for simplicity all $n \in \mathbb{N}$. Here, all those p_n are free of $\mathcal{A}_{\tilde{T}}$ and thus again we have that $W(t) - W(s)$ for $0 \leq s < t \leq \tilde{T}$ are free of p_n.

In both cases we get by (5)

$$\varphi(p_n \overline{V}(t)^2 p_n) = \varphi(p_n V(t)^2 p_n) \text{ for } t \in [0,\tilde{T}]. \tag{6}$$

Consequently, for both cases it holds that

$$\begin{aligned}
\infty > \|\overline{V}(T)^2\| &\geq \frac{1}{\varphi(p_n)}\|p_n\overline{V}(T)^2p_n\|_1 \\
&= \frac{1}{\varphi(p_n)}\|p_n(\overline{V}(\tilde{T})^2 + (\overline{V}(T)^2 - \overline{V}(\tilde{T})^2))p_n\|_1 \\
&\overset{(6)}{\geq} \left|\frac{1}{\varphi(p_n)}\|p_nV(\tilde{T})^2p_n\|_1 - \frac{1}{\varphi(p_n)}\|p_n(\overline{V}(T)^2 - \overline{V}(\tilde{T})^2)p_n\|_1\right| \\
&> n-1,
\end{aligned}$$

by the isometry (5). This is a contradiction.

2. Suppose $V(T)$ and hence $V(T)^2$ is not invertible. We find a sequence of projections (p_n) in the von Neumann algebra generated by the self-adjoint element $V(T)$ such hat

$$\alpha_n = \frac{1}{\varphi(p_n)}\|p_nV(T)^2p_n\|_1 \longrightarrow 0,\ n \to \infty. \tag{7}$$

We know that $\overline{V}(T)$ is invertible and therefore we get by the properties of p_n that

$$0 < \frac{1}{\|(\overline{V}(T)^2)^{-1}\|} \leq \frac{1}{\varphi(p_n)}\|p_n\overline{V}(T)^2p_n\|_1.$$

Since both processes are continuous mappings

$$\overline{V},V : [0,T] \longrightarrow \mathcal{A},$$

i.e. $\overline{V},V \in C([0,T],\mathcal{A})$, we can find a $\tilde{T} \in]0,T[$, s.t.

$$\left|\frac{1}{\|(\overline{V}(T)^2)^{-1}\|} - \frac{1}{\|(\overline{V}(\tilde{T})^2)^{-1}\|}\right| < \frac{1}{4\|(\overline{V}(T)^2)^{-1}\|}$$

and

$$\|V(T)^2 - V(\tilde{T})^2\| < \frac{1}{4\|(\overline{V}(T)^2)^{-1}\|}.$$

Thus, for all $n \in \mathbb{N}$:

$$\beta_n = \frac{1}{\varphi(p_n)}\|p_n(V(T)^2 - V(\tilde{T})^2)p_n\|_1 \leq \frac{1}{\varphi(p_n)}\varphi(p_n)\|V(T)^2 - V(\tilde{T})^2\| < \frac{1}{4\|(\overline{V}(T)^2)^{-1}\|}.$$

Again, since $V(T) = V(\tilde{T}) + (V(T) - V(\tilde{T}))$, we consider as in the proof to 1. two cases for the projections (p_n) and conclude by (5) that

$$\varphi(p_n \overline{V}(t)^2 p_n) = \varphi(p_n V(t)^2 p_n) \text{ for } t \in [0,\tilde{T}]. \tag{8}$$

Consequently, it holds that

$$\begin{aligned} \frac{1}{\varphi(p_n)} \|p_n V(T)^2 p_n\|_1 &= \frac{1}{\varphi(p_n)} \|p_n (V(\tilde{T})^2 + (V(T)^2 - V(\tilde{T})^2)) p_n\|_1 \\ &\overset{(8)}{\geq} \left| \frac{1}{\varphi(p_n)} \|p_n \overline{V}(\tilde{T})^2 p_n\|_1 - \frac{1}{\varphi(p_n)} \|p_n (V(T)^2 - V(\tilde{T})^2) p_n\|_1 \right| \\ &> \frac{3}{4\|(\overline{V}(T)^2)^{-1}\|} - \beta_n \geq \frac{1}{2\|(\overline{V}(T)^2)^{-1}\|} > 0. \end{aligned}$$

This is a contradiction to the assumption (7) and hence $V(T)$ is invertible.

□

3.9 Theorem. *Let $V_0 \in \mathcal{A}_+$ be given. Let $a : [0,\infty[\to \mathcal{A}_+$, $\sigma : [0,\infty[\to \langle V_0, id \rangle_+$ and $b > 0$ a constant such that (1) holds. Then the free SDE*

$$dV(t) = \left(\left(a(t) - \frac{\sigma^2(t)}{2} \right) \frac{1}{2} V^{-1}(t) - \frac{b}{2} V(t) \right) dt + \frac{\sigma(t)}{2} dW(t), \quad V(0) = V_0, \tag{9}$$

for $t \in [0,\infty[$, has a global solution $V \in C([0,\infty[, \mathcal{A}_+)$.

Proof. 1. In the first step we choose a maximal interval $[0,T[$, where a solution of the equation

$$\mathrm{d}V(t) = \left(\left(a(t) - \frac{\sigma^2(t)}{2} \right) \frac{1}{2} V^{-1}(t) - \frac{b}{2} V(t) \right) \mathrm{d}t + \frac{\sigma(t)}{2} \mathrm{d}W(t), \quad V(0) = V_0,$$

for $t \in [0,T[$, exists according to Kargin (2011). By Theorem 3.5 we know that the solution for

$$\mathrm{d}\overline{V}(t) = \left(\left(a(t) - \frac{\sigma^2(t)}{2} \right) \frac{1}{2} \overline{V}^{-1}(t) - \frac{b}{2} \overline{V}(t) \right) \mathrm{d}t + \frac{\sigma(t)}{2} \mathrm{d}B(t), \quad \overline{V}(0) = V_0,$$

for $t \in [0,\infty[$, exists globally. First we prove the following three isometries.

$$\begin{aligned} \|V(t)\|_2 &= \|\overline{V}(t)\|_2 \text{ for } t \in [0,T] \\ \|pV(t)^2p\|_1 &= \|p\overline{V}(t)^2p\|_1 \text{ for } t \in [0,T] \text{ and } p \in \langle V_0,\mathrm{id}\rangle_+ \\ \|pV(t)^2p\|_1 &= \|p\overline{V}(t)^2p\|_1 \text{ for } t \in [0,T] \text{ and } p \text{ free of } V(t) \end{aligned} \tag{10}$$

The third isometry directly follows from the first by freeness.

We approximate the solutions on $[t_0,T[$ and thus, select a partition $\mathcal{Z}$ of the interval $[t_0,T]$ namely $0 = t_0 < t_1 < \cdots < t_n = T$. We will omit the variable "t" in the expression of a and σ. The proof for the isometry

$$\|V(t)\|_2 = \|\overline{V}(t)\|_2 \text{ for } t \in [0,T]$$

is basically a complete induction. We will show it for $[t_0,t_1]$ and then the induction step from $[t_0,t_1]$ to $[t_1,t_2]$ to illustrate the procedure. We start with the interval $[t_0,t_1[$. For $t \in [t_0,t_1[$ we have

$$\overline{V}_{\mathcal{Z}}(t) = V_0 + \int_{t_0}^{t} \left(\left(a - \frac{\sigma^2}{2}\right)\frac{1}{2}\overline{V}_{\mathcal{Z}}^{-1}(s) - \frac{b}{2}\overline{V}_{\mathcal{Z}}(s)\right)\mathrm{d}s$$

and

$$V_{\mathcal{Z}}(t) = V_0 + \int_{t_0}^{t} \left(\left(a - \frac{\sigma^2}{2}\right)\frac{1}{2}V_{\mathcal{Z}}^{-1}(s) - \frac{b}{2}V_{\mathcal{Z}}(s)\right)\mathrm{d}s.$$

Thus we have by an easy approximation by step functions that $\overline{V}_{\mathcal{Z}}(t), V_{\mathcal{Z}}(t) \in \langle V_0,\mathrm{id}\rangle_+$ and

$$\varphi(V_{\mathcal{Z}}(t)^2) = \varphi(\overline{V}_{\mathcal{Z}}(t)^2).$$

By adding the classical Brownian motion term $\frac{\sigma}{2}B(t_1)$ and the free Brownian motion term $\frac{\sigma}{2}W(t_1)$, respectively we define

$$\begin{aligned} \overline{V}_{\mathcal{Z}}(t_1) &= \overline{V}_{\mathcal{Z}}(t_1-) + \frac{\sigma}{2}B(t_1) \\ V_{\mathcal{Z}}(t_1) &= V_{\mathcal{Z}}(t_1-) + \frac{\sigma}{2}W(t_1), \end{aligned}$$

where the first one is just a discretization of the global solution of Theorem 3.5 (which serves as our "reference solution") evaluated at t_1 and $V_{\mathcal{Z}}(t_1-)$ and $\overline{V}_{\mathcal{Z}}(t_1-)$, respectively is the local solution for $t \in [t_0,t_1[$. Since $\sigma \in \langle V_0,\mathrm{id}\rangle_+$ is free of $W(t_1)$ and $\varphi(W(t_1)) = 0$, we conclude with (Kargin, 2011,

p. 829) that

$$\varphi\left(\frac{\sigma}{2}B(t_1)\frac{\sigma}{2}B(t_1)\right) = \varphi\left(W(t_1)\frac{\sigma}{2}W(t_1)\frac{\sigma}{2}\right) = \frac{1}{4}\varphi(\sigma)^2 t_1$$

and see by the independence, resp. the freeness that

$$\varphi(\overline{V}_{\mathcal{Z}}(t_1)^2) = \varphi(V_{\mathcal{Z}}(t_1)^2).$$

On $[t_1, t_2[$ we again consider approximations of our solutions starting in $\overline{V}_{\mathcal{Z}}(t_1)$ and $V_{\mathcal{Z}}(t_1)$, (not equal but L_2-isometric) respectively:

$$\begin{aligned}
\overline{V}_{\mathcal{Z}}(t) &= \overline{V}_{\mathcal{Z}}(t_1) + \int_{t_1}^{t}\left(\left(a - \frac{\sigma^2}{2}\right)\frac{1}{2}\overline{V}_{\mathcal{Z}}^{-1}(s) - \frac{b}{2}\overline{V}_{\mathcal{Z}}(s)\right)\mathrm{d}s \\
\overline{V}_{\mathcal{Z}}(t) &= V_{\mathcal{Z}}(t_1) + \int_{t_1}^{t}\left(\left(a - \frac{\sigma^2}{2}\right)\frac{1}{2}V_{\mathcal{Z}}^{-1}(s) - \frac{b}{2}V_{\mathcal{Z}}(s)\right)\mathrm{d}s
\end{aligned}$$

Defining $\overline{X}(t) := \overline{V}_{\mathcal{Z}}(t)^2$, resp. $X(t) := V_{\mathcal{Z}}(t)^2$ and applying the trace we get a general ordinary differential equation (ODE) of the form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \alpha - by$$

with $\alpha = \varphi(a - \frac{\sigma^2}{2})$ and $y \in \{\varphi(\overline{X}(t)), \varphi(X(t))\}$ with the initial value

$$y(t_1) = \varphi(X(t_1)) = \varphi(\overline{X}(t_1)).$$

Since this ODE has a unique solution we get

$$\|\overline{V}_{\mathcal{Z}}(t)\|_2^2 = \varphi(X(t)) = \varphi(\overline{X}(t)) = \|V_{\mathcal{Z}}(t)\|_2^2 \text{ for } t \in [t_1, t_2[.$$

Again we add the classical Brownian motion term $\frac{\sigma}{2}(B(t_2) - B(t_1))$ and the free Brownian motion term $\frac{\sigma}{2}(W(t_2) - W(t_1))$, respectively and get

$$\begin{aligned}
\overline{V}_{\mathcal{Z}}(t_2) &= \overline{V}_{\mathcal{Z}}(t_2-) + \frac{\sigma}{2}(B(t_2) - B(t_1)) \\
V_{\mathcal{Z}}(t_2) &= V_{\mathcal{Z}}(t_2-) + \frac{\sigma}{2}(W(t_2) - W(t_1)).
\end{aligned}$$

As before by the independence, resp. freeness we conclude

$$\varphi(\overline{V}_{\mathcal{Z}}(t_2)^2) = \varphi(V_{\mathcal{Z}}(t_2)^2).$$

We may extend $\overline{V}_{\mathcal{Z}}$, resp. $V_{\mathcal{Z}}$ beyond the interval $[0,t_2[$. In the same respect as above we have for both processes extensions on $[0,t_n[$ such that

$$\varphi(\overline{V}_{\mathcal{Z}}(t)^2) = \varphi(V_{\mathcal{Z}}(t)^2) \text{ for all } t \in [0,t_n[.$$

Since these isometries hold by the continuity of the solutions for all partitions of $[0,T]$, we see that $V_{\mathcal{Z}}$ converges for $|\mathcal{Z}| \to 0$, i.e. if the length of the partition converges to 0:

$$\varphi(\overline{V}(t)^2) = \lim_{|\mathcal{Z}|\to 0} \varphi(\overline{V}_{\mathcal{Z}}(t)^2) = \lim_{|\mathcal{Z}|\to 0} \varphi(V_{\mathcal{Z}}(t)^2) = \varphi(V(t)^2) \text{ for all } t \in [0,T[.$$

We proceed by showing the second isometry of (10) and therefore introduce the equation

$$\begin{aligned} pVp(t) = pV_0p \quad &+ \quad \int_0^t \left(p\left(a - \frac{\sigma^2}{2}\right)p\frac{1}{2}(pVp)(s)^{-1} - p\frac{b}{2}pVp(s)\right)\mathrm{d}s \\ &+ \quad \int_0^t p\frac{\sigma}{2}\, p\, \mathrm{d}W(s)p, \quad pVp(0) = pV_0p \in p\mathcal{A}_+p, \end{aligned}$$

with $a \in \mathcal{A}_+$, $\sigma \in \langle V_0, \mathrm{id}\rangle_+$ and $b > 0$, all strictly positive, in $p\mathcal{A}p$. Since $p \in \langle V_0, \mathrm{id}\rangle_+$ by assumption, we have pbp, $p\sigma p \in \langle pV_0p, \mathrm{id}\rangle_+$ and the proof for the L_2-isometry above can be mimicked. For this we consider $\mathrm{d}(pVp(t))^2 = 2pVp(t)\,\mathrm{d}(pVp(t))$ to get a similar general ODE to the case above.

2. According to Kargin (2011), we know that the process $V(t)$ exists for $t < T$. The reference process $\overline{V}(t)$ exists globally with values in $\mathcal{A}$. Because of the isometry we can define

$$V(T) = L_2 - \lim_{t\to T} V(t) = V_0 + \int_0^T \Big(\frac{1}{2}(a - \frac{\sigma^2}{2})V^{-1}(t) - bV(t)\Big)dt + \frac{\sigma}{2}W(T).$$

Using Proposition 3.8 (1.) we can deduce that $V(T) \in \mathcal{A}$.

3. Step 2 told us that $V(T) \in \mathcal{A}$. We want to extend the (unique) solution $V(t)$ beyond T. To apply the basic result in Kargin (2011), we need the invertibility of $V(T)$. This would allow us an extension and the solution is global. Again by Proposition 3.8 (2.), we see that $V(T)$ is invertible. Therefore the solution to (9) is global. □

Having established the existence of a global solution to the square-root process we now apply the free Itô formula and get the global solution to the free CIR equation.

Let $X(t) = V(t)^2$. Then according to the free Itô formula:

$$\begin{aligned} \mathrm{d}X(t) &= (V(t)+\mathrm{d}V(t))^2 - (V(t))^2 = (\mathrm{d}V(t))^2 + \mathrm{d}V(t)V(t) + V(t)\mathrm{d}V(t) \\ &= \Big(2\Big(\frac{a-\frac{\sigma^2}{2}}{2V(t)} - \frac{b}{2}V(t)\Big)V(t) + \frac{\sigma^2}{2}\Big)\mathrm{d}t + \frac{\sigma}{2}V(t)\mathrm{d}W(t) + \frac{\sigma}{2}\mathrm{d}W(t)V(t) \\ &= (a - bV^2(t))\mathrm{d}t + \frac{\sigma}{2}V(t)\mathrm{d}W(t) + \frac{\sigma}{2}\mathrm{d}W(t)V(t) \\ &= (a - bX(t))\mathrm{d}t + \frac{\sigma}{2}\sqrt{X(t)}\mathrm{d}W(t) + \frac{\sigma}{2}\mathrm{d}W(t)\sqrt{X(t)}, \end{aligned}$$

for $t \in [0,\infty[$. Thus, we state

3.10 Theorem. *Let a,b,σ such that (1) and the assumptions made in (3.9) hold. Then the free SDE*

$$dX(t) = (a - bX(t))\,dt + \frac{\sigma}{2}\sqrt{X(t)}\,dW(t) + \frac{\sigma}{2}\,dW(t)\sqrt{X(t)}, \quad X(0) = X_0 \in \mathcal{A}_+,$$

for $t \in [0,\infty[$, has a global solution $X \in C([0,\infty[,\mathcal{A}_+)$.

3.11 Remark. *As mentioned in the introduction the above theorem guarantees the existence of a unique positive solution to the classical free CIR equation (2) as well and for a constant real-valued σ the solutions even coincide.*

4 Conclusion

We introduced the CIR equation to the world of free probability and made a contribution to the study of free SDEs. Initially developed as a tool for solving operator-theoretic problems free probability theory has been evolving into its own field of research inviting researchers from various disciplines such as finance, physics and signal processing to profit from and contribute to. As motivated in the introduction the connection between random matrix ensembles and their free operator-valued limit arouses the interest of many researchers by its attractive properties. Having introduced the free CIR equation(s) in this paper we paved the way for future applications to exploit this connection known as asymptotic freeness to process our setting for utilization in the real world. In particular the recent developments in big data and machine learning allow for the usage of very high-dimensional data as e.g. very large random matrices, such as portfolios incorporating a huge number of assets. We hope

that this paper will be used for portfolio optimization when the number of assets makes classical probabilistic approaches unfeasible.

Funding

This research received no external funding.

Conflicts of interest

The authors declare no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

An, G. and Gao, M. (2015). Poisson processes in free probability. *arXiv preprint*. https://arxiv.org/abs/1506.03130.

Anshelevich, M. (2002). Itô formula for free stochastic integrals. *Journal of Functional Analysis*, 188(1):292–315. https://doi.org/10.1006/jfan.2001.3849.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30. https://doi.org/10.1016/S0304-4076(95)01749-6.

Barndorff-Nielsen, O. E., Thorbjørnsen, S., et al. (2002). Self-decomposability and Lévy processes in free probability. *Bernoulli*, 8(3):323–366. https://www.jstor.org/stable/3318705.

Bayer, C., Friz, P. K., Gassiat, P., Martin, J., and Stemper, B. (2019). A regularity structure for rough volatility. *Mathematical Finance*. https://doi.org/10.1111/mafi.12233.

Biane, P. (1997). Free Brownian motion, free stochastic calculus and random matrices, in free probability theory. *Fields Institute Communications*, 12:1–19. https://ci.nii.ac.jp/naid/10004590009/.

Biane, P. (1998). Processes with free increments. *Mathematische Zeitschrift*, 227(1):143–174. https://doi.org/10.1007/PL00004363.

Biane, P. and Speicher, R. (1998). Stochastic calculus with respect to free Brownian motion and analysis on Wigner space. *Probability Theory and Related Fields*, 112(3):373–409. https://doi.org/10.1007/s004400050194.

Biane, P. and Speicher, R. (2001). Free diffusions, free entropy and free Fisher information. In *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, volume 37, pages 581–606. No longer published by Elsevier. https://doi.org/10.1016/S0246-0203(00)01074-8.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654. https://www.jstor.org/stable/1831029.

Bollerslev, T. and Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73(1):151–184. https://doi.org/10.1016/0304-4076(95)01736-4.

Bouchaud, J.-P. and Potters, M. (2009). Financial applications of random matrix theory: a short review. *arXiv preprint*. https://arxiv.org/abs/0910.1205.

Bru, M.-F. (1991). Wishart processes. *Journal of Theoretical Probability*, 4(4):725–751. https://doi.org/10.1007/BF01259552.

Comte, F., Coutin, L., and Renault, E. (2012). Affine fractional stochastic volatility models. *Annals of Finance*, 8(2-3):337–378. https://doi.org/10.1007/s10436-010-0165-3.

Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323. https://doi.org/10.1111/1467-9965.00057.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1:223–236. http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.16.5992.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407. https://www.jstor.org/stable/1911242.

El Euch, O. and Rosenbaum, M. (2019). The characteristic function of rough Heston models. *Mathematical Finance*, 29(1):3–38. https://doi.org/10.1111/mafi.12173.

El Euch, O., Rosenbaum, M., et al. (2018). Perfect hedging in rough Heston models. *The Annals of Applied Probability*, 28(6):3813–3856. https://projecteuclid.org/euclid.aoap/1538985636.

Fack, T. and Kosaki, H. (1986). Generalized s-numbers and τ-measurable operators. *Pacific Journal of Mathematics*, 123:269–300. https://projecteuclid.org/euclid.pjm/1102701004.

Fan, Z. (2006). Self-similarity of free stochastic processes. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 9(03):451–469. https://doi.org/10.1142/S0219025706002482.

Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics*, pages 173–182. https://www.jstor.org/stable/1969318.

Fink, H. and Schlüchtermann, G. (2018). Fractional Lévy Cox–Ingersoll–Ross and Jacobi processes. *Statistics & Probability Letters*, 142(C):84–91. https://doi.org/10.1016/j.spl.2018.07.004.

Gao, M. (2006). Free Ornstein-Uhlenbeck processes. *Journal of Mathematical Analysis and Applications*, 322(1):177–192. https://doi.org/10.1016/j.jmaa.2005.09.013.

Gao, M. et al. (2008). Free Markov processes and stochastic differential equations in von Neumann algebras. *Illinois Journal of Mathematics*, 52(1):153–180. https://projecteuclid.org/euclid.ijm/1242414126.

Gatheral, J., Jaisson, T., and Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6):933–949. https://doi.org/10.1080/14697688.2017.1393551.

Gouriéroux, C. (2006). Continuous time Wishart process for stochastic risk. *Econometric Reviews*, 25(2-3):177–217. https://doi.org/10.1080/14697688.2017.1393551.

Gourieroux, C. and Sufana, R. (2010). Derivative pricing with Wishart multivariate stochastic volatility. *Journal of Business & Economic statistics*, 28(3):438–451. https://www.jstor.org/stable/20750851.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327–343. https://doi.org/10.1093/rfs/6.2.327.

Hull, J. and White, A. (1990). Pricing interest-rate-derivative securities. *The Review of Financial Studies*, 3(4):573–592. https://doi.org/10.1093/rfs/3.4.573.

Jaber, E. A. and Euch, O. E. (2019). Multi-factor approximation of rough volatility models. *SIAM Journal of Financial Mathematics*, 10(2):309–349. https://doi.org/10.1137/18M1170236.

Kargin, V. (2011). On free stochastic differential equations. *Journal of Theoretical Probability*, 24(3):821–848. https://doi.org/10.1007/s10959-011-0341-z.

Keller-Ressel, M. and Steiner, T. (2008). Yield curve shapes and the asymptotic short rate distribution in affine one-factor models. *Finance and Stochastics*, 12(2):149–172. https://doi.org/10.1007/s00780-007-0059-z.

Kümmerer, B. and Speicher, R. (1992). Stochastic integration on the Cuntz algebra $\mathcal{O}^{\infty}$. *Journal of Functional Analysis*, 103(2):372–408. https://doi.org/10.1016/0022-1236(92)90126-4.

Maghsoodi, Y. (1996). Solution of the extended CIR term structure and bond option valuation. *Mathematical Finance*, 6(1):89–109. https://doi.org/10.1111/j.1467-9965.1996.tb00113.x.

Mandelbrot, B. B. (1997). The variation of certain speculative prices. In *Fractals and Scaling in Finance*, pages 371–418. Springer. https://doi.org/10.1007/978-1-4757-2763-0_14.

Mingo, J. A. and Speicher, R. (2017). *Free probability and random matrices*, volume 4. Springer. https://www.springer.com/de/book/9781493969418.

Murphy, G. J. (2014). *C*-algebras and operator theory*. Academic Press. https://doi.org/10.1016/C2009-0-22289-6.

Nica, A. and Speicher, R. (2006). *Lectures on the combinatorics of free probability*, volume 13. Cambridge University Press. https://doi.org/10.1017/CBO9780511735127.

Nourdin, I. and Taqqu, M. S. (2014). Central and non-central limit theorems in a free probability setting. *Journal of Theoretical Probability*, 27(1):220–248. https://link.springer.com/article/10.1007/s10959-012-0443-2.

Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance*, 3(1):15–102. https://doi.org/10.1016/0927-5398(95)00020-8.

Ryan, Ø. (2008). Free probability: Basic concepts, tools, applications, and relations to other fields. http://folk.uio.no/oyvindry/stocsem1.pdf.

Ryan, Ø. and Debbah, M. (2007). Free deconvolution for signal processing applications. In *International Symposium on Information Theory*, pages 1846–1850. IEEE. https://ieeexplore.ieee.org/document/4557490.

Schlüchtermann, G. and Yang, Y. (2016). Note on fractional CLKS-type stochastic differential equation–path-wise and in the Wick sense. https://www.researchgate.net/publication/299670926.

Speicher, R. (1990). A new example of 'independence' and 'white noise'. *Probability Theory and Related Fields*, 84(2):141–159. https://doi.org/10.1007/BF01197843.

Takesaki, M. (2002). *Theory of operator algebras I, Encyclopaedia of Mathematical Sciences*, volume 124. Springer. https://www.springer.com/de/book/9783540422488.

Takesaki, M. (2003). *Theory of operator algebras III, Encyclopaedia of Mathematical Sciences*, volume 127. Springer-Verlag, Berlin. https://www.springer.com/de/book/9783540429135.

Takesaki, M. (2013). *Theory of operator algebras II*, volume 125. Springer Science & Business Media. https://www.springer.com/gp/book/9783540429142.

Terp, M. (1981). L^p-spaces associated with von Neumann algebras. *Notes Copenhagen University*.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188. https://doi.org/10.1016/0304-405X(77)90016-2.

Voiculescu, D., Dykema, K. J., and Nica, A. (1992). *Free random variables*. Number 1. American Mathematical Society. https://books.google.de/books/about/Free_Random_Variables.html?id=r7BJAwAAQBAJ&redir_esc=y.

Voiculescu, D., Stammeier, N., and Weber, M. (2016). *Free probability and operator algebras*. https://www.ems-ph.org/books/book.php?proj_nr=208.

Yang, H. (2005). Calibration of the extended CIR model. *SIAM Journal on Applied Mathematics*, 66(2):721–735. https://www.jstor.org/stable/4096135.

III. The Impact of Sovereign Yield Curve Differentials on Value-at-Risk Forecasts for Foreign Exchange Rates

Fink, H.; Fuest, A.; Port, H. The Impact of Sovereign Yield Curve Differentials on Value-at-Risk Forecasts for Foreign Exchange Rates. Risks 2018, 6, 84.

Author Contributions[1]**:**

The general idea and conceptualization came from Holger Fink. He was responsible for supervision, as well as project administration. Andreas Fuest provided support concerning the methodology and was responsible for the functional data part in the literature review. Henry Port carried out the main work, including data curation, formal analysis, investigation, methodology, software, validation, visualization, writing of the original draft and writing review and editing.

[1] The explicit formulation from the corresponding manuscript is used.

risks

MDPI

Article

The Impact of Sovereign Yield Curve Differentials on Value-at-Risk Forecasts for Foreign Exchange Rates

Holger Fink [1,2], Andreas Fuest [2,3] and Henry Port [2,3,*]

1 Faculty Business Administration and International Finance, Nuertingen-Geislingen University, 72622 Nürtingen, Germany; holger.fink@hfwu.de
2 Center for Quantitative Risk Analysis, Department of Statistics, Ludwig-Maximilians-Universität München, 80799 Munich, Germany; andreas.fuest@stat.uni-muenchen.de
3 Institute of Statistics, Ludwig-Maximilians-Universität München, 80799 Munich, Germany
* Correspondence: henry.port@stat.uni-muenchen.de; Tel.: +49-89-2180-3522

Received: 19 June 2018; Accepted: 15 August 2018; Published: 20 August 2018

Abstract: A functional ARMA-GARCH model for predicting the value-at-risk of the EURUSD exchange rate is introduced. The model implements the yield curve differentials between EUR and the US as exogenous factors. Functional principal component analysis allows us to use the information of basically the whole yield curve in a parsimonious way for exchange rate risk prediction. The data analyzed in our empirical study consist of the EURUSD exchange rate and the EUR- and US-yield curves from 15 August 2005–30 September 2016. As a benchmark, we take an ARMA-GARCH and an ARMAX-GARCHX with the 2y-yield difference as the exogenous variable and compare the forecasting performance via likelihood ratio tests. However, while our model performs better in one situation, it does not seem to improve the performance in other setups compared to its competitors.

Keywords: value-at-risk; GARCH; yield curve; functional data; PCA

1. Introduction

Forecasting foreign exchange (FX) rates and especially risk inherent in such FX markets is of crucial importance to companies and individuals of Wall Street as well as Main Street. In particular, for portfolio managers, corporations and the like, managing FX risk is essential. Accordingly, a large amount of research is available in this field.

The work in, e.g., Meese and Rogoff (1983) analyzed FX rate models developed in the 1970s and came to the conclusion that these did not perform better than random walks, which exemplifies the difficulty in modeling spot FX rates. Even the addition of macroeconomic variables like expected inflation and interest rate differentials (among others) did not lead to significant improvements. Similar findings have been reported by Campbell and Clarida (1987), Meese and Rogoff (1988), Coughlin and Koedijk (1990), Edison and Pauls (1993), Chinn and Meese (1995) and Frankel and Rose (1995). Having said that, others disagree with the findings of the previous authors; see for instance Bjørnland and Hungnes (2006), Ang and Chen (2010), Chen and Tsang (2013) and Grisse and Nitschka (2015).

However, the above discussion might not be that surprising as it basically boils down to the question whether one believes in the efficient market hypothesis of Fama (1970). Nevertheless, forecasting volatility, which is of crucial importance to risk management, is another matter: e.g., Baillie and Bollerslev (1991) proposed a seasonal GARCH setup to model intraday FX rate volatility. In their findings, they experienced volatility to follow patterns that were very much alike over various hours of the day and to show a notable amount of serial correlation. For volatility based on daily FX spot levels, Vilasuso (2002) employed the fractionally-integrated GARCH by Baillie et al. (1996) and reported

a significant improvement in out-of-sample performance with respect to mean squared and mean absolute errors, as well as to accuracy, in comparison to the simpler GARCH or IGARCH models.

Moreover, the assumption of financial market log-returns being normally distributed has been criticized for a long time. Therefore, Hull and White (1998) introduced a model to estimate value-at-risk (VaR) that does not postulate a normal distribution, producing promising results. A similar approach was also adopted by Mittnik and Paolella (2000) and Kuester et al. (2006). The latter were able to identify a clear overall winner in their comparison of different VaR forecasting techniques, which turns out to be a combination of a heavy-tailed GARCH filter and extreme value theory approach.

Another approach to improve FX volatility and risk forecasting performance is to invoke external factors, as carried out by, e.g., Benavides and Capistràn (2012), who made use of implied volatility information from option prices. In line with the above-mentioned discussion on macroeconomic variables for spot models, several authors have incorporated yield curve information into FX risk models, as well; cf. Dominguez (1998), Neely (1999), Markiewicz (2012), Kočenda and Poghosyan (2009) and Ichiue and Koyama (2011). See furthermore Morana (2009), who analyzed the influence of macroeconomics, among which were interest rates, on FX rate volatility. When it comes to explicitly forecasting FX rate volatility employing macro variables, we mention the usage of neural networks as investigated, e.g., by Dunis and Huang (2002), who compared their model to GARCH models and reported superior results with their setup, in which they included yield curve data, among other inputs. Furthermore, we have Bauwens and Sucarrat (2010), who employed, among other macro variables, interest rates for exchange rate volatility forecasts.

For a more general analysis on the influence of macroeconomic factors on volatility forecasts, we refer to Christiansen et al. (2012), who found, in particular, more evidence for an influence of macro variables on FX rate volatility forecasts. Finally, the use of Euro deposit rates as a measure of performance for forecasts of FX volatility was introduced in West et al. (1993).

However, to the best of our knowledge, no one has yet investigated if using the complete sovereign yield curve differential (in contrast to just some specific maturities) does effectively improve VaR forecasting. The idea of using such functional exogenous parts in a GARCHX setup for volatility modeling has been introduced before by Fuest and Mittnik (2015). We extend this approach to an ARMAX-GARCHX-type setup, which allows the exogenous yield curve differential to influence both return and volatility.

In particular, we want to investigate the following for EURUSD as our lead example: Does the complete yield curve differential have significant effects on daily FX returns predictions? Does it improve VaR forecasts? Does the proposed mathematical machinery pay off when comparing it to a standard ARMA-GARCH model and an ARMAX-GARCHX with the 2y-yield difference as the exogenous variable? We will investigate these questions by the analysis of confidence intervals and by the evaluation of likelihood ratio tests.

Dealing with functional data, we conclude this section with a short survey on functional time series research. A comprehensive treatment of functional data analysis (FDA) is Ramsay and Silverman (2005). Models for functional time series have been pioneered in the seminal treatment of Bosq (2000), where a generalization of pure autoregressive models to the functional case was developed. More recently, dimension reduction techniques (Hörmann and Kokoszka 2010) and dynamic models based on such techniques (Aue et al. 2015; Hyndman and Shang 2009) have been proposed and investigated. The work in Klepsch et al. (2017) proposed functional ARMA models and Hörmann et al. (2013) a functional version of the ARCH model. Functional time series approaches have very recently also been employed to model yield curve dynamics Kowal et al. (2017a, 2017b). In our study, however, a scalar time series is of primary interest. The idea to map the full information inherent in functional data to first- and, possibly, higher-order moments of the conditional distribution of a scalar time series has been put forward in Fuest and Mittnik (2015) and Brockhaus et al. (2017).

The remainder of this paper is structured as follows: In Section 2, we will present the data used for our analysis and discuss the construction of the EUR- and US-yield curve. Section 3 starts with

a short introduction to functional principal component analysis, which is followed by presenting the method of implementation used in this investigation, as shown in Ramsay and Silverman (2005). Having described the background mathematics, we introduce our model and the estimation procedure applied later. In Section 4, we present our estimation results and discuss the implications. Finally we conduct one-day VaR forecasts and compare these to predictions coming from the competing models.

All calculations have been performed with MATLAB (R2014b, MathWorks, Natick, MA, USA) using the financial econometrics toolbox by Sheppard (2013) and the functional data toolbox by Ramsay (2014).

2. Data

In this section, we shall briefly present the data used in our analysis to come. In particular, the present paper considers daily prices of the EURUSD exchange rate and daily levels of the EUR- and US-yield curves. Our time period ranges from 15 August 2005–30 September 2016, which leaves us with a total sample of $n = 2905$ observations. Figure 1 shows the daily log-returns for the FX series, while Figure 2 visualizes the yield curve differential via the 2y-yields.

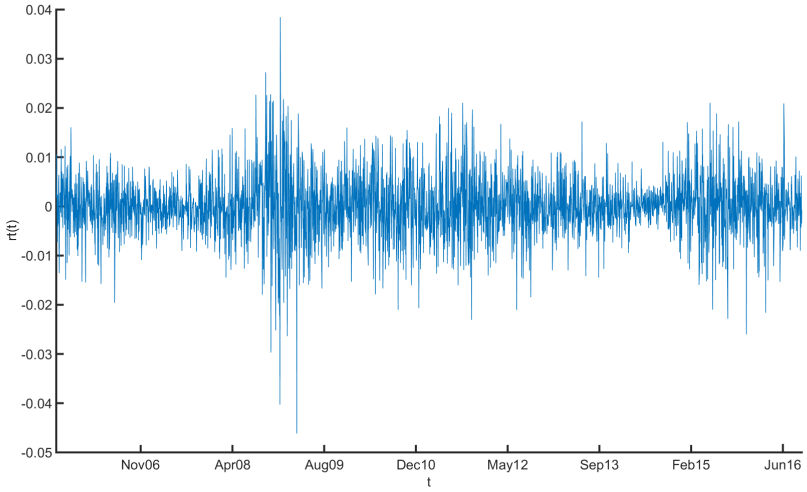


Figure 1. Log-returns of the EURUSD exchange rate from 16 August 2005–30 September 2016.

To be precise, we constructed the yield curve for Europe by using the EONIA (European overnight index average) offered rate for a maturity of 0 and the overnight index swap (OIS) rates based on EONIA with maturities of 1–12 months, 15 months, 18 months, 21 months and 2–10 years. The EONIA for the shortest maturity and the OIS for longer maturities are representative choices for constructing a Euro yield curve (cf. FBE (2008)), since overnight index swaps have established themselves as a benchmark proxy for the risk-free rate. (cf. Hull and White (2012) and Filipović and Trolle (2013), among others.)

For the US-yield curve, we used the federal funds rate for a maturity of 0 and yields obtained from US government bonds with maturities of 1, 3 and 6 months and 1–5, 7 and 10 years. Figure 3 visualizes the varying shape of the resulting curves. Finally, we converted the obtained discrete yield curves into functional data by using cubic B-splines. In the remainder of this paper, the terms "sovereign rate curve", "rate curve" and "yield curve" will be used interchangeably.

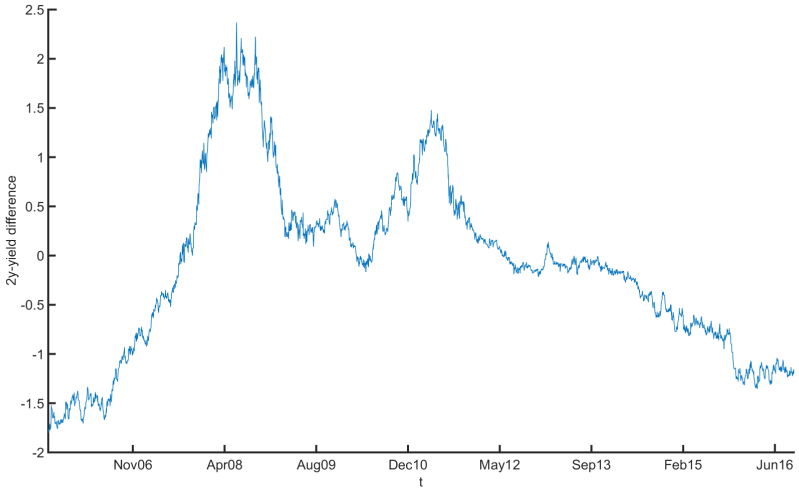


Figure 2. 2y-yield EURUSD differential from 15 August 2005–30 September 2016.

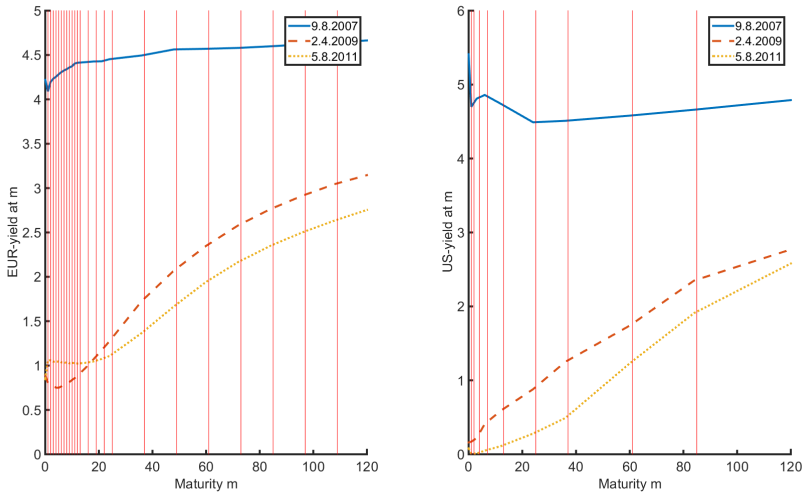


Figure 3. Yield curves for EUR (**left**) and the US (**right**). Parallels to the y-axis mark the maturities from the dataset.

3. Theory and Methods

In this section, we show how all the information contained in the yield curve at a given point in time can be mapped to the mean and variance parameters of the conditional distribution of exchange rate returns. The core technique we need for this purpose is functional principal component analysis (FPCA). We will sketch the theory and present the method of implementation that we used for our data, the implementation via basis expansion. As we will show, the estimated eigenfunctions from FPCA are also used to conveniently represent and estimate the functional parameters of our model. We will heavily draw on Ramsay and Silverman (2005), as well as Hörmann and Kokoszka (2012). For details on FPCA theory, we refer to the former and for a comprehensive treatment of functional regression models to the latter.

3.1. Functional Principal Components

We consider a stochastic process $(x_t)_{t\in\mathcal{T}}$. In order to simplify notation, we use lowercase letters for the processes, as well as for the realizations. For an ordered finite index set $\mathcal{T}$, we index our observations as $x_1, \ldots, x_{|\mathcal{T}|}$.

Now, consider the x_t as elements of the Hilbert space $\mathcal{L}^2([0,M])$. The Hilbert space $\mathcal{L}^2([0,M])$ is endowed with the inner product:

$$\langle x, y \rangle = \int_0^M x(s)y(s)\ \mathrm{d}s, \quad \forall x, y \in \mathcal{L}^2([0,M]),$$

so we have that x_t is square integrable.

Our curve valued process x_t exhibits a mean function, $\mu(s) = \mathbb{E}(x_t(s))$, and a covariance operator,

$$\mathcal{C} : \mathcal{L}^2([0,M]) \longrightarrow \mathcal{L}^2([0,M]), f \mapsto \int_0^M \mathrm{Cov}(r,s)f(r)\ \mathrm{d}r.$$

with covariance kernel $\mathrm{Cov}(r,s) = \mathrm{Covariance}(x_t(r), x_t(s))$.

Both mean function and covariance kernel are assumed to be constant over time $t \in \mathcal{T}$. Further note that for a $z \in \mathcal{L}^2([0,M])$, we have $\mathcal{C}(z) = \mathbb{E}[\langle x - \mu, z \rangle\ (x - \mu)]$.

Our goal is to find orthonormal weight functions, $\gamma_k \in \mathcal{L}^2([0,M])$, maximizing the unconditional variance of the scores $\xi_k = \langle \gamma_k, x_t - \mu \rangle = \int_0^M \gamma_k(s)(x_t(s) - \mu(s))\ \mathrm{d}s$, i.e.,

$$\mathrm{Var}(\xi_k) = \int_0^M \int_0^M \gamma_k(r)\mathrm{Cov}(r,s)\gamma_k(s)\ \mathrm{d}r\ \mathrm{d}s. \tag{1}$$

The weight functions, γ_k, are subject to orthonormality constraints, i.e., to:

$$\langle \gamma_k, \gamma_l \rangle = \int_0^M \gamma_k(s)\gamma_l(s)\ \mathrm{d}s = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l. \end{cases} \tag{2}$$

Then, the weight functions γ_k happen to be the eigenfunctions of the covariance operator $\mathcal{C}$. $\mathcal{C}$ is a bounded, symmetric, positive operator, and its eigenfunctions γ_k form an orthonormal basis of $\mathcal{L}^2([0,M])$. The so-called Fredholm (integral) equations:

$$\int_0^M \mathrm{Cov}(r,s)f(s)\ \mathrm{d}s = \lambda_k f(r) \tag{3}$$

give us the γ_k, which are ordered ascendingly as their corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. Finally, these eigenfunctions maximize (1).

Further note that $\mathrm{Cov}(r,s)$ is a Mercer kernel. In particular, this means that it can be represented in terms of the eigenvectors, γ_k, and the eigenvalues, λ_k, of the covariance operator $\mathcal{C}$,

$$\mathrm{Cov}(r,s) = \sum_{k=1}^{\infty} \lambda_k \gamma_k(r)\gamma_k(s).$$

Now, applying the theorem of Karhunen–Loève, we are able to represent the centered (yield) curve process, $x_t - \mu$, via its eigenbasis $(\gamma_k)_k$,

$$x_t - \mu = \sum_{k=1}^{\infty} \xi_{k,t}\gamma_k,$$

where the eigenfunctions, γ_k, are the eigenvectors of the linear operator $\mathcal{C}$, and the scores, $\xi_{k,t}$, have the properties:

$$\mathbb{E}(\xi_{k,t}) = 0,\ \mathrm{Var}(\xi_{k,t}) = \mathbb{E}(\xi_{k,t}^2) = \lambda_k \text{ and } \mathbb{E}(\xi_{k,t} \cdot \xi_{l,t}) = 0 \text{ for } k \neq l.$$

Recall that the eigenvalues, λ_k, are ordered in a non-decreasing order, so that the functional principal components are sorted by their contribution to the variation of the x_t's.

Let K be the smallest number necessary to explain a certain amount of the curves' total variation, then we can approximate our curve by:

$$x_t \approx \mu + \sum_{k=1}^{K} \xi_{k,t}\gamma_k.$$

Consider a yield curve x_t : Maturities $\rightarrow$ Yields, observed at time t. The above result means that we can approximate the information inherent in the yield curve over the entire range of observed maturities by a (possibly small) set of K scalar values. Note, however, that real-world observed yield curve data are only available for certain maturities m, which are, in most cases, not even structured in an equidistant way.[1] As is common practice in functional data analysis, we use spline interpolation of the observed yields to obtain equidistant maturities.

The resulting approximations of the yield curve can then be implemented and analyzed in our functional principal component analysis. There are several possibilities of implementing FPCA, of which we chose the method of "basis expansion" in our application, as is presented in Ramsay and Silverman (2005), which we briefly describe in the following.

We start by expressing our data, given by a matrix:

$$\begin{pmatrix} x_1(0) & \cdots & \cdots & x_1(M) \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ x_{|\mathcal{T}|}(0) & \cdots & \cdots & x_{|\mathcal{T}|}(M) \end{pmatrix},$$

in terms of a basis $\mathcal{B}$ in order to obtain a curve x_t. Thus, for each $t \in \mathcal{T}$, we can represent x_t by:

$$x_t(s) = \sum_{k=1}^{|\mathcal{B}|} c_{t,k}\psi_k(s),$$

which is equivalent to:

$$\begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_{|\mathcal{T}|} \end{pmatrix} = \boldsymbol{C} \cdot \boldsymbol{\psi} = \begin{pmatrix} c_{1,1} & \cdots & \cdots & c_{1,|\mathcal{B}|} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ c_{|\mathcal{T}|,1} & \cdots & \cdots & c_{|\mathcal{T}|,|\mathcal{B}|} \end{pmatrix} \cdot \begin{pmatrix} \psi_1 \\ \vdots \\ \vdots \\ \psi_{|\mathcal{B}|} \end{pmatrix},$$

where $|\mathcal{B}|$ denotes the number of basis functions of the chosen basis $\mathcal{B}$.

We then have:

$$\mathrm{Cov}(r,s) = |\mathcal{T}|^{-1}\boldsymbol{\psi}(r)^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{\psi}(s).$$

Now, define the symmetric matrix:

$$\boldsymbol{W} = \int_0^M \boldsymbol{\psi}(s)\boldsymbol{\psi}(s)^{\mathrm{T}}\,\mathrm{d}s = \left(\int_0^M \psi_i(s)\psi_j(s)\,\mathrm{d}s\right)_{i,j}.$$

---

[1] "Observed" yield curves are actually estimates obtained from observed bond prices. In the present paper, as in almost all of the literature (see for example Diebold and Li (2006)), we treat the yield curve data as if they had been observed directly.

We assume that the eigenfunctions γ_k are representable using the basis $\mathcal{B}$ in the form:

$$\gamma_k = \sum_{l=1}^{|\mathcal{B}|} \alpha_{k,l}\psi_l = \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\alpha_k},$$

where $\boldsymbol{\alpha_k} = (\alpha_{k,1}, \ldots, \alpha_{k,|\mathcal{B}|})$. Now, we express the Fredholm Equation (3) in terms of our basis $\mathcal{B}$ by considering the left-hand side of (3) with $f = \gamma_k$,

$$\int_0^M \mathrm{Cov}(r,s)\gamma_k(s)\,\mathrm{d}s = \int_0^M |\mathcal{T}|^{-1}\boldsymbol{\psi}(r)^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{\psi}(s)\boldsymbol{\psi}(s)^{\mathrm{T}}\boldsymbol{\alpha_k}\,\mathrm{d}s = \boldsymbol{\psi}(r)^{\mathrm{T}}|\mathcal{T}|^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{W}\boldsymbol{\alpha_k}.$$

Employing the basis representation of γ_k, the Fredholm equations then form:

$$\boldsymbol{\psi}(r)^{\mathrm{T}}|\mathcal{T}|^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{W}\boldsymbol{\alpha_k} = \lambda_k\boldsymbol{\psi}(r)^{\mathrm{T}}\boldsymbol{\alpha_k}$$

and we obtain:

$$|\mathcal{T}|^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{W}\boldsymbol{\alpha_k} = \lambda_k\boldsymbol{\alpha_k}.$$

Considering the orthogonality correspondence:

$$\langle\gamma_k, \gamma_l\rangle = 0 \;\Leftrightarrow\; \boldsymbol{\alpha_k}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\alpha_l} = 0,$$

we note that for $\|\gamma_k\| = 1$, we have $\boldsymbol{\alpha_k}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\alpha_k} = 1$, resulting in the same correspondence from above for normality. By defining $\boldsymbol{u_k} := \boldsymbol{W}^{1/2}\boldsymbol{\alpha_k}$, we arrive at the symmetric eigenvalue problem:

$$|\mathcal{T}|^{-1}\boldsymbol{W}^{1/2}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{W}^{1/2}\boldsymbol{u_k} = \lambda_k\boldsymbol{u_k}.$$

Solving this problem gives us the $\boldsymbol{u_k}$, and in turn, we get:

$$\boldsymbol{\alpha_k} = \boldsymbol{W}^{-1/2}\boldsymbol{u_k}.$$

The scores $\xi_{k,t}$ are then obtained by:

$$\xi_{k,t} = \langle x_t - \mathbb{E}(x_t), \boldsymbol{\alpha_k}\rangle.$$

Next, we shortly discuss how the mean, eigenfunctions and the corresponding scores can be estimated from data. As mentioned above, we observe values of the yield curves x_t on a (possibly non-equidistant) grid of maturities. By spline interpolation, we obtain a discrete version of x_t on an equidistant grid of maturities $0, \ldots, M$. Therefore, the realization of each x_t is an $M+1$-dimensional vector. Using interpolation, the number $J+1$ of grid points can be chosen arbitrarily, and one obtains an equidistant grid of the form $m_0 = 0, m_1 = M/J, m_2 = 2M/J, \ldots, m_{J-1} = (J-1)M/J, m_J = M$. The more complex the variation pattern of the curves, the higher J has to be chosen. For yield curves, the variation pattern is typically rather simple and the curves rather smooth, so that a moderate M should be reasonable. We found that a grid distance of one month is sufficient for our purpose, resulting in the choice $J = 120$.

The estimates of the mean function and the covariance kernel are given by:

$$\hat{\mu}(m_j) = \frac{1}{|\mathcal{T}|}\sum_{t=1}^{|\mathcal{T}|} x_t(m_j), \quad m_j \in [0, M]$$

and:

$$\widehat{\mathrm{Cov}} = \frac{1}{|\mathcal{T}|}\boldsymbol{X^c}^{\mathrm{T}}\boldsymbol{X^c},$$

where $|\mathcal{T}|$ is the number of observations (e.g., the number of observed yield curves x_t), $\boldsymbol{x}_t = [x_t(0), \ldots, x_t(M)]^{\mathrm{T}}$, $\hat{\boldsymbol{\mu}} = [\hat{\mu}(0), \ldots, \hat{\mu}(M)]^{\mathrm{T}}$ and $\boldsymbol{X}^c = [\boldsymbol{x}_1 - \hat{\boldsymbol{\mu}}, \ldots, \boldsymbol{x}_{|\mathcal{T}|} - \hat{\boldsymbol{\mu}}]^{\mathrm{T}}$.

In this framework, the eigenvalues $\hat{\lambda}_k$ and eigenvectors $\hat{\gamma}_k$ of $\widehat{\mathrm{Cov}}$ can be estimated using standard software for PCA. By means of numerical integration, we get empirical scores:

$$\hat{\xi}_{k,t} = \int_0^M (x_t(m) - \hat{\mu}(m))\hat{\gamma}_k(m)\mathrm{d}m.$$

The covariance operator $\mathcal{C}$ itself can be estimated by:

$$\hat{\mathcal{C}}(z) = \frac{1}{|\mathcal{T}|}\sum_{t=1}^{|\mathcal{T}|}\langle x_t - \hat{\mu}, z\rangle\,(x_t - \hat{\mu}) \quad \text{for } z \in \mathcal{L}^2([0, M]).$$

3.2. Econometric Model

We now define our model, which incorporates curve-valued information into a (scalar) ARMA-GARCH framework.

Definition 1 (ARMA(1,1)FunX-logGARCH(1,1)FunX process)**.** *Let x_t be drawn from a curve-valued exogenous process. Then, r_t follows an ARMA(1,1)FunX-logGARCH(1,1)FunX process if:*

$$\begin{aligned} r_t &= \alpha_0 + \alpha r_{t-1} + \beta\epsilon_{t-1} + \epsilon_t + \int_0^M \lambda(m)x_{t-1}\,dm, \\ \epsilon_t &= \sigma_t e_t, \\ \log\sigma_t^2 &= \omega + \gamma\log\epsilon_{t-1}^2 + \delta\log\sigma_{t-1}^2 + \int_0^M \rho(m)x_{t-1}\,dm, \end{aligned}$$

with $e_t \sim WN(0,1)$.

In our case, we have:

- r_t is the FX rate EURUSD
- $z_t^{\mathrm{US}}(\cdot)$ is the sovereign rate curve for the US
- $z_t^{\mathrm{EUR}}(\cdot)$ is the sovereign rate curve for EUR.
- $x_t(\cdot) = z_t^{\mathrm{EUR}}(\cdot) - z_t^{\mathrm{US}}(\cdot)$

Estimation is rendered feasible by expressing all functional elements of the model within an ARMAX(1,1)-log-GARCHX(1,1) framework, where the X-part is given by the K-dimensional approximations of functional observations and parameters. Concerning the resulting Gaussian QML estimators, there is some asymptotic theory for the GARCHX (see Han (2015), Han and Kristensen (2014)) and ARMAX models (see Hannan et al. (1980)). However, according to Sucarrat et al. (2016), Francq et al. (2013) and Hansen et al. (2012), there is none for the log-GARCHX case.

Remark 1. *The generalization of our setup to an ARMAX(A,B)-log-GARCHX(1,1) or even an ARMAX(A,B)-log-GARCHX(C,D), allowing for an arbitrary number of lags, is straightforward. For the sake of brevity and conciseness, we will nevertheless work with the ARMAX(1,1)-log-GARCHX(1,1) case in this paper, since the novelty and core idea of our model lies within the exogenous part.*

As introduced in Section 3.1, we can approximate our curve-valued process x_t by its first K principal components:

$$x_t \approx \mu + \sum_{k=1}^{K}\xi_{k,t}\gamma_k.$$

As in Fuest and Mittnik (2015), it is convenient to assume the existence of some $K < \infty$ such that:

$$\begin{aligned} \int_0^M \sum_{k=K+1}^{\infty} \gamma_k(m) x_t(m)\, \mathrm{d}m &= 0 \\ \Leftrightarrow \int_0^M \sum_{k=K+1}^{\infty} \sum_{l=1}^{\infty} \gamma_k(m)\gamma_l(m)\xi_{l,t}\, \mathrm{d}m &= 0 \\ \Leftrightarrow \int_0^M \sum_{k=K+1}^{\infty} \sum_{l=K+1}^{\infty} \gamma_k(m)\gamma_l(m)\xi_{l,t}\, \mathrm{d}m &= 0. \end{aligned}$$

In other words, we assume that the number of principal components that actually have an effect on the return (ARMA-part) or the variance (GARCH-part) is finite. As they are ordered by their contribution to the curves' variation, the leading K components are the ones that explain the variation best. However, there might still be dependencies between the first K and the remaining components. We (realistically) assume that these lead/lag effects of components $K+1, \ldots$ on the first K components' scores are negligible.

Estimation of the model can then be accomplished in three steps:

1. Estimation of the curved valued process x_t via an orthonormal FPC expansion:

$$\hat{x}_t = \hat{\mu} + \sum_{k=1}^{K} \hat{\xi}_{k,t}\hat{\gamma}_k,$$

   where the true values of K, μ and γ_k are unknown and the $\hat{\xi}_{k,t}$ are obtained via numerical integration (see Section 3.1).
2. Estimation of the ARMA-FunX parameters using the scores $\hat{\xi}_{k,t}$ for $k = 1, \ldots, K$ and $t = 1, \ldots, |\mathcal{T}|$ from Step 1 and the return data by Gaussian QML.
3. Gaussian QML estimation of the GARCH-FunX parameters using the scores $\hat{\xi}_{l,t}$ for $l = 1, \ldots, L$, $t = 1, \ldots, |\mathcal{T}|$ from Step 1 and the estimated errors from Step 2.

$$\begin{aligned} r_t &= \alpha_0 + \alpha r_{t-1} + \beta\epsilon_{t-1} + \epsilon_t + \sum_{k=1}^{K} b_k \hat{\xi}_{k,t-1}, \\ \epsilon_t &= \sigma_t e_t, \\ \log\sigma_t^2 &= \omega + \gamma \log\epsilon_{t-1}^2 + \delta \log\sigma_{t-1}^2 + \sum_{l=1}^{L} c_l \hat{\xi}_{l,t-1}, \end{aligned}$$

where $e_t \sim WN(0,1)$ and $\xi_{k,t}$ is the score of the k-th principal component of the functional PC representation of x_t.

Remark 2. *To ensure the stationarity of our process, we restrict our parameters as follows:*

- $|\alpha + \beta| < 1$
- $|\gamma + \delta| < 1$

Additionally, we employ the following assumptions:

- *We force past volatility to influence present volatility positively, so we choose $\gamma > 0$. (see Francq et al. (2013).)*
- *Past errors should positively influence present volatility, leading to the choice $\delta > 0$.*

We estimate the ARMA(1,1)-FunX parameters by means of non-linear least squares using the function **armaxfilter** from the MFE toolbox by Kevin Sheppard. The conditional working distribution of the logarithmic GARCH(1,1)-FunX is given by:

$$\epsilon_t | \mathcal{F}_{t-1}^{Y} \sim N(0, \exp(\omega + \gamma \log \epsilon_{t-1}^2 + \delta \log \sigma_{t-1}^2 + \int_0^M \rho(m) x_{t-1}(m) \, \mathrm{d}m)),$$

where $\mathcal{F}_{t-1}^{Y}$ denotes the information set consisting of past returns plus yield curve differences at $t-1$.

The Gaussian quasi-log-likelihood is then given by:

$$l(\boldsymbol{\epsilon}, \boldsymbol{x}; \omega, \gamma, \delta, \rho) = -\frac{1}{2} \sum_{t=2}^{|\mathcal{T}|} \left(\sigma_t^2 + \frac{\epsilon_t^2}{\sigma_t^2} \right),$$

where $\boldsymbol{\epsilon}$ is the vector of estimated errors from the ARMA(1,1)-FunX estimation and $\boldsymbol{x}$ is the "matrix" of yield curve differences.

Drawing the attention to the integrals in our model equations, we want to find finite representations of these infinite-dimensional terms. Employing the results of Section 3.1, we are able to estimate the integrals in our model equation, which are of the form[2]:

$$\int_0^M \lambda(s) x_t(s) \, \mathrm{d}s. \tag{4}$$

In order to do so, we consider the approximation of our process, x_t, by means of the first K components of the Karhunen–Loève representation,

$$x_t(s) \approx \mu(s) + \sum_{k=1}^{K} \xi_{k,t} \gamma_k(s),$$

where the true μ and the true γ_k are unknown. Furthermore, we assume that our weight function λ is representable by the eigenfunctions γ_k, which means we can approximate λ by:

$$\lambda(s) \approx \sum_{j=1}^{K} b_j \gamma_j(s).$$

Then, we can approximate: (4) by

$$\int_0^M \lambda(s) x_t(s) \, \mathrm{d}s \approx \int_0^M \sum_{j=1}^{K} b_j \gamma_j(s) (\mu(s) + \sum_{k=1}^{K} \xi_{k,t} \gamma_k(s)) \, \mathrm{d}s.$$

We obtain:

$$\sum_{k,j} \xi_{k,t} b_j \int_0^M \gamma_k(s) \gamma_j(s) \, \mathrm{d}s + \sum_{j=1}^{K} b_j \int_0^M \mu(s) \gamma_j(s) \, \mathrm{d}s.$$

Since the γ_k form an orthonormal basis of $\mathcal{L}^2([0, M])$, we have:

$$\sum_{k=1}^{K} \xi_{k,t} b_k + \sum_{j=1}^{K} b_j \int_0^M \mu(s) \gamma_j(s) \, \mathrm{d}s.$$

---

[2] To simplify notation, we write x_t instead of x_{t-1}.

Moreover, since $\mu = \mathbb{E}(x_t)$ is assumed to be constant for all $t \in \mathcal{T}$, we have:

$$\sum_{k=1}^{K} \xi_{k,t} b_k + \text{const},$$

where the estimators $\hat{\xi}_{k,t}$ of $\xi_{k,t}$ are obtained as described in Section 3.1.

As a consequence, the task of estimating the infinite-dimensional integral (4) translates into estimating $b_1, \ldots, b_k$. Moreover, note that estimating the return r_t for a centered process x_t or a possibly uncentered process x_t only differs in the constant α_0 or ω respectively in our model. Therefore, we can assume w.l.o.g. that $\mu = \mathbb{E}(x_t) = 0$. For a centered process x_t, we have const. $= 0$.

All in all, estimating our model Definition 1 simplifies to estimating the coefficients from the basic ARMA- and GARCH-parts and further $K + L$ parameters for the integral parts, where K, L can be chosen arbitrarily according to the application.

4. Results

In this section, we present an application of the model introduced in Section 3.2 to VaR forecasting for the EURUSD exchange rate.

4.1. Model Fit

We start by considering the model fit of our new ARMAFunX-GARCHFunX setup taking a classical ARMA-GARCH and an ARMAX-GARCHX with just the 2y-yield differential as the exogenous variable and Gaussian errors for both as benchmarks. Table 1 shows the parameter estimates for all three models. To assess the models' convergence, which was discussed in Section 3.2, we carried out a simulation study; For each setup, we simulated 10,000 paths of length 2903 using the initially estimated parameters and re-estimated these based on each path. Table 1 now additionally contains the means of these bootstrapped estimates and 95%-confidence intervals.

Considering the latter, we found that apparently the 2y-yield differential has no significant effect on the ARMAX-GARCHX setup, contradicting the industry's common belief.[3] However, for the functional setup, the situation looks much brighter even though there are parts without significance, as well. To reduce dimensionality and ensure the models' stabilities, we set all non-significant parameters (expect the intercepts) to zero and re-estimated all setups, which led to Table 2.[4] By evaluating the models' in-sample fit via AIC and BIC (Table 3), it can directly be seen that the ARMAFunX-GARCHFunX exhibited the lowest AIC and BIC, i.e., fit the data best, with the 2y-ARMAX-GARCHX coming in second. Restricting the models to significant parameters only (Table 4) improved the fit for all models, but the 2y-ARMAX-GARCHX.

[3] Traces of this assumption are scattered all over the Internet, but we restrain from quoting web pages.
[4] Note that we will work with the full models in the following, as explained in Section 4.2.

Table 1. Parameter estimates, bootstrapped means and bootstrapped 95%-confidence intervals using 10,000 simulated paths of length 2903.

	ARMA-GARCH			ARMAFunX-GARCHFunX			2y-ARMAX-GARCHX		
Parameters	**Estimate**	**Mean**	**Confidence Interval**	**Estimate**	**Mean**	**Confidence Interval**	**Estimate**	**Mean**	**Confidence Interval**
α_0	1.12×10^{-5}	3.34×10^{-5}	$[-5.12 \times 10^{-4}, 6.06 \times 10^{-4}]$	1.21×10^{-5}	1.59×10^{-5}	$[-1.71 \times 10^{-4}, 2.22 \times 10^{-4}]$	2.02×10^{-5}	2.08×10^{-5}	$[-3.24 \times 10^{-4}, 3.65 \times 10^{-4}]$
α	6.09×10^{-1}	1.51×10^{-3}	$[-9.37 \times 10^{-1}, 9.41 \times 10^{-1}]$	5.82×10^{-1}	5.10×10^{-1}	$[-1.99 \times 10^{-1}, 8.34 \times 10^{-1}]$	4.05×10^{-1}	4.03×10^{-1}	$[3.09 \times 10^{-1}, 4.91 \times 10^{-1}]$
β	-6.09×10^{-1}	-1.90×10^{-3}	$[-9.43 \times 10^{-1}, 9.38 \times 10^{-1}]$	-6.11×10^{-1}	-5.41×10^{-1}	$[-8.57 \times 10^{-1}, 1.68 \times 10^{-1}]$	-4.23×10^{-1}	1.62×10^{-3}	$[-9.67 \times 10^{-2}, 1.02 \times 10^{-1}]$
b_1	–	–	–	-1.48×10^{-3}	-1.47×10^{-3}	$[-2.33 \times 10^{-3}, -5.94 \times 10^{-4}]$	-1.28×10^{-2}	-1.48×10^{-3}	$[-8.93 \times 10^{-3}, 6.21 \times 10^{-3}]$
b_2	–	–	–	-1.46×10^{-4}	-1.80×10^{-4}	$[-1.91 \times 10^{-3}, 1.48 \times 10^{-3}]$	–	–	–
b_3	–	–	–	6.80×10^{-4}	6.77×10^{-4}	$[-1.91 \times 10^{-3}, 3.27 \times 10^{-3}]$	–	–	–
ω	-2.10×10^{-2}	-3.21×10^{-2}	$[-7.81 \times 10^{-2}, -6.78 \times 10^{-4}]$	-2.25×10^{-2}	-5.12×10^{-2}	$[-8.26 \times 10^{-2}, -1.40 \times 10^{-3}]$	-4.48×10^{-2}	-6.91×10^{-2}	$[-1.32 \times 10^{-1}, -1.62 \times 10^{-2}]$
γ	2.15×10^{-2}	2.11×10^{-2}	$[1.54 \times 10^{-2}, 2.69 \times 10^{-2}]$	2.05×10^{-2}	1.86×10^{-2}	$[1.27 \times 10^{-2}, 2.46 \times 10^{-2}]$	2.22×10^{-2}	1.97×10^{-2}	$[1.38 \times 10^{-2}, 2.60 \times 10^{-2}]$
δ	9.73×10^{-1}	9.72×10^{-1}	$[9.64 \times 10^{-1}, 9.80 \times 10^{-1}]$	9.74×10^{-1}	9.73×10^{-1}	$[9.66 \times 10^{-1}, 9.83 \times 10^{-1}]$	9.70×10^{-1}	9.70×10^{-1}	$[9.60 \times 10^{-1}, 9.81 \times 10^{-1}]$
c_1	–	–	–	6.50×10^{-2}	6.61×10^{-2}	$[2.85 \times 10^{-2}, 1.06 \times 10^{-1}]$	-2.06×10^{-1}	-2.05×10^{-1}	$[-5.48 \times 10^{-1}, 1.22 \times 10^{-1}]$
c_2	–	–	–	-8.93×10^{-2}	-9.12×10^{-2}	$[-1.60 \times 10^{-1}, -2.38 \times 10^{-2}]$	–	–	–
c_3	–	–	–	1.07×10^{-1}	1.01×10^{-1}	$[-5.41 \times 10^{-2}, 2.56 \times 10^{-1}]$	–	–	–

Table 2. Parameter estimates, bootstrapped means and bootstrapped 95%-confidence intervals using 10,000 simulated paths of length 2903 restricting the models from Table 1 to significant parameters only.

	ARMA-GARCH			ARMAFunX-GARCHFunX			2y-ARMAX-GARCHX		
Parameters	**Estimate**	**Mean**	**Confidence Interval**	**Estimate**	**Mean**	**Confidence Interval**	**Estimate**	**Mean**	**Confidence Interval**
α_0	3.18×10^{-5}	3.56×10^{-5}	$[-4.44 \times 10^{-4}, 5.05 \times 10^{-4}]$	3.19×10^{-5}	3.27×10^{-5}	$[-3.50 \times 10^{-4}, 4.06 \times 10^{-4}]$	3.11×10^{-5}	3.04×10^{-5}	$[-3.97 \times 10^{-4}, 4.55 \times 10^{-4}]$
α	–	–	–	–	–	–	5.98×10^{-3}	5.68×10^{-3}	$[-3.51 \times 10^{-2}, 4.50 \times 10^{-2}]$
β	–	–	–	–	–	–	–	–	–
b_1	–	–	–	-1.21×10^{-3}	-1.21×10^{-3}	$[-2.06 \times 10^{-3}, -3.63 \times 10^{-4}]$	–	–	–
b_2	–	–	–	–	–	–	–	–	–
b_3	–	–	–	–	–	–	–	–	–
ω	-1.78×10^{-2}	-2.41×10^{-2}	$[-7.00 \times 10^{-2}, 5.07 \times 10^{-3}]$	-2.21×10^{-2}	-3.04×10^{-2}	$[-8.08 \times 10^{-2}, 1.33 \times 10^{-3}]$	-2.37×10^{-2}	-3.60×10^{-2}	$[-8.24 \times 10^{-2}, -2.31 \times 10^{-3}]$
γ	2.43×10^{-2}	2.57×10^{-2}	$[1.93 \times 10^{-2}, 3.24 \times 10^{-2}]$	2.33×10^{-2}	2.37×10^{-2}	$[1.73 \times 10^{-2}, 3.05 \times 10^{-2}]$	2.11×10^{-2}	2.06×10^{-2}	$[1.51 \times 10^{-2}, 2.65 \times 10^{-2}]$
δ	9.70×10^{-1}	9.68×10^{-1}	$[9.58 \times 10^{-1}, 9.76 \times 10^{-1}]$	9.71×10^{-1}	9.70×10^{-1}	$[9.60 \times 10^{-1}, 9.78 \times 10^{-1}]$	9.73×10^{-1}	9.72×10^{-1}	$[9.64 \times 10^{-1}, 9.81 \times 10^{-1}]$
c_1	–	–	–	5.47×10^{-2}	5.37×10^{-2}	$[1.36 \times 10^{-2}, 9.31 \times 10^{-2}]$	–	–	–
c_2	–	–	–	-1.16×10^{-1}	-1.17×10^{-1}	$[-1.84 \times 10^{-1}, -4.77 \times 10^{-2}]$	–	–	–
c_3	–	–	–	–	–	–	–	–	–

Table 3. logL, AIC and BIC, corresponding to the full models from Table 1.

Model	logL	AIC	BIC
ARMA-GARCH	10,826	−21,639	−21,604
ARMAFunX-GARCHFunX	10,868	−21,711	−21,640
2y-ARMAX-GARCHX	10,841	−21,666	−21,619

Table 4. logL, AIC and BIC, corresponding to the restricted models from Table 2.

Model	logL	AIC	BIC
ARMA-GARCH	10,832[5]	−21,656	−21,632
ARMAFunX-GARCHFunX	10,863	−21,712	−21,670
2y-ARMAX-GARCHX	10,829	−21,649	−21,619

4.2. VaR Backtesting

In the next step, we wanted to assess the out-of-sample-fit of our setups and investigate further whether the yield curve differential can effectively improve risk forecasts for the EURUSD exchange rate. For computational simplicity and to ensure better comparability, we shall from now on work with the full models depicted in Table 1 to avoid identifying the best model for each window separately. In particular, we shall apply the well-known methods from Christoffersen (1998).[6]

Now, using a 500-day rolling window, we estimate the full models from Table 1 and calculate one-day ahead VaR forecasts via:

$$\widehat{\mathrm{VaR}}_{t|t-1}(p) = \hat{\mu}_t + \Phi^{-1}(p)\hat{\sigma}_t$$

for $1-p \in \{99\%,\ 97.5\%,\ 95\%\}$, the standard normal cumulative distribution function Φ and with $\hat{\mu}_t$ being the estimated mean obtained from the ARMA parts and $\hat{\sigma}_t$ the volatility estimated via GARCH, respectively.

This leaves us with 2404 VaR forecasts for each model. Table 5 depicts the percentage of violations and the likelihood ratio tests for unconditional coverage, independence and conditional coverage as proposed by Christoffersen (1998). Figure 4 shows the standard 5%-VaR forecasts compared to the actual log-returns for all three models.

---

[5] The peculiarity of having a higher logL for the nested model in comparison to the full model of Table 3 arises due to using a two-step procedure instead of estimating jointly.

[6] Although, since then, various alternative backtests have been established as, e.g., in Ziggel et al. (2014) or Wied et al. (2016). However, such new approaches would deviate too much from the core idea of the present paper, which is why we stick to the classical procedure of Christoffersen (1998).

Table 5. VaR prediction performance with a window size of 500 days for ARMA-GARCH, ARMAFunX-GARCHFunX and 2y-ARMAX-GARCHX for $1-p \in \{99\%, 97.5\%, 95\%\}$. Bold numbers denote significant values at the 5%-level.

Model	p	% Viol.	LR_{uc}	LR_{ind}	LR_{cc}
ARMA-GARCH	1%	1.46×10^{-02}	$\mathbf{4.42 \times 10^{00}}$	1.03×10^{00}	5.49×10^{00}
	2.5%	2.79×10^{-02}	7.84×10^{-01}	6.19×10^{-01}	1.46×10^{00}
	5%	5.20×10^{-02}	1.99×10^{-01}	4.15×10^{-02}	3.48×10^{-01}
ARMAFunX-GARCHFunX	1%	1.50×10^{-02}	$\mathbf{5.21 \times 10^{00}}$	1.10×10^{00}	$\mathbf{6.34 \times 10^{00}}$
	2.5%	2.83×10^{-02}	1.02×10^{00}	$\mathbf{3.96 \times 10^{00}}$	5.04×10^{00}
	5%	5.07×10^{-02}	2.82×10^{-02}	$\mathbf{7.21 \times 10^{00}}$	$\mathbf{7.34 \times 10^{00}}$
2y-ARMAX-GARCHX	1%	1.58×10^{-02}	$\mathbf{6.96 \times 10^{00}}$	1.22×10^{00}	$\mathbf{8.21 \times 10^{00}}$
	2.5%	2.83×10^{-02}	1.02×10^{00}	5.66×10^{-01}	1.65×10^{00}
	5%	5.20×10^{-02}	1.99×10^{-01}	1.23×10^{00}	1.53×10^{00}

In the following, we briefly recall the basic idea of these statistics.

- Firstly, we have the unconditional coverage (uc) test, which assumes the independence of the violations and tests the hypothesis that the empirical percentage of violations is equal to the expected p.
- The independence test (ind) checks for the independence of violations or detects clustering, respectively.
- Finally, there is the conditional coverage (cc) test that compares the empirical percentage of violations and the expected percentage as the unconditional coverage test does, but considers a possible dependence structure of the violations. We may treat it as a combination of the former two tests.
- The statistics LR_{uc} and LR_{ind} for the uc test and the ind test are χ^2-distributed with one degree of freedom, whereas the LR_{cc}, the one for the cc test, is χ^2-distributed with two degrees of freedom.

Summing things up, Table 5 does not show an improvement in using the yield curve data, apart from the 5%-VaR, where our functional model delivers a better unconditional coverage statistics than the benchmark, which in this case means that the expected and the empirical numbers of violations are very close. Even then, the ARMAXFunX-GARCHFunX exhibits a clustering of violations and therefore a worse independence test and conditional coverage test statistics.

As mentioned before, the LR_{cc} can be seen as a combination of the LR_{uc} and the LR_{ind}. Therefore, in most cases, LR_{cc} had a high value whenever LR_{uc} had one as well. The exceptions were the cases in which the independence test statistics was unusually high. High values of LR_{ind} hinted towards a clustering of violations. This phenomenon can especially be observed at the 5%-VaR from the ARMAFunX-GARCHFunX, where the unconditional coverage test statistics was rather small and well below the significance level, but because of the clustering, we had a significant LR_{cc}.

Finally, we want to note that we carried out robustness checks regarding the actual number of principal components for the functional setup. We experienced an increase of clustering of violations while adding more principal components to our model, with the change from two to three principal components having the biggest impact on the LR_{ind}. Note that, since LR_{cc} can be seen as a combination of LR_{uc} and LR_{ind}, this influenced the conditional coverage, as well. This phenomenon appeared to be capped at three principal components, which was observable in particular for the 5%-VaR. Interestingly, using two principal components instead of just one achieved worse results for all $p \in \{1\%, 2.5\%, 5\%\}$. The performances of the models with three and four principal components did not differ much, apart from a noticeably smaller LR_{ind} for the 2.5%-VaR. Finally, we want to point out that our model, which used three principal components, had the smallest LR_{uc} for the 5%-VaR compared to incorporating 1, 2 or 4 principal components.

Consequently, we can further confirm that employing yield curve information seemed to be able to improve VaR forecasting for FX rates in some situations.

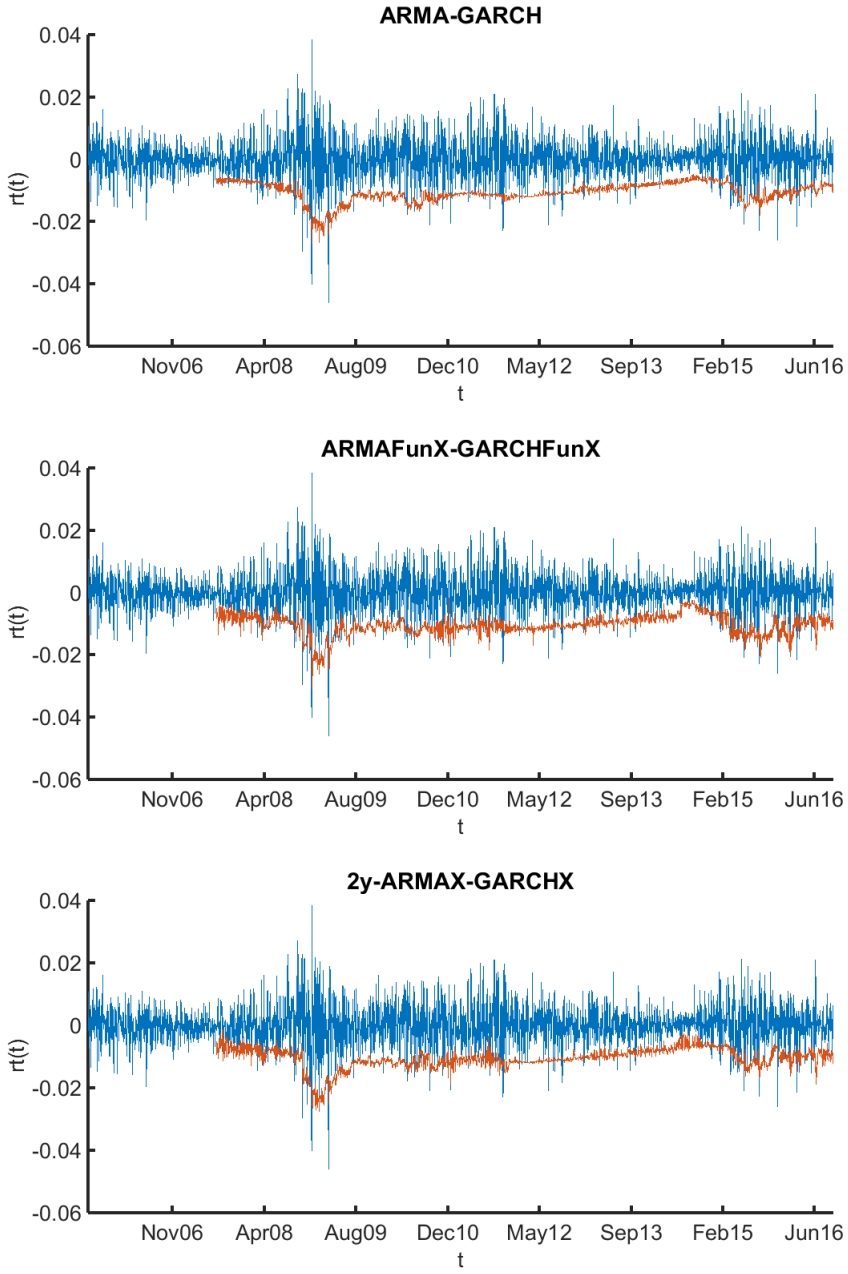


Figure 4. ARMA-GARCH, ARMAFunX-GARCHFunX, 2y-ARMAX-GARCHX: 1-day prediction: 5%-VaR.

5. Discussion

This paper introduces a new approach in implementing macroeconomic data, in particular yield curve information in FX rate models. As we have shown in detail in Section 1, the general idea of such an approach can be found throughout the literature. However, according to the efficient market hypothesis, interest rate differentials cannot be used for forecasting FX rates in a sensible fashion. We do not argue this and will not delve into a discussion about the failure of the uncovered interest rate parity. Therefore, we utilize our ARMAXFunX-GARCHFunX for forecasting risk, in particular the VaR of the EURUSD rate.

The way our model sticks out and improves on older approaches is the use of functional principal component analysis to implement the whole shape of the yield curve as an exogenous variable.

Applying principal components in this way has been done in a similar fashion before. We refer to Fuest and Mittnik (2015), who presented the GARCH-FunXL model and used the functional approach to model asset price volatility. The information contained in the stock exchange's limit order book is extracted by means of functional principal component analysis, and its impact on asset price volatility is analyzed. The work in Fuest and Mittnik (2015) reported an improved forecasting performance compared to models without liquidity impact.

We extend their GARCH-FunXL setup by allowing for an additional ARMAXFunX part and invoke this more generalized setup to forecast the VaR of FX rates.

The data used for our yield curves consist of overnight index swaps on the EONIA and the EONIA for shortest maturity and the US treasury yields, with the Fed Funds rate for the shortest maturity. Interpolation and taking differentials[7] gives us the data matrix of the yield curve differentials used for our analysis. We transform this matrix of yield differentials into functional data, which means that we get a stochastic process $(x_t)_{t\in\mathcal{T}}$, where each x_t maps maturities to yields. The theory for functional principal component analysis, which is needed to understand the basic idea of the model, is introduced in Section 3.

Although our functional model exhibits clustering of violations, we experienced a smaller unconditional coverage statistics for the 5%-VaR, which means that the estimated number of violations was very close to the expected number. Summing up, we confirm that employing yield curve information is able to improve VaR forecasting for FX rates in some situations. This established connection between yield curve data and FX risk goes with the findings in the literature: Dominguez (1998) modeled FX rate volatility via a GARCH(1,1) setup that incorporated, among other variables, overnight index rate differentials corresponding to the FX rates in question as exogenous variables, an approach similar to ours, and found the interest rate differentials to have a significant impact on FX rate volatility. We further mention Neely (1999), who analyzed the connection of realignment and conditional volatility in target zone exchange rate systems via a jump-diffusion GARCH model, where they used interest rate differentials and a proxy for the domestic yield curve to model a country-specific realignment probability or jump probability. Letting the jump-intensity vary with dependence on the interest rates in this way allows the model to detect realignments before they happen. With this setup, the precision of forecasts can be improved. As for a connection between macroeconomic volatility and FX rate volatility, we recall Morana (2009). He used principal component analysis to extract the relevant factors for macroeconomic volatility and FX rate volatility, respectively, and showed that there is an influence from macro to FX, especially for the long-term. Among other macro variables, he used short-term interest rates, in particular three-month government bills. An improvement in out-of-sample forecasting performance, using this setup, was reported. This result supports our findings, since we confirm that the yield curve information contains valuable information for forecasting FX rate volatility, which is confirmed by Dunis and Huang (2002), as mentioned in the Introduction. Finally, we mention Bauwens and Sucarrat (2010), who compared the forecasting performance of several models of exchange rate volatility in the GETS (general-to-specific) approach and did find evidence of an improvement by the inclusion of macro variables, such as interest rates.

Summing up, this points towards an influence of macroeconomic or, in particular, yield curve data on modeling FX risk. One property that most basic models of FX rates or FX rate volatility share though is that an increase in the exogenous information used always leads to a higher number of variables. Implementing the exogenous data as functional data and using the power of functional

---

[7] We also take daily differences to ensure the stationarity of our yield curve process.

principal component analysis allow us to contain almost all of the information contained within any dataset by about three principal components, no matter the size of the matrix.

Coming to a conclusion, we find that the ARMAXFunX-GARCHFunX enriches the family of ARMAX-GARCHX models by allowing the implementation of the information contained in a data matrix via its principal components or its principal component scores, respectively, no matter how big it is or, in our case, how many maturities we want to implement, a feature that is, as we have seen, desirable when modeling and forecasting exchange rate volatility.

However, when it comes to actually using our newly-proposed forecasting setup, we see just some minor improvements on a classical ARMAX-GARCHX benchmark containing just the 2y-yield as the exogenous variable for, e.g., the 5%-VaR unconditional coverage, but are not able to detect a more general advantage of the functional setup. However, considering different lags and (fat-tailed) error distributions might be a fruitful start for future research in this area and might justify the usage of, for example, more maturities, as discussed above.

Furthermore, we want to stress the fact that the data in the present study are based on first-step estimates of the yield curve, which are evaluated at a rather small number of maturities. Recall the construction of our yield curves: The EUR- and the US-yield curves were equivalently constructed by using the available data as presented in Section 2 as follows. For each observation date, we used the available data points (maturities) as nodes for spline interpolation. With this procedure, a yield curve consisting of 121 maturities for EUR, as well as for the US is obtained. Taking the componentwise difference of both yield curves, which are basically data matrices, we acquire the yield curve differential of the EUR- and the US-yield curve.

Such estimates might be over-simplifying in the present context. Using the primary information on the YTMs of single securities underlying the curve estimates, additional modes of variation may be uncovered that are potentially useful to predict FX returns.

Author Contributions: The general idea and conceptualization came from H.F. He was responsible for supervision, as well as project administration. A.F. provided support concerning the methodology and was responsible for the functional data part in the literature review. H.P. carried out the main work, including data curation, formal analysis, investigation, methodology, software, validation, visualization, writing of the original draft and writing review and editing.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

Ang, Andrew, and Chen Joseph. 2010. Yield Curve Predictors of Foreign Exchange Returns. Paper presented at AFA 2011 Denver Meetings, Denver, CO, USA, January 7–9.

Aue, Alexander, Diogo Dubart Norinho, and Siegfried Hörmann. 2015. On the prediction of stationary functional time series. *Journal of the American Statistical Association* 110: 378–92. [CrossRef]

Baillie, Richard T., and Tim Bollerslev. 1991. Intra-day and inter-market volatility in foreign exchange rates. *The Review of Economic Studies* 58: 565–85. [CrossRef]

Baillie, Richard T., Tim Bollerslev, and Hans Ole Mikkelsen. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74: 3–30. [CrossRef]

Bauwens, Luc, and Genaro Sucarrat. 2010. General-to-specific modelling of exchange rate volatility: A forecast evaluation. *International Journal of Forecasting* 26: 885–907. [CrossRef]

Benavides, Guillermo, and Carlos Capistràn. 2012. Forecasting exchange rate volatility: The superior performance of conditional combinations of time series and option implied forecasts. *Journal of Empirical Finance* 19: 627–39. [CrossRef]

Bjørnland, Hilde C., and Håvard Hungnes. 2006. The importance of interest rates for forecasting the exchange rate. *Journal of Forecasting* 25: 209–21. [CrossRef]

Bosq, Denis. 2000. *Linear Processes in Function Spaces: Theory and Applications*. Berlin: Springer.

Brockhaus, Sarah, Andreas Fuest, Andreas Mayr, and Sonja Greven. 2017. Signal regression models for location, scale and shape with an application to stock returns. *Journal of the Royal Statistical Society/Series C (Applied Statistics)* 67: 665–86. [CrossRef]

Campbell, John Y., and Richard H. Clarida. 1987. The dollar and real interest rates. *Carnegie-Rochester Conference Series on Public Policy* 27: 103–39. [CrossRef]

Chen, Yu-chin, and Kwok Ping Tsang. 2013. What does the yield curve tell us about exchange rate predictability? *Review of Economics and Statistics* 95: 185–205. [CrossRef]

Chinn, Menzie D., and Richard A. Meese. 1995. Banking on currency forecasts: How predictable is change in money? *Journal of International Economics* 38: 161–78. [CrossRef]

Christiansen, Charlotte, Maik Schmeling, and Andreas Schrimpf. 2012. A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics* 27: 956–77. [CrossRef]

Christoffersen, Peter F. 1998. Evaluating interval forecasts. *International Economic Review* 39: 841–62. [CrossRef]

Coughlin, Cletus C., and Kees Koedijk. 1990. What do we know about the long-run real exchange rate? *Federal Reserve Bank of St. Louis Review* 72. [CrossRef]

Diebold, Francis X., and Canlin Li. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* 130: 337–64. [CrossRef]

Dominguez, Kathryn M. 1998. Central bank intervention and exchange rate volatility. *Journal of International Money and Finance* 17: 161–90. [CrossRef]

Dunis, Christian L., and Xuehuan Huang. 2002. Forecasting and trading currency volatility: An application of recurrent neural regression and model combination. *Journal of Forecasting* 21: 317–54. [CrossRef]

Edison, Hali J., and B. Dianne Pauls. 1993. A re-assessment of the relationship between real exchange rates and real interest rates: 1974–1990. *Journal of Monetary Economics* 31: 165–87. [CrossRef]

European Banking Federation (FBE). 2008. *EONIA Swap Index: The Derivatives Market Reference Rate for the Euro*. Brussels: European Banking Federation.

Fama, Eugenel. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25: 383–417. [CrossRef]

Filipović, Damir, and Anders B. Trolle. 2013. The term structure of interbank risk. *Journal of Financial Economics* 109: 707–33. [CrossRef]

Francq, Christian, Olivier Wintenberger, and Jean-Michael Zakoïan. 2013. GARCH models without positivity constraints: Exponential or log GARCH? *Journal of Econometrics* 177: 34–46. [CrossRef]

Frankel, Jeffrey A., and Andrew K. Rose. 1995. Empirical research on nominal exchange rates. In *Handbook of International Economics*. Edited by G. M. Grossman and Kenneth Rogoff. Amsterdam: North Holland Publishing Company, vol. 3, pp. 1689–729.

Fuest, Andreas, and Stefan Mittnik. 2015. Modeling Liquidity Impact on Volatility: A GARCH-FunXL Approach. Avaliable online: http://dx.doi.org/10.2139/ssrn.3038947 (accessed on 16 August 2018).

Grisse, Christian, and Thomas Nitschka. 2015. On financial risk and the safe haven characteristics of swiss franc exchange rates. *Journal of Empirical Finance* 32: 153–64. [CrossRef]

Han, Heejon. 2015. Asymptotic properties of GARCH-X processes. *Journal of Financial Econometrics* 13: 188–221. [CrossRef]

Han, Heejon, and Dennis Kristensen. 2014. Asymptotic theory for the QMLE in GARCH-X models with stationary and nonstationary covariates. *Journal of Business & Economic Statistics* 32: 416–29.

Hannan, Edward J., William T. M. Dunsmuir, and Manfred Deistler. 1980. Estimation of vector ARMAX models. *Journal of Multivariate Analysis* 10: 275–95. [CrossRef]

Hansen, Peter Reinhard, Zhuo Huang, and Howard Howan Shek. 2012. Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics* 27: 877–906. [CrossRef]

Hörmann, Siegfried, Lajos Horváth, and Ron Reeder. 2013. A functional version of the ARCH model. *Econometric Theory* 29: 267–88. [CrossRef]

Hörmann, Siegfried, and Piotr Kokoszka. 2010. Weakly dependent functional data. *The Annals of Statistics* 38: 1845–84. [CrossRef]

Hörmann, Siegfried, and Piotr Kokoszka. 2012. Functional time series. In *Handbook of Statistics: Time Series Analysis: Methods and Applications*, 1st ed. Edited by Tata Subba Rao, Suhasini Subba Rao and C. R. Rao. Amsterdam: North Holland Publishing Company, vol. 30, p. 157.

Hull, John, and Alan White. 1998. Value at risk when daily changes in market variables are not normally distributed. *Journal of Derivatives* 5: 9–19. [CrossRef]

Hull, John, and Alan White. 2012. Libor VS OIS: The derivatives discounting dilemma. *Journal of Investment Management* 11: 14–27. [CrossRef]

Hyndman, Rob J., and Han Lin Shang. 2009. Forecasting functional time series. *Journal of the Korean Statistical Society* 38: 199–211. doi:10.1016/j.jkss.2009.06.002. [CrossRef]

Ichiue, Hibiki, and Kentaro Koyama. 2011. Regime switches in exchange rate volatility and uncovered interest parity. *Journal of International Money and Finance* 30: 1436–50. [CrossRef]

Klepsch, Johannes, Claudia Klüppelberg, and Taoran Wei. 2017. Prediction of functional ARMA processes with an application to traffic data. *Econometrics and Statistics* 1: 128–49. [CrossRef]

Kočenda, Evžen, and Tigran Poghosyan. 2009. Macroeconomic sources of foreign exchange risk in new EU members. *Journal of Banking & Finance* 33: 2164–73.

Kowal, Daniel R., David S. Matteson, and David Ruppert. 2017. A Bayesian multivariate functional dynamic linear model. *Journal of the American Statistical Association* 112: 733–44. [CrossRef]

Kowal, Daniel R., David S. Matteson, and David Ruppert. 2017. Functional autoregression for sparsely sampled data. *Journal of Business & Economic Statistics*. [CrossRef]

Kuester, Keith, Stefan Mittnik, and Marc S. Paolella. 2006. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics* 4: 53–89. [CrossRef]

Markiewicz, Agnieszka. 2012. Model uncertainty and exchange rate volatility. *International Economic Review* 53: 815–44. [CrossRef]

Meese, Richard, and Kenneth Rogoff. 1983. Empirical exchange rate models of the seventies-Do they fit out of sample? *Journal of International Economics* 14: 3–24. [CrossRef]

Meese, Richard, and Kenneth Rogoff. 1988. Was it real? The exchange rate-interest differential relation over the modern floating-rate period. *The Journal of Finance* 43: 933–48. [CrossRef]

Mittnik, Stefan, and Marc S. Paolella. 2000. Conditional density and value-at-risk prediction of Asian currency exchange rates. *Journal of Forecasting* 19: 313–33. [CrossRef]

Morana, Claudio. 2009. On the macroeconomic causes of exchange rate volatility. *International Journal of Forecasting* 25: 328–50. [CrossRef]

Neely, Christopher J. 1999. Target zones and conditional volatility: The role of realignments. *Journal of Empirical Finance* 6: 177–92. [CrossRef]

Ramsay, James O. 2014. FDA Toolbox. Avaiable online: http://www.psych.mcgill.ca/misc/fda/index.html (accessed on 16 August 2018).

Ramsay, James O., and Bernard W. Silverman. 2005. *Functional Data Analysis*. New York: Springer.

Sheppard, Kevin. 2013. MFE Toolbox. Avaiable online: https://www.kevinsheppard.com/MFE_Toolbox (accessed on 16 August 2018).

Sucarrat, Genaro, Steffen Grønneberg, and Alvaro Escribano. 2016. Estimation and inference in univariate and multivariate log-GARCH-x models when the conditional density is unknown. *Computational Statistics & Data Analysis* 100: 582–94.

Vilasuso, Jon. 2002. Forecasting exchange rate volatility. *Economic Letters* 76: 59–64. [CrossRef]

West, Kenneth D., Hali J. Edison, and Dongchul Cho. 1993. A utility-based comparison of some models of exchange rate volatility. *Journal of International Economics* 35: 23–45. [CrossRef]

Wied, Dominik, George N.F. Weiß, and Daniel Ziggel. 2016. Evaluating value-at-risk forecasts: A new set of multivariate backtests. *Journal of Banking & Finance* 72: 121–32.

Ziggel, Daniel, Tobias Berens, Geroge N. F. Weiß, and Dominis Wied. 2014. A new set of improved value-at-risk backtests. *Journal of Banking & Finance* 48: 29–41.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 19ten Mai, 2020

Henry Alvorado Port