Quantitative Results for Multivariate Models

High-Dimensional Central Limit Theorems and Structural Modelling of Electricity Markets

Lota Copić

PhD Dissertation

AARHUS UNIVERSITY

danske commodities

Quantitative Results for Multivariate Models: High-Dimensional Central Limit Theorems and Structural Modelling of Electricity Markets

PhD dissertatioin by
Lota Copić

Affiliated with:

Aarhus University
Department of Mathematics
Ny Munkegade 118
8000 Aarhus C, Denmark

Danske Commodities
Risk Management
Kalkværksvej 16
8000 Aarhus C, Denmark

Supervisors:
Prof. Andreas Basse-O'Connor & Prof. Jan Pedersen, *Aarhus University*
Tor Bonde & Mikkel Graversen, *Danske Commodities*

Submitted to the Graduate School of Natural Sciences, Aarhus, December 2025.

Raviju, za beskrajnu ljubav, podršku i inspiraciju.

Abstract

This thesis investigates two distinct topics in modern statistics: Gaussian approximation theory in high dimensions and structural modelling of electricity markets.

The first research topic studies high-dimensional Gaussian limits and quantitative bounds on the associated approximation error. In high-dimensional regimes, a central challenge is obtaining error bounds that explicitly track the dependence on both the sample size n and the dimension d, rather than treating d as fixed. Working within the Malliavin–Stein framework, we derive bounds on distances between a d-dimensional random vector whose components admit Wiener–Itô chaos expansions and a corresponding Gaussian vector. These bounds are explicit in their dependence on d up to a factor reflecting possible near-singularity of the covariance matrix of the Gaussian limit. Applications of our bounds yield fully explicit (n, d)-dependent results in two settings: quantitative Breuer–Major theorems for high-dimensional vectors of non-affine functionals of multivariate stationary Gaussian sequences, and quantitative Gaussian approximation bounds for the central limit theorem term arising in parameter estimation for stable $\mathrm{VAR}(p)$ processes.

The second research topic studies structural modelling of day-ahead electricity prices in wholesale markets with high renewable infeed. We develop a structural model that incorporates renewable generation into a multi-fuel bid-stack framework. We treat renewables as priority dispatch and capture their effect through residual demand, effectively adjusting the demand curve to account for renewable output. Under standard short-run assumptions (price-inelastic demand and market equilibrium), we obtain a general pricing representation and highlight two special cases: a single-fuel market and a multi-fuel setting with proportional fuel prices, under which pricing reduces to a parsimonious multiplicative form. Empirically, using German day-ahead data from 2016–2023, we estimate a library of structural specifications within this framework and evaluate performance via rolling-origin out-of-sample testing, benchmarking against standard machine-learning baselines and a simple reduced-form model. A theory-consistent one-fuel specification driven by residual demand, gas prices, and carbon emissions allowance prices delivers the most robust out-of-sample performance.

Resumé

Denne afhandling handler om to problemer i moderne statistik: sandsynlighedsteoretisk approksimationsteori i høje dimensioner og strukturel modellering, der balancerer prædiktiv evne med gennemsigtighed og praktisk anvendelighed.

Det første problem undersøger højdimensionelle Gaussiske grænseværdier og kvantitative begrænsninger på den tilhørende approksimationsfejl. I højdimensionelle systemer er det en central udfordring at opnå approksimationsbegrænsninger, der eksplicit viser afhængigheden af både stikprøvestørrelsen n og dimensionen d, i stedet for at betragte d som fast. Inden for rammerne af Malliavin–Stein-metoden udleder vi begrænsninger på afstande mellem en d-dimensionel stokastisk vektor, hvis komponenter har Wiener–Itô kaos-udviklinger, og en tilsvarende Gaussisk vektor. Disse begrænsninger er eksplicitte i deres afhængighed af d, bortset fra en faktor, der afspejler en eventuel næsten-singularitet af kovariansmatricen for den Gaussiske grænseværdi. Anvendelserne giver, så vidt vi ved, nye resultater med fuldt eksplicit (n, d)-afhængighed i to situationer. Disse inkluderer kvantitative Breuer–Major-sætninger for højdimensionelle vektorer af ikke-affine funktionaler af multivariate stationære Gaussiske sekvenser, samt kvantitative Gaussiske approksimationsbegrænsninger for det centrale grænseværdisætnings-led, der opstår ved parameterestimering i stabile VAR(p)-processer.

Det andet problem undersøger strukturel modellering af day-ahead elpriser i markeder med høj andel af vedvarende energi. Vi udvikler en strukturel model, der inkorporerer vedvarende energiproduktion i en ramme baseret på en udbudskurve med flere fossile brændsler (multi-fuel bid-stack). Vi behandler vedvarende energi som prioriteret produktion og fanger dens effekt gennem residualefterspørgslen, hvilket effektivt justerer efterspørgselskurven for at tage højde for produktionen fra vedvarende kilder. Under standardantagelser (prisuelastisk efterspørgsel og markedsligevægt) udleder vi en generel repræsentation for prisfastsættelsen og fremhæver to specialtilfælde: et marked med ét brændsel og en multi-brændsels-setting med proportionale brændselspriser, hvor prisfastsættelsen reduceres til en simpel multiplikativ form. Empirisk estimerer vi en sammenfatning af strukturelle specifikationer inden for denne ramme ved brug af tyske day-ahead data fra 2016–2023. Vi evaluerer modellernes performance gennem rolling-origin out-of-sample tests og benchmarker dem mod standard maskinlærings-baselines samt en simpel reduceret form model. En teorikonsistent model baseret på ét brændsel, drevet af residualefterspørgsel, gaspriser og priser på CO2-kvoter, leverer den mest robuste præstation out-of-sample.

Contents

Preface

The present dissertation is the result of my three-year PhD studies at Aarhus University. It comprises two research papers:

Paper A Quantitative Bounds for High-Dimensional Random Vectors on Gaussian Spaces

Paper B Fitting Structural Models in Electricity Markets

Paper A is written in collaboration with Andreas Basse-O'Connor and David Kramer-Bang, and is in preparation for submission. Paper B is written jointly with my supervisors Andreas Basse-O'Connor, Jan Pedersen, and Tor Bonde. The data-analysis section was included in my progress report mid-way through my studies. This paper has been submitted to a journal.

The first half of my PhD focused on an applied problem: developing a simple yet useful model linking electricity price behaviour to its underlying drivers. The second half shifted focus to a theoretical direction, exploring the Malliavin–Stein method and its applications to convergence rates in high-dimensional limit theorems.

Reflecting over the past three years, I would like to express my deepest gratitude. First and foremost, I thank my university supervisors, Jan Pedersen and Andreas Basse-O'Connor, for their invaluable support throughout this journey. I am equally grateful to Innovation Fund Denmark for funding this research and to Danske Commodities for making this PhD opportunity possible. My sincere thanks also go to my company supervisors Tor Bonde and Mikkel Graversen, and my Risk Modelling team for their constant encouragement and inclusion, despite my parallel academic track.

I am grateful to my fellow PhD students and post-docs at Aarhus University for lasting friendships that have brightened this journey. I extend special thanks to Giovanni Peccati and his Luxembourg research group for their warm welcome, generous time, and invaluable discussions, which profoundly enriched my engagement with the topics of Paper A and my overall research experience. Special thanks to David for always making time, whether for mathematical questions, long calculations, or running discussions, and for his inspiring dedication, which made collaborating with him both motivating and fulfilling.

Finally, no words could ever fully express my gratitude to my family–in every sense of the word. To my core, those on the other side of the continent, to my dearest friends, and to Martin: your support and encouragement, regardless of the distance, have been the foundation of this journey. I leave this here, knowing that words fall short of what you mean to me.

Thank you!

Introduction

This chapter lays the necessary groundwork for the two foundational papers, Paper A and Paper B, that constitute this thesis. Given that these papers address distinct topics and employ different methodologies, this introductory chapter is accordingly divided into two dedicated parts: Section 1.1 and Section 1.2.

In Section 1.1, the research topic concerns limit theorems for high-dimensional random vectors on Gaussian spaces. The main problem studied within this topic is how collections of many random variables behave jointly as the dimension increases, and the results provide tools for deriving quantitative approximation bounds. Paper A focuses on quantitative bounds for high-dimensional Gaussian approximations, which are essential in understanding the rates of convergence in high-dimensional central limit theorems.

Section 1.2 examines the modelling of wholesale electricity markets. It outlines the merit-order principle and demonstrates how this mechanism gives rise to the structural models that characterise day-ahead price formation. Paper B proposes and empirically assesses a tractable structural framework for day-ahead electricity prices, combining theoretical modelling with data-driven analysis to evaluate the model's ability to explain observed market dynamics.

1.1 High-Dimensional Gaussian Approximations

Central limit theorems (CLTs) are fundamental to probability theory. In its classical form, the CLT asserts that a suitably normalised sum converges in distribution to a Gaussian limit. Indeed, let $(X_i)_{i\geqslant 1}$ be independent and identically distributed (i.i.d.) random variables with mean zero and finite variance σ^2, and define $S_n = n^{-1/2}\sum_{i=1}^n X_i$. Then the CLT yields that $S_n \xrightarrow{d} Z$, where Z follows a Gaussian distribution with mean zero and variance σ^2, denoted $Z \sim \mathcal{N}(0, \sigma^2)$. In other words, for sufficiently large n, the sum S_n is approximately Gaussian, and the Gaussian variable Z serves as its *Gaussian (normal) approximation*. This naturally leads to the following question:

> **Question 1.** How accurate is this Gaussian approximation? More precisely, can we quantify the error incurred by replacing S_n with its Gaussian approximation Z?

To answer Question 1, the standard approach is to adopt a distance $d_{\mathscr{H}}$ between the two probability distributions and establish a quantitative (upper) bound of the form

$$d_{\mathscr{H}}(S_n, Z) := \sup_{h\in\mathscr{H}} |\mathbb{E}[h(S_n)] - \mathbb{E}[h(Z)]| \leqslant f(n), \tag{1.1}$$

where $\mathscr{H}$ is a rich class of test functions such that the above expectations are finite, $\mathbb{E}$ denotes expectation, and $(f(n))_{n\geqslant 1}$ is a positive sequence, called the rate of convergence, such that $f(n) \to 0$, as $n \to \infty$. We refer to $d_{\mathscr{H}}(S_n, Z)$ as the Gaussian approximation error in the distance $d_{\mathscr{H}}$, and to the inequalities of the form (1.1) as quantitative bounds with rate function f. An example of such a distance $d_{\mathscr{H}}$ on $\mathbb{R}$ is the Kolmogorov distance d_{Kol}, for which $\mathscr{H}$ is the family of indicator functions of all half-lines $(-\infty, x]$, $x \in \mathbb{R}$. In this case, one usually writes $\mathbb{P}(S_n \leqslant x)$ instead of $\mathbb{E}[\mathbb{1}_{(-\infty,x]}(S_n)]$. A CLT equipped with an explicit error bound $d_{\mathscr{H}} \leqslant f(n)$ is often called a *quantitative CLT*.

Establishing explicit convergence rates turns CLTs from qualitative asymptotic statements into practically usable approximations. Indeed, in many applications one works with a sequence of statistics $(F_n)_{n\geqslant 1}$ and a corresponding Gaussian approximation Z (typically after suitable centring and scaling), and one wishes to control the approximation error in a metric such as $d_{\mathscr{H}}$. A bound of the form $d_{\mathscr{H}}(F_n, Z) \leqslant f(n)$ provides a practical criterion for choosing n, which ensures that the Gaussian approximation error is below a prescribed tolerance. For any n satisfying this criterion, practitioners may use the Gaussian approximation Z and its explicit quantiles and tail probabilities, for instance to construct confidence intervals and to set critical values for hypothesis tests.

A central benchmark for quantitative CLTs, is the seminal contribution of Berry [6] and Esseen [20, 21], commonly referred to as the Berry–Esseen inequality. Consider the normalised sum S_n and the limit Z defined above. Assuming a finite third moment $\mathbb{E}[|X_1|^3] < \infty$, the Berry–Esseen inequality yields for all $n \in \mathbb{N}$,

$$d_{\mathrm{Kol}}(S_n, Z) = \sup_{x\in\mathbb{R}} |\mathbb{P}(S_n \leqslant x) - \mathbb{P}(Z \leqslant x)| \leqslant C\frac{\mathbb{E}[|X_1|^3]}{\sqrt{n}}, \tag{1.2}$$

where $C > 0$ is a universal constant. The result (1.2) shows that the convergence rate in the CLT is of the order $n^{-1/2}$, thus providing a direct answer to Question 1. Furthermore, this convergence rate is considered optimal in the sense that symmetric Bernoulli random variables $(X_k)_{k\geqslant 1}$ (think fair coin tossing) achieve a lower bound with the same rate but different constant, see [22]. Hence, we cannot expect any rate of convergence to be faster than this optimal rate.

Foundational quantitative results date back to the 1940s and 1950s. The theory is well-developed for both univariate and multivariate settings, provided the dimension remains fixed. However, the recent emergence of high-dimensional data, in which the dimension d of the considered random vector may grow as fast, or even faster than the sample size n, has posed new challenges for classical probabilistic and statistical methods.

Two central difficulties arise. First, classical CLT results often assume independence or weak dependence structures, whereas high-dimensional settings frequently exhibit complex dependence patterns (as in neural networks). Second, in classical theory, the dimension d was treated as a fixed constant and absorbed into the constant terms of convergence bounds. In the high-dimensional setting, the critical challenge is to derive quantitative bounds $f(n, d)$ whose dependence on the dimension d is sufficiently mild so that $f(n, d) \to 0$ as $d, n \to \infty$, i.e.

Question 2.
(1) Given sequence of d-dimensional random vectors $(\boldsymbol{F}_n)_{n\geqslant 1}$ that satisfy a CLT when d is fixed, can we derive convergence rates $f(n,d)$ that ensure asymptotic normality even when d grows with n? Furthermore, can the obtained bounds accommodate nontrivial dependence structures?
(2) Specifically, can we derive quantitative bounds on the Gaussian approximation error $d_{\mathscr{H}}(\boldsymbol{F}, \boldsymbol{Z})$ for $\boldsymbol{F} \in \mathbb{R}^d$ and a Gaussian $\boldsymbol{Z} \in \mathbb{R}^d$ that make the dependence on the dimension d explicit?

A solution to Question 2 requires fixing a distance $d_{\mathscr{H}}$ and proving a quantitative bound $d_{\mathscr{H}}(\boldsymbol{S}_n, \boldsymbol{Z}) \leqslant f(n,d)$ with explicit dependence on d, while aiming to preserve close to optimal dependence on the sample size n. Three distances that are often used in this context are the hyper-rectangular $d_{\mathcal{R}}$, convex $d_{\mathscr{C}}$, and 1-Wasserstein $d_{\mathcal{W}}$ distance, which differ in the choice of test function class $\mathscr{H}$. For the hyper-rectangular distance, $\mathscr{H} = \{\mathbb{1}_A : A \in \mathcal{R}\}$, where $\mathcal{R}$ is the set of d-dimensional hyper-rectangles in $\mathbb{R}^d$. For the convex distance, $\mathscr{H} = \{\mathbb{1}_A : A \in \mathscr{C}\}$, where $\mathscr{C}$ is the family of all convex subsets of $\mathbb{R}^d$, and for the Wasserstein distance $\mathscr{H} = \mathrm{Lip}(1)$, where $\mathrm{Lip}(1)$ denotes the class of Lipschitz continuous functions $h : \mathbb{R}^d \to \mathbb{R}$ with Lipschitz constant at most 1, that is $|h(x) - h(y)| \leqslant |x - y|$ for all $x, y \in \mathbb{R}^d$. These distances are related. For instance, for random vectors $\boldsymbol{F}$ and $\boldsymbol{Z}$, it is known that $d_{\mathscr{C}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C\sqrt{d_{\mathcal{W}}(\boldsymbol{F}, \boldsymbol{Z})}$, for a positive constant C, and since $\mathcal{R} \subset \mathscr{C}$, the hyper-rectangular distance is always bounded above by the convex distance. However, in many statistical problems, one often considers the hyper-rectangular distance. Examples of such statistical problems include multiple hypothesis testing and simultaneous confidence intervals for many parameters at once, where the sets of accepted or covered values naturally have the shape of hyper-rectangles, see for example [11, 12] and references therein.

Addressing Question 2 constitutes a central theme in modern probability and high-dimensional statistics. Similar difficulties appear in many modern statistical models, such as nonlinear time series, functionals of Gaussian fields, and stochastic systems with strong temporal or spatial dependence. In these settings, classical Fourier methods generally fail to yield sharp, dimension-explicit error bounds. This need has given rise to the development of general techniques for Gaussian approximations that accommodate dependence and remain effective in high dimensions.

In what follows, we introduce two frameworks, Stein's method, and its combination with Malliavin calculus. We then present the main contributions of Paper A, which provide three answers to Question 2: one to part (2), and two to part (1). More precisely, we obtain general bounds on $d_{\mathscr{H}}(\boldsymbol{F}, \boldsymbol{Z})$ for random vectors on Gaussian spaces in the three distances mentioned above (hyper-rectangular, convex and 1-Wasserstein), which allow us to derive concrete rates of convergence in the multivariate Breuer–Major theorem, and subsequently, in parameter estimation for VAR(p) processes. To the best of our knowledge, the applications presented have not appeared previously in the literature. After each of those results, we discuss the different components that the obtained bounds consist of, and the relations to literature. Finally, a remark: the results in Paper A are quite technical, since the primary contribution lies in making all bounds explicit and tractable. Consequently, several details are simplified in this presentation, and all such instances are clearly indicated. For the rest of this section, we use common notation $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to denote that $\boldsymbol{X}$ is a d-dimensional random vector following a Gaussian (normal) distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Usually, we will assume that $\boldsymbol{X}$ has values in $\mathbb{R}^d$, and that $\boldsymbol{X}$ is centered, meaning that the mean vector is the zero vector in $\mathbb{R}^d$, i.e. $\boldsymbol{\mu} = \mathbf{0} \in \mathbb{R}^d$.

Stein's Method

The classical approach to Gaussian approximation relies on characteristic functions, exemplified by the seminal work of Berry and Esseen. In the case of independent random variables, the characteristic function of a sum factorises into a product of characteristic functions, a multiplicative structure that makes the analysis tractable and leads to sharp quantitative bounds. For dependent random variables, however, this structure breaks down, and Fourier methods often yield sub-optimal rates or become difficult to apply.

A fundamental shift occurred in 1972 with Stein's seminal work [49]. Stein observed that a random variable Z follows the standard normal distribution $\mathcal{N}(0,1)$ if and only if

$$\mathbb{E}[f'(Z)] = \mathbb{E}[Zf(Z)],$$

holds for all sufficiently regular test functions f. This characterisation implies that the distance between the distributions of an arbitrary random variable F and a standard normal random variable Z can be expressed through quantities of the form $\mathbb{E}[f'(F) - Ff(F)]$.

More precisely, for a given test function $h : \mathbb{R} \to \mathbb{R}$, one considers the *Stein equation*

$$f'(x) - xf(x) = h(x) - \mathbb{E}[h(Z)],$$

whose solution f_h satisfies $\mathbb{E}[h(F)] - \mathbb{E}[h(Z)] = \mathbb{E}[f_h'(F) - Ff_h(F)]$. Comparing to (1.1), obtaining quantitative bounds reduces to bounding

$$d_{\mathscr{H}}(F, Z) = \sup_{h \in \mathscr{H}} |\mathbb{E}[f_h'(F) - Ff_h(F)]|.$$

For standard classes of test functions h, such as bounded, Lipschitz continuous, or smooth functions, the solution f_h exists and is unique within an appropriate function space, see [37, Proposition 3.2.2].

These ideas admit a natural generalisation to the multivariate setting. A d-dimensional random vector $\boldsymbol{Z}$ with values in $\mathbb{R}^d$ follows the Gaussian distribution $\mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$ if and only if

$$\mathbb{E}\big[\langle \boldsymbol{Z}, \nabla f(\boldsymbol{Z})\rangle_{\mathbb{R}^d}\big] = \mathbb{E}\big[\mathrm{tr}\big(\boldsymbol{\Sigma}\mathrm{Hess} f(\boldsymbol{Z})\big)\big], \tag{1.3}$$

for all functions $f \in C^2(\mathbb{R}^d)$ with bounded gradient and Hessian; see [37, Lemma 4.1.3]. Here Hess$f$ denotes the Hessian matrix of second-order partial derivatives, and $\mathrm{tr}(\boldsymbol{M})$ denotes the trace of a matrix $\boldsymbol{M} \in \mathbb{R}^{d\times d}$. In analogy with the univariate case, the characterisation (1.3) leads to a multivariate Stein equation, whose solution f_h exists for a wide range of test function families $\mathscr{H}$ of interest. For a random vector $\boldsymbol{F}$ in $\mathbb{R}^d$, the Gaussian approximation error in the distributional distance $d_{\mathscr{H}}$ can be written as

$$d_{\mathscr{H}}(\boldsymbol{F}, \boldsymbol{Z}) = \sup_{h \in \mathscr{H}} |\mathbb{E}[\mathrm{tr}(\boldsymbol{\Sigma}\mathrm{Hess} f_h(\boldsymbol{F})) - \langle \boldsymbol{F}, \nabla f_h(\boldsymbol{F})\rangle_{\mathbb{R}^d}]|. \tag{1.4}$$

Thus, bounding the Gaussian approximation error $d_{\mathscr{H}}(\boldsymbol{F}, \boldsymbol{Z})$ reduces to controlling the right-hand side uniformly in h in (1.4).

A central object in quantitative bounds for $d_{\mathscr{H}}(\boldsymbol{F}, \boldsymbol{Z})$ is the *Stein kernel*. For a centered random vector $\boldsymbol{F}$ in $\mathbb{R}^d$, a Stein kernel is a matrix-valued function $\tau^{\boldsymbol{F}} : \mathbb{R}^d \to \mathbb{R}^{d\times d}$, where $\mathbb{E}[|\tau^{\boldsymbol{F}}_{i,j}(\boldsymbol{F})|] < \infty$ for all $i, j \in \{1, \ldots, d\}$, and such that

$$\mathbb{E}\big[\langle \boldsymbol{F}, \nabla f(\boldsymbol{F})\rangle_{\mathbb{R}^d}\big] = \mathbb{E}\big[\mathrm{tr}\big(\tau^{\boldsymbol{F}}(\boldsymbol{F})\mathrm{Hess} f(\boldsymbol{F})\big)\big], \tag{1.5}$$

for all $f : \mathbb{R}^d \to \mathbb{R}$, $f \in C^2$ with bounded gradient and Hessian. The random matrix $\tau^{\boldsymbol{F}}(\boldsymbol{F})$ is often called the Stein matrix. Intuitively, $\tau^{\boldsymbol{F}}(\boldsymbol{F})$ plays the role of a *stochastic covariance matrix*. In particular, for $\boldsymbol{Z} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$ one can choose $\boldsymbol{\tau}^{\boldsymbol{Z}}(\boldsymbol{Z}) \equiv \boldsymbol{\Sigma}$, and then (1.5) reduces to the Gaussian characterisation (1.3).

Therefore, obtaining quantitative Gaussian approximation reduces to controlling the size of the discrepancy between the random Stein matrix $\boldsymbol{\tau}^{\boldsymbol{F}}(\boldsymbol{F})$ and the target covariance matrix $\boldsymbol{\Sigma}$. It is now known that, for general dependent random vectors admitting a Stein kernel, high-dimensional Gaussian approximation can be bounded explicitly in terms of these discrepancy terms. Indeed, Fang and Koike in [23, Thm 1.1] obtain the following bound.

Theorem 1.1 (Fang–Koike, 2021)**.** *Let $\boldsymbol{F}$ be a centered d-dimensional random vector that has a Stein kernel $\tau^{\boldsymbol{F}}$, and let $\boldsymbol{Z} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$. Then there exists a positive constant that depends on $\boldsymbol{\Sigma}$ such that*

$$d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C(\boldsymbol{\Sigma}) \log(d) \mathbb{E}[\|\boldsymbol{\Sigma} - \boldsymbol{\tau}^{\boldsymbol{F}}(\boldsymbol{F})\|_{\max}], \tag{1.6}$$

where $\|\boldsymbol{M}\|_{\max} = \max_{1\leqslant i,j\leqslant d} |\boldsymbol{M}_{ij}|$ denotes the max-norm for a matrix $\boldsymbol{M} \in \mathbb{R}^{d\times d}$.

The term $\mathbb{E}[\|\boldsymbol{\Sigma} - \boldsymbol{\tau}^{\boldsymbol{F}}(\boldsymbol{F})\|_{\max}]$ is precisely a measurement of the discrepancy between the Stein matrix $\boldsymbol{\tau}^{\boldsymbol{F}}(\boldsymbol{F})$ and the covariance matrix $\boldsymbol{\Sigma}$.

As for related literature, Stein's method has proved highly successful for obtaining dimension-explicit bounds in multivariate Gaussian approximation. A key advance was made by Götze [24], who applied Stein's method to sums of independent d-dimensional random vectors. This was later refined by Bentkus [4, 5], who obtained a Berry–Esseen bound of order $d^{1/4}n^{-1/2}$ in the convex distance. For this setting, this is the best currently known dependence on the dimension d in convex distance, and it provides an example where the rate still tends to zero even when d grows with n, thus answering Question 2. For normalised sums of independent, centered d-dimensional random vectors with an additional assumption on the boundedness of the random vectors, Chernozhukov, Chetverikov, and Koike [13] obtain rates of order $\log^{3/2}(d)n^{-1/2}\log n$ in the hyper-rectangular distance, which are optimal in n up to logarithmic factors and have a remarkable dimensional dependence. Namely the bound scales as $\log(d)$ as d grows. All these results address sums of independent random vectors and provide answers to Question 2 within their respective frameworks. However, since data in practice are often dependent, it is also essential to obtain corresponding bounds for the dependent random vectors. To overcome these challenges in the dependent setting, one possible approach combines Stein's method with tools from Malliavin calculus.

Malliavin–Stein Method

Although Stein's method provides a powerful analytical framework for bounding distances, the construction of a Stein kernel for dependent functionals often requires additional mathematical structure. For functionals on Gaussian spaces, this can be achieved by combining Stein's method with Malliavin calculus.

Malliavin calculus was introduced by Paul Malliavin in 1976 in connection with a concrete problem in PDE theory: to obtain a purely probabilistic proof of Hörmander's theorem on the smoothness of densities for hypoelliptic diffusions, without resorting to classical PDE techniques. Malliavin's answer was a calculus of variations on Wiener space, providing a rigorous framework to differentiate random functionals (functions of Brownian paths) by the means of the Malliavin derivative DF.

Historically, the domain of Malliavin calculus was density regularity theory. A pivotal development in its use for limit theorems occurred in 2005 when Nualart and Peccati established their celebrated Fourth Moment Theorem [45]: functionals in Wiener chaos converge to a Gaussian random variable if and only if their fourth moment converges to that of the Gaussian variable. This revealed that Malliavin calculus could be used for distributional convergence, not just density smoothness. Nourdin and Peccati in [36] then showed that combining Stein's method with Malliavin calculus allows explicit construction of Stein kernels for Gaussian functionals, that is, random variables which are functions of underlying Gaussian random variables. The Malliavin derivative naturally produces a canonical Stein kernel (via $\langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}$, see below), which captures the covariance structure in the form needed for Stein's method. Later, Nourdin, Peccati and Réveillac [38], extended this to the multivariate setting.

For a comprehensive treatment of Malliavin–Stein method, Nourdin and Peccati's monograph [37] presents key results in Malliavin calculus, and proves a handful of classical results: the Fourth Moment Theorem, the classical Berry–Esseen result (1.2), and the Breuer–Major theorem in depth, to name a few. For further reading primarily on Malliavin calculus, see monographs [43, 44]. In this section, we develop tools for deriving quantitative bounds on the Gaussian approximation error and introduce, in a non-technical way, some basic notions from Malliavin calculus.

Formally, let $X = \{X(h) : h \in \mathfrak{H}\}$ be an isonormal Gaussian process over a real separable Hilbert space $\mathfrak{H}$, that is, a centered Gaussian family with covariance $\mathbb{E}[X(h)X(g)] = \langle h, g\rangle_{\mathfrak{H}}$. For a positive integer q, let $\mathfrak{H}^{\otimes q}$, and $\mathfrak{H}^{\odot q}$ denote the q-fold tensor product and the q-fold symmetric tensor product of $\mathfrak{H}$, respectively. Any square-integrable random variable $F \in L^2(\Omega)$ admits a Wiener–Itô chaos expansion

$$F = \mathbb{E}[F] + \sum_{q\geqslant 1} I_q(f_q), \quad \text{with} \quad f_q \in \mathfrak{H}^{\odot \mathfrak{H}}, \tag{1.7}$$

where I_q denotes the q-th multiple Wiener integral. Heuristically, the term $I_q(f_q)$ behaves like a homogeneous polynomial of degree q in the underlying Gaussian coordinates. It is obtained by integrating a kernel $f_q \in \mathfrak{H}^{\odot q}$ against q increments of the Gaussian noise, and thus represents the q-th order "noise component" of F. In this sense, Wiener chaos is the orthogonal decomposition of the space of square-integrable random variables $L^2(\Omega)$, into subspaces of different noise orders, where the "noise component" of order $q \in \mathbb{N}$ is called the Wiener chaos of order q. A valid question is "To which random variables can we apply the Malliavin derivative operator?", and unfortunately, the answer is not all square-integrable random variables, namely F belonging to the domain $\mathbb{D}^{1,2}$ of Malliavin derivative is a stronger condition, i.e. $\mathbb{D}^{1,2} \subset L^2(\Omega)$. For details see Section 4.1 or [37]. Malliavin calculus provides differential rules for such functionals as well as an integration by parts formula involving the pseudo-inverse L^{-1} of the Ornstein–Uhlenbeck generator.

For a centered vector $\boldsymbol{F} = (F_1, \ldots, F_d)$ with values in $\mathbb{R}^d$, and $F_i \in \mathbb{D}^{1,2}$ and a sufficiently smooth test function $f : \mathbb{R}^d \to \mathbb{R}$, the Malliavin integration by parts identity yields

$$\mathbb{E}\big[\langle \boldsymbol{F}, \nabla f(\boldsymbol{F})\rangle_{\mathbb{R}^d}\big] = \sum_{i,j=1}^{d} \mathbb{E}\left[\partial_{ij}^2 f(\boldsymbol{F})\langle DF_i, -DL^{-1}F_j\rangle_{\mathfrak{H}}\right]. \tag{1.8}$$

The formula (1.8) is the fundamental result for the Malliavin–Stein method. Denote the matrix $M_{\boldsymbol{F}}(i,j) := \langle DF_i, -DL^{-1}F_j\rangle_{\mathfrak{H}}$, for all $i, j \in \{1, \ldots, d\}$. Comparing (1.8) with the Stein kernel identity (1.5) and taking the conditional expectation with respect to the σ-algebra generated by $\boldsymbol{F}$, we obtain the promised explicit Stein kernel for the law of $\boldsymbol{F}$:

$$\tau_{ij}^{\boldsymbol{F}}(\boldsymbol{F}) = \mathbb{E}[\langle DF_i, -DL^{-1}F_j\rangle_{\mathfrak{H}} \mid \boldsymbol{F}] = \mathbb{E}[M_{\boldsymbol{F}}(i,j) \mid \boldsymbol{F}], \qquad 1 \leqslant i, j \leqslant d. \tag{1.9}$$

For a more detailed mathematical statement, see [40, Prop. 3.7]. In particular, in connection to the bound (1.6), it can be shown that

$$\mathbb{E}[\|\boldsymbol{\Sigma} - \tau^{\boldsymbol{F}}(\boldsymbol{F})\|_{\max}] \leqslant \mathbb{E}[\|\boldsymbol{\Sigma} - M_{\boldsymbol{F}}\|_{\max}].$$

This follows from conditional Jensen's inequality and the tower property, see also proof of Theorem 4.1 in Paper A.

Identity (1.9) shows that bounding the Gaussian approximation error $d_{\mathscr{H}}(\boldsymbol{F}, \boldsymbol{Z})$ for functionals on a Gaussian space reduces to controlling the fluctuations of $M_{\boldsymbol{F}}$ around the target covariance matrix $\boldsymbol{\Sigma}$. Since $\tau^{\boldsymbol{F}}(\boldsymbol{F}) = \mathbb{E}[M_{\boldsymbol{F}} \mid \boldsymbol{F}]$, the same control also bounds deviations of $\tau^{\boldsymbol{F}}(\boldsymbol{F})$ from $\boldsymbol{\Sigma}$. This principle underlies a large body of quantitative limit theorems obtained via the Malliavin–Stein method. See, for instance (1.10) and (1.11).

Nourdin, Peccati and collaborators established the following two key quantitative bounds: one in convex [42, Thm 2.1] and one in the 1-Wasserstein distance [38, Thm 3.5] stated in Theorem 1.2, in (1.10) and (1.11) respectively.

Theorem 1.2 (Nourdin–Peccati–Yang, 2022 & Nourdin–Peccati–Réveillac, 2010)**.** *Fix a centered random vector vector $\boldsymbol{F} = (F_1, \ldots, F_d)$ in $\mathbb{R}^d$ with components in $\mathbb{D}^{1,2}$, and $\boldsymbol{Z} \sim \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ invertible. Then there exists a constant depending on $\boldsymbol{\Sigma}$, such that*

$$d_{\mathscr{C}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C(\boldsymbol{\Sigma}) d^{41/24} \sqrt{\mathbb{E}[\|\boldsymbol{\Sigma} - M_{\boldsymbol{F}}\|_{\text{H.S.}}^2]}, \quad \textit{and} \tag{1.10}$$

$$d_{\mathcal{W}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C(\boldsymbol{\Sigma})\sqrt{d}\sqrt{\mathbb{E}\left[\|\boldsymbol{\Sigma} - M_{\boldsymbol{F}}\|_{\text{H.S.}}^2\right]}, \tag{1.11}$$

where $\|\boldsymbol{M}\|_{\text{H.S.}} = \sqrt{\operatorname{tr}(\boldsymbol{M}\boldsymbol{M}^\top)}$ denotes the Hilbert–Schmidt norm of a matrix $\boldsymbol{M} \in \mathbb{R}^{d\times d}$.

We comment on the role of $C(\boldsymbol{\Sigma})$ in the discussion point (II) below Theorem 1.3. Note that $\mathbb{E}[\|\boldsymbol{\Sigma} - M_{\boldsymbol{F}}\|_{\text{H.S.}}^2]^{1/2}$ measures the discrepancy between $M_{\boldsymbol{F}}$ and $\boldsymbol{\Sigma}$.

The bounds in Theorems 1.1 and 1.2, are explicit in the dimension d up to discrepancy terms: $\mathbb{E}[\|\boldsymbol{\Sigma} - \boldsymbol{\tau}^{\boldsymbol{F}}(\boldsymbol{F})\|_{\max}]$ and $\mathbb{E}[\|\boldsymbol{\Sigma} - M_{\boldsymbol{F}}\|_{\text{H.S.}}^2]$, respectively. These terms quantify deviations from the Gaussian target covariance matrix $\boldsymbol{\Sigma}$ (via $\tau^{\boldsymbol{F}}(\boldsymbol{F})$ in Theorem 1.1 and $M_{\boldsymbol{F}}$ in Theorem 1.2). Consequently, they provide a dimension-explicit answer to Question 2(2) up to the control of the discrepancy term. However, in many high-dimensional settings, the discrepancy term hides additional dimensional dependence. Hence, obtaining fully explicit and tractable rates requires further structure, which motivates the structural assumption introduced in Theorem 1.3.

Dimensional Explicit Quantitative Bounds

This section presents the first of the contributions from Paper A, that provides an answer to Question 2(2). The existing Malliavin–Stein bounds in the multivariate setting are typically formulated in non-explicit form, (recall the discussions surrounding (1.6), (1.11) and (1.10)), or are derived under the restriction that each component belongs to a fixed Wiener chaos ([42, Cor 1.3] and [23, Corollary 1.2]). Theorem 1.3 addresses this gap by providing dimension-explicit quantitative bounds for random vectors $\boldsymbol{F}$ on a Gaussian space, under appropriate regularity assumptions, in three distances: hyper-rectangular, convex and 1-Wasserstein.

We begin by stating Theorem 1.3, and then discuss the main structural components of the resulting bounds, namely their dependence on the dimension d and on the smallest eigenvalue σ_* of the covariance (or correlation) matrix of the Gaussian limit.

To present d-dependence for the hyper-rectangular distance, define a function

$$\psi_{\alpha,\beta}(d) := \log_+^{1/(2\beta)}(d) e^{k_1 \log_+^{1/(2\beta)}(d)}, \text{ for all } d \in \mathbb{N}, \tag{1.12}$$

where $\log_+(x) = \max\{|\log(x)|, 1\}$ for $x \in \mathbb{R}$, and where the parameters $\alpha \in \mathbb{R}$ and $\beta \in [1/2, 1]$ control the growth rate. The constant k_1 depends on β: specifically, $k_1 \approx 0.368$ when $\beta = 1/2$, and for $\beta > 1/2$, the value $k_1 = k_1(\alpha, \beta)$ is determined explicitly, see (1.4) in Paper A.

The key observation is that for $\beta \in (1/2, 1]$, the exponents satisfy $1/(2\beta) \in [1/2, 1)$, and consequently $\psi_{\alpha,\beta}$ grows sub-polynomially in d, that is, $\psi_{\alpha,\beta}(d) \leqslant d^{\varepsilon}$ for all $\varepsilon > 0$. In contrast, the critical case $\beta = 1/2$ yields polynomial growth in d with order $d^{0.368}$. In Theorem 2.1 from Paper A, we obtain the following result.

Theorem 1.3. *Fix a centered random vector $\boldsymbol{F} = (F_1, \dots, F_d)$ in $\mathbb{R}^d$ where $F_i \in \mathbb{D}^{1,2}$ and also admits a Wiener–Itô chaos expansion (1.7) for all $i \in \{1, \dots, d\}$. Let $\boldsymbol{Z} \sim \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{\Sigma})$ with matching covariance $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{F})$. Assume $\boldsymbol{\Sigma}$ is invertible, with the associated correlation matrix $\boldsymbol{\Lambda}$. Let $\sigma_*(\boldsymbol{M})$ and $\sigma^*(\boldsymbol{M})$ denote the minimal and respectively maximal eigenvalue of a matrix $\boldsymbol{M}$. Assume for $\gamma, \alpha \in \mathbb{R}$ and $\beta \in [1/2, 1]$ that*

$$\|f_{i,p} \otimes_r f_{j,q}\|_{\mathfrak{H}^{\otimes(p+q-2r)}} \leqslant \frac{\gamma e^{\alpha p} e^{\alpha q}}{(p!q!)^{\beta}}, \qquad \textit{for all } i, j \in \{1, \dots, d\}, \tag{1.13}$$

and all p, q, r natural numbers where $p, q \geqslant 1$, not simultaneously 1, and $1 \leqslant r \leqslant \min\{p, q\} - \mathbb{1}_{\{p=q\}}$. Additionally, if $\beta = 1/2$, for (1.14) assume that $\alpha < \alpha_0 \approx -2.846$, and for (1.15) that $\alpha < \log(1/2) - e^{1/(2e)} \approx -1.895$. Then, there exists a finite constant $C_{\boldsymbol{\theta}} > 0$, depending *only on multivariate parameter* $\boldsymbol{\theta} = (\alpha, \beta)$*, such that*

$$d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C_{\boldsymbol{\theta}} \psi_{\alpha,\beta}(d) \gamma \log_+(\gamma) \frac{\log_+(\sigma_*(\boldsymbol{\Lambda}))}{\sigma_*(\boldsymbol{\Lambda})}, \tag{1.14}$$

$$d_{\mathscr{C}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C_{\boldsymbol{\theta}} d^{65/24} \gamma \frac{1}{\sigma_*(\boldsymbol{\Lambda})^{3/2}}, \quad \text{and} \quad d_{\mathcal{W}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C_{\boldsymbol{\theta}} d^{3/2} \gamma \frac{\sigma^*(\boldsymbol{\Sigma})^{1/2}}{\sigma_*(\boldsymbol{\Sigma})}. \tag{1.15}$$

We begin with two technical comments: (i) the constant α_0 is given explicitly in (1.3) in Paper A, and (ii) in Paper A we present versions of the result (namely Theorem 4.1 and Corollary 4.2) without the assumption that the covariance of the Gaussian vector equals the covariance of $\boldsymbol{F}$. Namely, the result holds for a general invertible $\boldsymbol{\Sigma}$, which simply adds an additional term to the bounds that accounts for the discrepancy between the two matrices. We now turn to the two central aspects of this discussion:

(I) The three bounds in Theorem 1.3 exhibit different dependencies on the dimension d. The hyper-rectangular distance bound (1.14) is unique among the three in its sensitivity to the regularity parameters α and β (see (1.12)). Specifically, for $\beta \in (1/2, 1]$, the rate $\psi_{\alpha,\beta}(d)$ grows sub-polynomially in d, however for $\beta = 1/2$ and $\alpha < \alpha_0$, the rate becomes polynomial of order d^{k_1} with $k_1 \leqslant 0.368$. Consequently, stronger regularity assumptions on $\boldsymbol{F}$ directly improves the dimensional dependence of the hyper-rectangular bound. Such behaviour is absent from both the convex and 1-Wasserstein bound. Importantly, regardless of regularity, the hyper-rectangular bound achieves substantially better dimension dependence than the convex and 1-Wasserstein bounds, which have fixed polynomial rates of $d^{65/24}$ and $d^{3/2}$, respectively.

(II) The factor $(\sigma_*(\boldsymbol{M}))^{-1}$ appears in all bounds for $\boldsymbol{M} \in \{\boldsymbol{\Sigma}, \boldsymbol{\Lambda}\}$, and it diverges as $\boldsymbol{M}$ approaches singularity. If $\sigma_*(\boldsymbol{M})$ is bounded away from 0 by an absolute constant, terms themselves reduce to a universal constant. The $(\sigma_*(\boldsymbol{M}))^{-1}$ terms are precisely the terms that appear in the bounds (1.6), (1.10), and (1.11) through $C(\boldsymbol{\Sigma})$. Note, however, that even under uniform control of $\sigma_*(\boldsymbol{\Sigma})$, the bound in (1.11) also involves the maximal eigenvalue $\sigma^*(\boldsymbol{\Sigma})$. Consequently, the bound may still deteriorate with the dimension through the overall scale of $\boldsymbol{\Sigma}$, and controlling non-singularity alone does not prevent dimension-dependent constants.

The hyper-rectangular bound (1.14), which accommodates general vectors $\boldsymbol{F}$ without fixed chaos restriction, is novel to our knowledge. It builds upon work by Fang and Koike, and generalises [23, Cor. 1.2], which required all components to belong to a Wiener chaos of *fixed order*. The convex and 1-Wasserstein bounds (1.15) refine results [38, Thm 3.5] and [42, Thm 2.1]. Under the assumption (1.13), we provide an explicit estimate of the Stein kernel term, $\mathbb{E}[\|\boldsymbol{\Sigma} - M_{\boldsymbol{F}}\|^2_{\text{H.S.}}]^{1/2}$, which is the key to achieving dimension-explicit dependence.

Having the explicit bounds from Theorem 1.3 allows us to prove an important motivational result, namely the Breuer–Major theorem.

Multivariate Quantitative Breuer–Major

In this section, we present a special case of our statements involving the multivariate Breuer–Major theorem from Paper A, namely Theorem 2.3 and Corollary 2.4. The Breuer–Major theorem is a fundamental result in Gaussian analysis, guaranteeing convergence to a Gaussian limit of random vectors $\boldsymbol{S}_n := n^{-1/2}\sum_{k=1}^{n}\Phi(\boldsymbol{G}_k)$, for a centered stationary K-dimensional Gaussian sequence $(\boldsymbol{G}_k)_{k\in\mathbb{Z}}$, and a nonlinear function $\Phi : \mathbb{R}^K \to \mathbb{R}^d$, under appropriate assumptions. Unlike the standard Berry–Esseen result (1.2) for independent random variables, Breuer–Major accommodates strong dependence through a Hermite rank condition on the non-affine functional Φ, under an appropriate summability condition on the covariance structure. This framework extends classical limit theorems to dependent Gaussian data, making it fundamental for analysing non-affine functionals of time series and other dependent structures.

The seminal work [7] by Breuer and Major is concerned with the qualitative statement of this problem, i.e. they show weak convergence in the case when $d = 1$ and $K = 1$. Their result is later extended to multivariate inputs by Arcones in [1] to cover functionals of the form $\Phi : \mathbb{R}^K \to \mathbb{R}$, and further to vector-valued outputs $\Phi : \mathbb{R}^K \to \mathbb{R}^d$ in the continuous setting by Nualart and Tiva [46].

The quantitative bounds are well-established for the following three cases: (i) scalar-valued functions with $K = 1$ (e.g., [41, 47]), (ii) scalar-valued functions with $K \geqslant 1$ in [39, Thm 2.1], however their result is not explicit in K nor d, and (iii) $\mathbb{R}^d$-valued functions with $K = 1$ (e.g., [38, 42]). To the best of our knowledge, multivariate extensions for general maps $\Phi : \mathbb{R}^K \to \mathbb{R}^d$ with explicit dependence on the input dimension K, output dimension d and sample size n, are currently unknown, despite their importance for high-dimensional inference where one needs quantitative control of Gaussian approximations rather than purely qualitative CLTs. Corollary 1.4 provides such bounds, which give an answer to this gap, and hence an answer to Question 2(1) for the class of nonlinear functionals considered here.

Let $(\boldsymbol{G}_k)_{k\in\mathbb{Z}}$ be a centered i.i.d. stationary Gaussian sequence in $\mathbb{R}^K$ with $\boldsymbol{G}_k = (\boldsymbol{G}_k^{(j)})_{1\leqslant j\leqslant K} \sim \mathcal{N}_K(\boldsymbol{0}, \boldsymbol{I}_K)$, and let $\Phi = (\varphi_1, \ldots, \varphi_d)$ be such that the components φ_i are square-integrable and have finite Hermite rank $m_i \geqslant 2$. Here, square-integrable $\varphi_i : \mathbb{R}^K \to \mathbb{R}$ means $\mathbb{E}[|\varphi_i(\boldsymbol{G}_1)|^2] < \infty$. Each such φ_i admits a Hermite expansion, $\varphi_i(\boldsymbol{x}) = \sum_{i=m_i}^{\infty}\varphi_{i,q}(\boldsymbol{x})$, where $\varphi_{i,q}(\boldsymbol{x})$ is the projection of φ_i onto the space generated by multivariate Hermite polynomials of total degree q, that is the q-th Wiener chaos. The Hermite rank $m_i \in \mathbb{N}$ is the smallest $q \geqslant 1$ such that $\varphi_{i,q} \not\equiv 0$, that is, all projections of degrees $0, \ldots, m_i - 1$ vanish. Hence, having Hermite rank of at least 2, essentially means that there are no constant or linear terms in the Hermite expansion of φ_i. in particular, $\varphi_i(\boldsymbol{G}_1)$ is not a Gaussian random variable. For a detailed discussion of these notions, see Section 4.1 in Paper A. For the described set-up, Paper A yields the following:

Corollary 1.4. *Assume $\boldsymbol{\Sigma}_n := \mathrm{Cov}(\boldsymbol{S}_n)$ is invertible, and let $\boldsymbol{Z}_n \sim \mathcal{N}_d(0, \boldsymbol{\Sigma}_n)$. Furthermore, assume that $\mathbb{E}[\Phi(\boldsymbol{G}_k)] = 0$ for all $k \in \mathbb{Z}$. Let $c_1 \geqslant 0$, $\alpha \in \mathbb{R}$ and $\beta \in [1/2, 1]$, and define the multivariate parameter $\boldsymbol{\theta} = (c_1, \alpha, \beta)$. If $\beta = 1/2$, for the hyper-rectangular distance bound assume that $\alpha < \alpha_0 - \log_+(K)/2 < 0$ and for the other two distances assume that $\alpha < -1/(2e) - \ln(2\sqrt{K}) < 0$. Assume, that*

$$\mathbb{E}[\varphi_{i,q}(\boldsymbol{G}_1)^2]^{1/2} \leqslant \frac{c_1 e^{\alpha q}}{(q!)^{\beta-1/2}}, \quad \textit{for all } q \in \mathbb{N}. \tag{1.16}$$

Finally, for positive constants $k_2 \approx 0.768$ and c_2 depending on $\boldsymbol{\theta}$, define

$$\zeta(d,K) := \begin{cases} \log_+^{1/(2\beta)}(d) e^{c_2 K^{1/2\beta} \log_+^{1/(2\beta)}(d)}, & \beta \in (1/2, 1], \\ \log_+(d) d^{k_2}, & \beta = 1/2. \end{cases} \tag{1.17}$$

Then, there exist positive constants $C_{K,\boldsymbol{\theta}}$ depending on K and $\boldsymbol{\theta}$, such that

$$\begin{aligned} d_{\mathcal{R}}(\boldsymbol{S}_n, \boldsymbol{Z}_n) &\leqslant C_{K,\boldsymbol{\theta}} \zeta(d,K) n^{-1/2} \log_+(n) \frac{\log_+(\sigma_*(\boldsymbol{\Lambda}))}{\sigma_*(\boldsymbol{\Lambda})}, \\ d_{\mathscr{C}}(\boldsymbol{S}_n, \boldsymbol{Z}_n) &\leqslant C_{K,\boldsymbol{\theta}} d^{65/24} n^{-1/2} \frac{1}{\sigma_*(\boldsymbol{\Lambda})^{3/2}}, \ \textit{and} \\ d_{\mathcal{W}}(\boldsymbol{S}_n, \boldsymbol{Z}_n) &\leqslant C_{K,\boldsymbol{\theta}} d^{3/2} n^{-1/2} \frac{\sigma^*(\boldsymbol{\Sigma})^{1/2}}{\sigma_*(\boldsymbol{\Sigma})}. \end{aligned} \tag{1.18}$$

First, note that in the boundary case of $\beta = 1/2$, the factor α, describing the growth of Hermite components of every φ_i, depends on the input dimension K. Specifically, as the dimension K of the underlying process $(\boldsymbol{G}_k)k \in \mathbb{Z}$ increases, the parameter α must decrease for the result to hold, equivalently, $\varphi i, q$ must become more regular. The function ζ, described in (1.17), grows sub-polynomially in d for $\beta \in (1/2, 1]$, and polynomially in the boundary case $\beta = 1/2$. The order of the polynomial growth $k_2 \approx 0.768$, which is less than polynomial growths order 65/24 for the convex distance, and for 3/2 for 1-Wasserstein distance in (1.18).

In Corollary 1.4, we treat the i.i.d. case of the multivariate Gaussian sequences. However, the results in Paper A hold for Gaussian sequences with a non-trivial covariance structure. This makes it possible to consider long-range dependent Gaussian fields, which are of interest in various applications, see [31, 17, 34, 2]. Taking $K = 1$, a canonical example of a long-range dependent sequence is fractional Gaussian noise. For our general statements, see Theorem 2.3 for the full generality, and Corollary 2.4 for an application under assumption (1.16) and a decaying covariance structure (such as that of fractional Gaussian noise). As a sanity check, when we specialise to the case where $(G_k)_{k \in \mathbb{Z}}$ is fractional Gaussian noise with Hurst parameter $H \in (0,1)$, our bound in Corollary 2.4 recovers the existing n-rates for the 1-Wasserstein distance [38, Thm 4.1] and for the convex [39, Ex. 2.6], up to a $\log_+(n)$ factor in the case $H = 1/2$.

The Breuer–Major theorem has been extensively studied from a theoretical perspective [8, 35, 33], and, as mentioned above, the same structure often arises in statistical applications. In particular, stable vector autoregressive (VAR) models are among the central tools for analysing multivariate time series in econometrics, finance, macroeconomics, and many applied sciences [25, 32]. Parameter estimation of the coefficients in a stable VAR model with Gaussian innovations has Hermite rank exactly 2, and fits naturally into the Breuer–Major framework.

Parameter Estimation in Vector Autoregressive Processes

Multivariate time series in applications are often modelled using VAR processes, which provide a flexible linear framework for describing how several variables influence each other over time. In such models, each component today is expressed as a linear combination of past values of all components, so past movements in one variable can feed into the future evolution of the others. For instance, one may model how today's values of several variables (such as time spent listening to upbeat music and a daily mood score) depend on yesterday's values of both variables, thereby capturing questions like whether increased music listening today is associated with higher mood tomorrow or whether a low mood leads to different listening patterns on the following day.

Standard inference for VAR models is often based on the asymptotic normality of the multivariate ordinary least squares (OLS) estimator of the coefficient matrices under stability and moment conditions, as described in classical time series textbooks (see [32]). However, these results are typically qualitative: they state that normality holds in the limit, but do not provide explicit bounds. For reliable confidence intervals and hypothesis tests, one needs quantitative bounds on the asymptotic normality, and in Paper A we derive explicit convergence rates for the *CLT component* of the OLS estimator. To our knowledge, the rate of convergence in the asymptotic normality for parameter estimators in VAR models (see (1.19)) is currently unknown. Below we explain what being the CLT component of an OLS estimator means, and we present a simplified case of our Theorem 3.4 from Paper A, followed by a short discussion and relations to existing literature.

When proving asymptotic normality of a parameter estimator $\widehat{\boldsymbol{\theta}}$ (showing that its distribution converges to a Gaussian law as the sample size tends to infinity (1.19)), a particularly convenient situation is when the estimator can be expressed as a smooth function of empirical averages. In such cases, one aims to establish

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma}), \quad \text{as } n \to \infty, \tag{1.19}$$

for a positive definite matrix $\boldsymbol{\Sigma}$. A common strategy is to rewrite

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \boldsymbol{N}_n \boldsymbol{S}_n,$$

where $\boldsymbol{N}_n \xrightarrow{\mathbb{P}} \boldsymbol{N}$ by an appropriate law of large numbers (LLN) result, and where $\boldsymbol{S}_n \xrightarrow{\mathcal{D}} \boldsymbol{S}$ is for a Gaussian $\boldsymbol{S}$, as $n \to \infty$, usually by a CLT for suitably centered and scaled empirical averages. Multivariate versions of Slutsky's theorem then ensure that $\boldsymbol{N}_n \boldsymbol{S}_n \xrightarrow{\mathcal{D}} \boldsymbol{N}\boldsymbol{S}$ as $n \to \infty$, so that $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normal with covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{N}\boldsymbol{S})$. Hence, the CLT component of an OLS estimator refers to $\boldsymbol{S}_n$, which satisfies a CLT $\boldsymbol{S}_n \xrightarrow{\mathcal{D}} \boldsymbol{S}$, where $\boldsymbol{S}$ is a Gaussian random vector. For a concrete illustration in the context of the OLS estimator of the coefficient matrices in a stable VAR model with p lags (VAR(p)), see [32, Sec. 3.2.2].

Consider a d-dimensional stable VAR(1) process, that is

$$\boldsymbol{y}_k = \boldsymbol{A}\boldsymbol{y}_{k-1} + \boldsymbol{u}_k, \qquad \text{for } k \in \mathbb{Z}, \tag{1.20}$$

with coefficient matrix $\boldsymbol{A} \in \mathbb{R}^{d \times d}$ and the innovation process $(\boldsymbol{u}_k)_{k\in\mathbb{Z}}$ such that $\boldsymbol{u}_k \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{I}_d)$. Let $\boldsymbol{\mathcal{Y}} := (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ for fixed $n \in \mathbb{N}$. Furthermore, let $\widehat{\boldsymbol{A}}$ be the multivariate least squares estimator of the coefficient matrix $\boldsymbol{A}$, and $\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$ their vector forms respectively. Denote by $\boldsymbol{S}_n^{\widehat{\beta}}$ the CLT component of $\widehat{\boldsymbol{\beta}}$. Combining the above logic with the framework of [32, Sec. 3.2.2], yields the following:

$$\boldsymbol{S}_n^{\widehat{\beta}} = \frac{1}{\sqrt{n}} (\boldsymbol{\mathcal{Y}} \otimes \boldsymbol{I}_d) \xrightarrow{\mathcal{D}} \mathcal{N}_{d^2p}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathcal{Y}} \otimes \boldsymbol{I}_d),$$

where $\boldsymbol{\Sigma}_{\mathcal{Y}}$ denotes the covariance matrix of $\boldsymbol{\mathcal{Y}}$, and $\otimes$ denotes the Kronecker product of two matrices. Let $\|\boldsymbol{M}\|_{\mathrm{op}} = \sup_{\|x\|_2=1} \|\boldsymbol{M}x\|_2$ denote the operator norm of a matrix $\boldsymbol{M}$, where $\|\cdot\|_2$ denote the usual Euclidean norm. Corollary 2.4 from Paper A yields the following result:

Corollary 1.5. *Consider* VAR(1)*-setting* (1.20)*, described above. Let* $\boldsymbol{Z} \sim \mathcal{N}_{d^2}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\mathcal{Y}} \otimes \boldsymbol{I}_d)$*, with* $\boldsymbol{\Lambda}$ *denoting its correlation matrix.Assume* $\|\boldsymbol{A}\|_{\mathrm{op}} < 1$*, and denote*

$$\mathfrak{A}(\boldsymbol{A}) := \begin{cases} (\|\boldsymbol{A}\|_{\mathrm{op}} - \|\boldsymbol{A}\|_{\mathrm{op}}^2)^{-1}, & \textit{if } \|\boldsymbol{A}\|_{\mathrm{op}} < (\sqrt{5}-1)/2, \\ \{(1 - \|\boldsymbol{A}\|_{\mathrm{op}})(1 + \|\boldsymbol{A}\|_{\mathrm{op}}^2)\}^{-1}, & \textit{otherwise.} \end{cases}$$

Then, there exists a universal constant $C > 0$*, such that*

$$\begin{aligned} d_{\mathcal{R}}(\boldsymbol{S}_n^{\widehat{\boldsymbol{\beta}}}, \boldsymbol{Z}) &\leqslant C \log_+^2(d) \log_+ \log_+(d) \frac{\log_+(n)}{\sqrt{n}} \frac{\log_+(\sigma_*)}{\sigma_*} \mathfrak{A}(\boldsymbol{A})^{3/2} \log_+(\mathfrak{A}(\boldsymbol{A})), \\ d_{\mathscr{C}}(\boldsymbol{S}_n^{\widehat{\boldsymbol{\beta}}}, \boldsymbol{Z}) &\leqslant C d^{65/24} n^{-1/2} \frac{1}{\sigma_*(\boldsymbol{\Lambda})^{3/2}} \mathfrak{A}(\boldsymbol{A})^{3/2}, \ \textit{and} \\ d_{\mathcal{W}}(\boldsymbol{S}_n^{\widehat{\boldsymbol{\beta}}}, \boldsymbol{Z}) &\leqslant C d^3 n^{-1/2} \frac{\sigma^*(\boldsymbol{\Sigma}_{\mathcal{Y}})}{\sigma_*(\boldsymbol{\Sigma}_{\mathcal{Y}})} \mathfrak{A}(\boldsymbol{A})^{3/2}. \end{aligned}$$

Why only consider the CLT component? The LLN part involves self-normalised random matrices, and obtaining quantitative Berry–Esseen-type bounds for such self-normalised terms in high-dimensional settings is difficult and an open problem. Recent progress was achieved by Chang et al. [10], who study the Gaussian approximation of the coordinate-wise maximum of high-dimensional self-normalized sums and derive an explicit Berry–Esseen bound, for a problem-specific distance, that scales as $\log^{5/4}(d) n^{-1/8}$, under suitable moment conditions. Remark that this is suboptimal in n-dependence, as the optimal scaling would be $n^{-1/2}$. A full treatment of the self-normalisation problem is beyond the scope of Paper A, as our primary goal here is to demonstrate how our methodology can be applied in a high-dimensional time-series setting.

Related work on the error bounds for stable VAR(p) estimators exists in the univariate case, however the high-dimensional case, is currently an open question: In the continuous-time scalar setting, Kim and Park [28] obtain an optimal Berry–Esseen bound of order $T^{-1/2}$ for the normalised drift estimator $\sqrt{T}(\widehat{\theta} - \theta)$ in a one-dimensional Ornstein–Uhlenbeck process (the continuous-time analogue of a stable VAR(1) with Gaussian innovations). Douissi et al. [19] derive optimal rates of order $n^{-1/2}$ for suitably normalised parameter estimation for a univariate AR(1) process (univariate VAR(1) process). For the high-dimensional continuous-time rates, Ciołek et al. [14] obtain oracle inequalities and error bounds for Dantzig and Lasso estimators of the drift in a multivariate Ornstein–Uhlenbeck process, under sparsity assumptions on the drift matrix, but do not study Berry–Esseen type bounds for the estimator.

Closing Remarks

In Paper A we derive quantitative bounds for distances between a random vector whose components admit a Wiener–Itô chaos expansion and a Gaussian random vector. The resulting bounds are fully explicit in their dependence on the dimension of the random vectors, up to a multiplicative term that reflects the (possible) degeneracy of the covariance matrix of the Gaussian limit; an analogous term is present in related work in the area. Applying these bounds to two concrete settings (Breuer–Major type functionals and parameter estimation in VAR(p) processes) yields, to our knowledge, new results that are fully explicit in their dependence on the sample size n and the dimension d. We also establish a version of Theorem 1.3 for the case where the components have a finite chaos expansion. In this situation, for the hyper-rectangular distance, the dimension dependence can be sharpened, and the corresponding bounds are strictly better than in the general infinite-chaos case.

Our proofs combine techniques from Malliavin calculus and Stein's method. The starting point, for proving our general bounds are three inequalities [23, Thm 1.1], [42, Thm 2.1] and [38, Thm 3.5] involving the hyper-rectangular, the convex and the 1-Wasserstein distance, respectively. These three results provide general bounds via the use of Stein kernels, which, in the setting of multiple Wiener–Itô integrals, get remarkably explicit. The challenging part of the proofs is deconstructing these Stein kernels into more manageable components: we leverage combinatorial ideas from [37, Lem. 6.2.1] combined with insights from [30, Lem. A.1 and Prop. A.2] about sub-Gaussian random variables and [29, Lem. 2.2]. Furthermore, to establish convergence of the series arising in the proofs, the main ingredient is Stirling's inequality (4.18) used in combination with our structural assumption (1.13).

In the applications, the proofs reduce to estimating norms of kernel contractions arising in the quantities that enter assumption (1.13), and then invoking our general bounds. In the case of the OLS estimator of the coefficient matrices in a VAR(p) model, the first step is to represent the relevant CLT component as a finite sum of multiple Wiener–Itô integrals with explicit kernels, and to control the norms of the contractions of these kernels. In fact, the CLT component in this setting fits into a Breuer–Major-type framework.

1.2 Day-Ahead Electricity Price Formation and Structural Modelling

This section summarises the market setting and modelling framework for Paper B, and reviews how day-ahead electricity prices are formed in wholesale markets. Structural models of electricity spot-price[1] formation express prices as functions of observable market drivers, most notably fuel costs, emissions allowance prices, and renewable availability. Understanding how spot prices depend on these inputs matters for market participants and policy-makers, not least because it underpins risk management and hedging in energy markets. As electrification expands across households, transport, industry, and critical infrastructure, short-term power prices can matter for a wider range of cost and operational decisions. Modelling electricity prices is therefore a natural step towards quantifying price risk and supporting robust trading and hedging decisions in modern power markets.

Unlike most commodities, electricity cannot be stored, it must be generated and consumed almost in real time. System operators therefore require that supply and demand are balanced continuously. This need for real-time balance is reflected in a set of markets trading electricity for different delivery times. The central short-term market is typically the day-ahead auction: participants submit bids and offers for delivery in each hour of the

[1] In this thesis, the term "spot price" refers to the day-ahead price. This usage follows common market terminology.

following day, and the market clears at the *day-ahead spot price* S_t for each delivery hour[2] t. As delivery approaches, intraday markets allow participants to adjust their positions as forecasts change (for example, wind and solar forecasts). In addition, longer-dated forward markets (forwards, futures, and related derivatives) allow firms to hedge price risk over horizons from weeks to years. Many additional products exist, but these three layers capture the main timing structure.

A key development over the last decade has been the increasing *share of renewable generation* in the electricity mix. Renewable generation differs from conventional thermal generation: output from gas- or coal-fired plants is generally controllable up to capacity, whereas output from wind and solar depends on weather-variables (e.g., wind speed) and is therefore variable and less predictable. Since many renewables have low marginal operating costs, they are typically offered into the market at low prices and are accepted whenever available. This changes price formation through the *merit-order* mechanism: generation is dispatched in order of increasing marginal cost, and the last accepted (marginal) unit sets the day-ahead spot price. As renewable generation increases, the demand served by conventional plants becomes more volatile, oscillating between periods of near-zero prices (when renewable output is abundant) and periods of elevated prices (when renewable output is limited and expensive conventional capacity is required).

In what follows, "renewables" denotes wind and solar generation, which dominate German electricity supply and day-ahead price formation (the market analysed in Paper B). Hydro, biomass, and other renewable technologies, while present, do not materially affect short-run price dynamics in this setting.[3]

Price Setting via the Merit-Order

In liberalised wholesale day-ahead markets, generators submit supply offers and consumers submit demand bids for each hour of the following day. These offers and bids are collected in a day-ahead auction administered by a market operator, such as EPEX SPOT. Supply offers are ordered from lowest to highest price to construct the aggregate supply curve (the *bid curve* or *bid stack*), while demand bids are ordered from highest to lowest price to form the demand curve. The intersection of these curves identifies the accepted quantity and the corresponding market-clearing price for the hour.

In efficient markets, generators' offers are assumed to reflect marginal costs. These costs depend primarily on the fuel price, the unit's efficiency (often summarised by its heat rate), and the cost of emissions allowances. The *price-setting fuel* is the fuel used by the marginal unit, i.e. the last unit required to meet demand.

At the hourly level, electricity demand is commonly treated as *price-inelastic*. Demand responds weakly to hourly price changes within the relevant time scale, so structural electricity price models typically assume demand as fixed and represent it by a vertical demand curve. Under this approximation, short-run price formation is driven predominantly by the supply side.

Wind and solar plants have very low marginal operating costs, so they typically appear at the bottom of the bid stack and are accepted first when available. When renewable

[2] 15 minutes (as of 1 October 2025).

[3] The technological composition of renewable generation varies across electricity markets; findings here reflect German conditions.

output is higher, less conventional generation is needed to meet demand. The market therefore stops earlier in the bid stack, the marginal accepted unit is cheaper, and the clearing price is lower. This is the *merit-order effect*.

Figure 1.1 schematically illustrates the merit-order effect. The shaded blocks represent the conventional bid stack ordered by marginal cost, and the dotted block represents renewable generation.

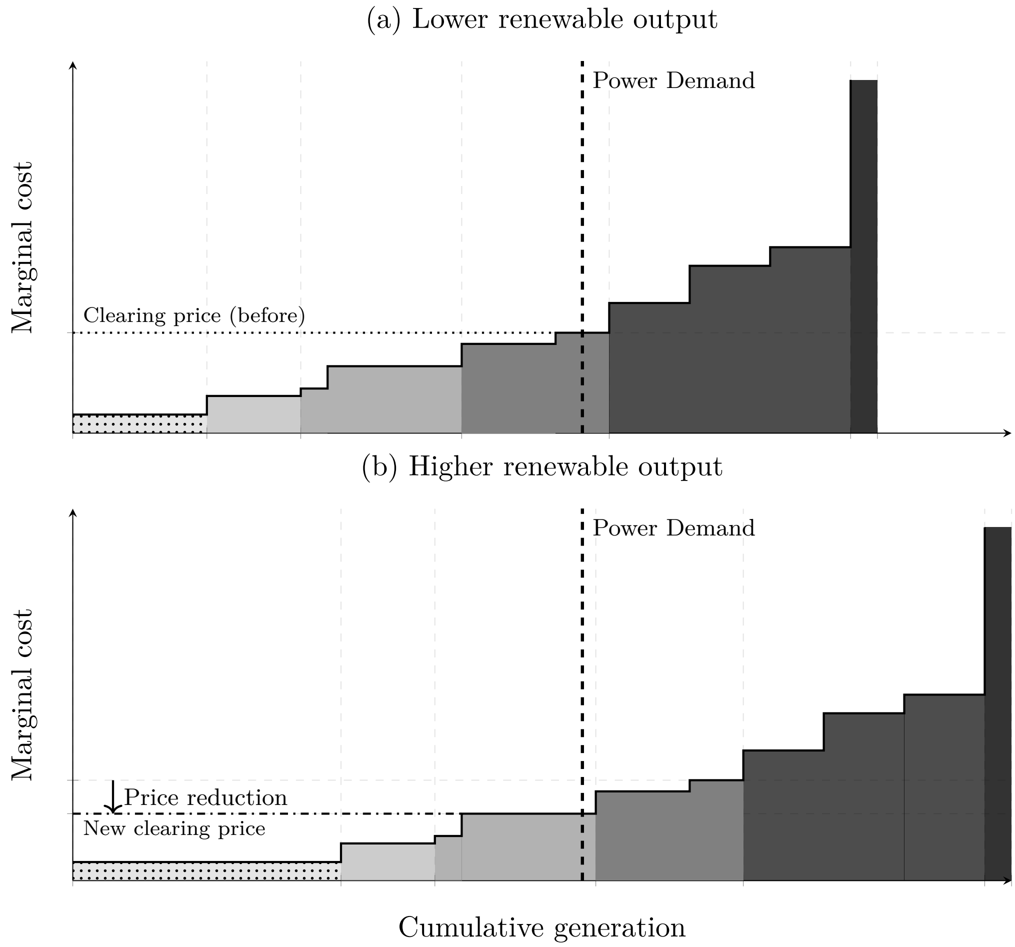


Figure 1.1: Merit-order effect (schematic). Conventional bids form an increasing supply curve (bid stack). For fixed demand (vertical dashed line), the clearing price is set by the marginal accepted unit. Higher renewable generation reduces the demand that must be served by conventional sources (bottom panel), so the clearing point shifts to a lower marginal cost and the clearing price falls.

A convenient summary variable for this mechanism is *residual demand*, defined as $R_t := D_t - W_t$, where D_t denotes total demand and W_t denotes renewable generation in hour t. Since renewables are typically dispatched first due to their low marginal costs, residual demand is the quantity that must be covered by conventional generation. In particular, higher renewable infeed reduces R_t, consistent with the shift in the clearing point in Figure 1.1.

Modelling Approaches in the Literature

The literature on electricity spot prices is commonly organised into three broad classes (see, e.g., [18, 51]): *fundamental*, *reduced-form*, and *structural* models. The distinctions are largely driven by the modelling objective. Fundamental models focus on physical realism, reduced-form models focus on mathematical tractability, and structural models aim to stay economically interpretable while still being usable in practice.

Fundamental models approximate the market-clearing mechanism by solving optimisation problems subject to engineering constraints such as capacity limits, ramp-rate constraints, start-up costs, and network restrictions. They are well suited for scenario analysis and can capture substantial operational detail, but they are computationally demanding and often require extensive re-specification as market design and generation mixes evolve. As a result, obtaining a full price distribution typically requires solving the optimisation problem many times, which can be computationally expensive, particularly for risk management applications that require information about the tails of the distribution.

Reduced-form models specify the spot price directly as a stochastic process (e.g., mean-reverting diffusion with jumps). Their analytical tractability has made them standard in derivative pricing and risk management. However, economic drivers, such as fuel prices, carbon prices, and renewable infeed, typically enter only indirectly (if at all), for instance through time-varying parameters or regime-switching specifications. As a result, reduced-form models often provide limited transparency for questions that are naturally expressed in terms of fundamentals, such as the effect of a change in gas prices or renewable output on the spot price.

Structural models, the focus of Paper B, bridge these approaches by explicitly deriving the spot price $S_t = b(X_t)$ from market equilibrium conditions, where X_t collects the fundamental drivers, such as demand, fuel and emissions costs, and renewable availability. Hence, there are two taks for obtaining a structural dynamic model for spot prices: (i) choosing the function b, typically derived from market equilibrium conditions, and (ii) modelling the dynamics of the driving variables X_t.

Early work by Barlow [3] derives spot-price dynamics from a deterministic supply curve and stochastic demand, capturing stylised features like price spikes. Howison and Coulon [26] and Carmona et al. [9] identify fuel prices F_t as key drivers of medium- to long-term dynamics and propose multi-fuel bid-stack models $S_t = b(D_t, F_t)$ based on merit-order principle. Extensions by Pirrong and Jermakyan [48] and de Maere d'Aertrycke and Smeers [16] develop two-factor models with demand and fuel costs (initially gas, later multiple fuels) as main variables.

Further variants modify the demand component to reflect market design and renewable integration: Wagner [50] model residual demand dynamics within a Barlow-type set-up, and Coskun and Korn [15] employ a Jacobi process to capture bounded demand dynamics in the German market. Several other approaches to structural electricity spot-price modelling exist. For instance, Ziel and Steinert [52] provide a data-driven structural representation of price formation by estimating supply and demand curves from auction bids and using their market-clearing intersection to forecast day-ahead electricity prices. While this approach is grounded in market equilibrium, it does not focus on the causal relationship between spot prices and underlying fundamental drivers such as fuel costs or renewable infeed, which is a key distinction from the parametric structural models reviewed above.

To summarise, the structural literature has developed multi-fuel bid-stack models [9] and residual-demand approaches [50] largely in parallel. Paper B has two primary contributions. First, we extend multi-fuel bid stacks to incorporate renewables via residual demand and characterise when multi-fuel pricing reduces to single-fuel forms (under efficiency and proportional fuel prices). Second, using extensive rolling-origin testing on German market data, we demonstrate that a theory-consistent one-fuel structural model using gas prices, EUA prices, and residual demand as explanatory variables delivers the most robust out-of-sample performance, matching or exceeding both multi-fuel variants and machine-learning benchmarks for the considered dataset.

Paper B: Approach and Main Findings

The aim of Paper B is to build a structural and tractable model of day-ahead electricity prices that remains easy to estimate and, importantly, easy to interpret. Instead of maximizing predictive accuracy via highly flexible machine-learning models, we focus on a parsimonious specification grounded in the price-setting logic of the market. The objective is a model that practitioners can understand, stress-test, and use for scenario analysis.

Modelling assumptions. To obtain a tractable structural model based on the merit-order mechanism, we adopt three standard approximations: (i) hourly demand is treated as price-inelastic; (ii) offers are treated as marginal-cost-based; and (iii) renewables are treated as having zero marginal cost and are therefore dispatched first. The resulting price is the market-clearing price, the price at which accepted supply equals demand in the auction (also called market equilibrium). Under these assumptions, the day-ahead outcome follows merit-order dispatch, with the price set by the marginal accepted unit. These assumptions simplify strategic behaviour, market power, and operational constraints inherent in real markets.

Key Modelling Idea: the Market Bid Stack with Renewables. Building on the bid-stack approach of Carmona et al. [9], we represent the conventional supply stack through fuel-specific bid curves b_i, for fuel i, and their aggregation into a market bid stack b. We denote by $b(\cdot, f)$ the *conventional* market bid stack, i.e. the supply curve constructed from non-renewable offers only (the stack that would apply in the absence of renewables). We interpret $b_i(\xi, f_i)$ as the marginal cost of producing an additional unit when the fuel price is f_i and ξ units have already been dispatched for fuel i.

Renewable generation enters as near-zero marginal cost supply that is dispatched first. As a consequence, renewables affect the clearing price mainly by reducing the demand that must be met by conventional plants. Writing total demand as D_t and renewable generation as W_t, the operative demand variable becomes residual demand R_t.

Formally, Paper B shows that under inelastic demand and market equilibrium, the day-ahead price equals the conventional market bid stack evaluated at residual demand:

$$S_t = b(R_t, F_t).$$

In other words, $b(r, f)$ denotes the marginal cost at supply level r, given fuel prices $f = (f_1, \ldots, f_n)$. Here renewables enter only through R_t; the function b itself describes conventional supply.

In the special case of a one-fuel market, this representation reduces to the multiplicative form

$$S_t = F_t\, q(R_t),$$

where $q(\cdot)$ is the (fuel-specific) heat-rate function that captures how marginal efficiency varies along the fuel-specific bid stack. More generally, when multiple fuels are present but their prices move proportionally—for instance, coal and gas prices satisfying $C_t = \alpha G_t$—the same multiplicative structure persists:

$$S_t = G_t \tilde{q}(R_t),$$

where $\tilde{q}$ is a heat-rate function determined by the marginal fuel mix. The full characterisation is given in Theorem 2.4 in Paper B.

Empirical Strategy. The theory provides the framework $S_t = b(R_t, F_t)$. We use out-of-sample performance to select a robust specification within this framework. We therefore estimate and compare 12 semi-parametric variants, differing in (i) which drivers are included and (ii) the functional form used for the bid stack

We estimate parameters by ordinary least squares and model nonlinear components using natural cubic splines (four internal knots placed at the 20%, 40%, 60%, and 80% quantiles of the residual-demand distribution). This keeps the economic structure explicit while allowing the data to inform the curvature of the stack. Natural cubic splines are recommended by James et al. [27] as a convenient method for capturing nonlinear relationships.

Model performance is evaluated using rolling-origin out-of-sample testing on German day-ahead auction data from 2016–2023. We use the day-ahead price and fundamentals for a fixed delivery hour (12:00), resulting in one observation per day (weekday sample; $n = 1{,}996$ after exclusions). For each training period, models are estimated on one or more consecutive years and evaluated on the following year.

To benchmark the structural specifications, we also include a naïve benchmark $S_t = S_{t-1}$, and simple off-the-shelf machine-learning baselines (including `MLPRegressor` and `AdaBoostRegressor`) trained on the same fundamental feature set as the structural models.

Main Empirical Findings. Across evaluation windows, the strongest and most stable performers are simple multiplicative specifications that combine residual demand with a single fuel,

$$S_t = \overline{G}_t\, q(R_t) \quad \text{and} \quad S_t = \overline{C}_t\, q(R_t),$$

where $\overline{G}_t$ and $\overline{C}_t$ denote carbon-adjusted gas and coal prices. Among these, the gas-based model is the most robust overall, especially in the more recent part of the sample. More general multi-fuel variants can fit in-sample well but tend to be less stable out of sample, which is suggestive of overfitting when the sample size is limited.

Theorem 2.4 states that, under marginal-cost-based offers, the price takes a multiplicative form in specific settings. In a one-fuel market, the clearing price can be written as $S_t = F_t q(R_t)$, i.e. a fuel-price term and a residual-demand term. In a multi-fuel market, the same factorisation continues to hold when fuel prices move proportionally, so the price can still be expressed using a single fuel-price index and a residual-demand function. Hence, even when more than one fuel can be marginal, residual demand together with one (possibly carbon-adjusted) fuel-price index can capture the main price-setting mechanism.

A practical implication is that transparent, theory-guided models can match (and in some folds exceed) the predictive accuracy of flexible machine-learning baselines on this dataset, while remaining easier to interpret and to communicate in terms of residual demand and marginal input costs.

Table 1.1 summarises the best-performing specifications, and Table 1.2 reports their RMSE values across rolling-origin evaluation periods.

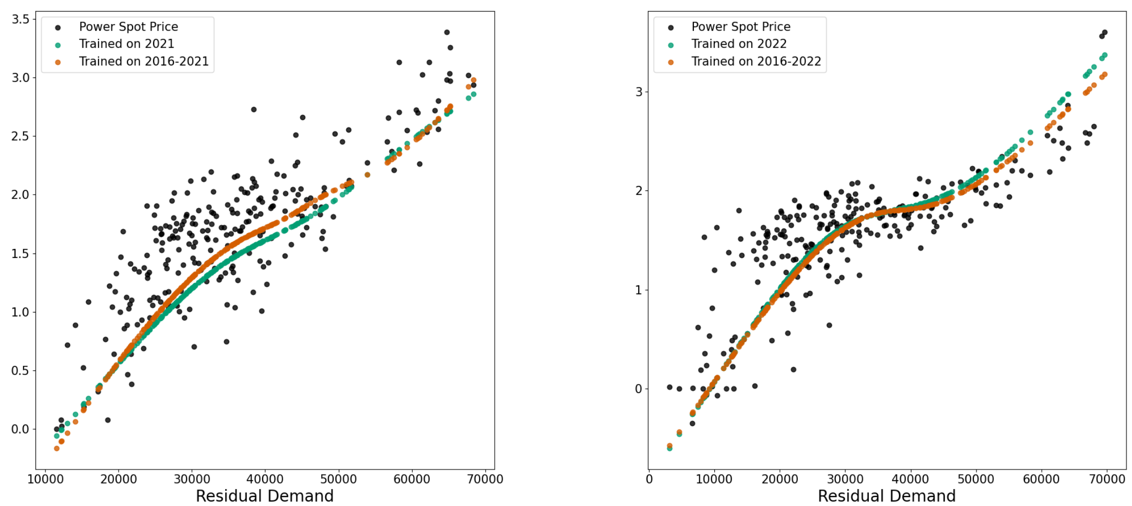
Model	Formula
$M_{1.1}^{G}$	$S_t = \overline{G}_t q(R_t)$
$M_{2.2}^{G}$	$S_t = G_t q(R_t)$
$M_{1.1}^{C}$	$S_t = \overline{C}_t q(R_t)$
M_1^{ML}	`MLPRegressor`

Table 1.1: Structural models sublibrary

Trained on	**Predicting**	M_1^{ML}	$M_{1.1}^{C}$	$M_{1.1}^{G}$	$M_{2.2}^{G}$
2016-2017	2018	11.11	8.39	8.16	8.22
2016-2018	2019	12.79	7.57	10.38	15.08
2016-2019	2020	12.29	7.53	10.75	16.13
2016-2020	2021	32.26	52.52	23.59	24.96
2018-2020	2021	37.00	55.48	24.37	27.91
2019-2020	2021	37.58	57.15	36.04	52.90
2016-2021	2022	63.53	135.02	63.99	56.04
2018-2021	2022	52.69	134.96	63.50	54.74
2019-2021	2022	56.59	134.33	61.92	52.70
2020-2021	2022	53.10	133.42	66.29	56.46
2016-2022	2023	33.72	55.68	21.31	27.36
2018-2022	2023	33.12	57.54	21.52	27.41
2019-2022	2023	34.87	59.80	21.66	27.40
2020-2022	2023	30.19	60.50	21.95	28.14
2021-2022	2023	30.33	65.22	21.63	27.59

Table 1.2: RMSE values for selected models.

The Estimated Heat Rate Function. Figure 1.2 shows the estimated heat-rate function q from the best-performing model $S_t = \overline{G}_t q(R_t)$. The estimate differs from common parametric shapes (such as exponential or power-function). A plausible explanation for the steep increase when the residual demand is low is the enrty of additional marginal technologies (e.g., lignite), which introduce jumps in the theoretical supply stack; the spline provides a smoothed approximation of these features. However, it is important to note that the endpoints of the residual-demand range contain relatively few observations. Because natural cubic splines impose linear behavior at the boundaries, the steep endpoint behavior may partly reflect this constraint rather than underlying structure. Predictions at very low and high residual demand should therefore be interpreted with caution.



(a) Single-year training (2021, green) vs. multi-year training (2016–2021, orange)

(b) Single-year training (2022, green) vs. multi-year training (2016–2022, orange).

Figure 1.2: Estimated function q from model ($S_t = \overline{G}_t q(R_t)$). The plotted curves correspond to the implied ratio $q(R_t) = S_t/\overline{G}_t$, shown for different training windows to assess stability.

References

[1] M. A. Arcones. Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.*, 22(4):2242–2274, 1994. ISSN 0091-1798,2168-894X. URL http://links.jstor.org/sici?sici=0091-1798(199410)22:4<2242:LTFNFO>2.0.CO;2-L&origin=MSN.

[2] S. Bai and M. S. Taqqu. Multivariate limit theorems in the context of long-range dependence. *J. Time Series Anal.*, 34(6):717–743, 2013. ISSN 0143-9782,1467-9892. doi: 10.1111/jtsa.12046. URL https://doi.org/10.1111/jtsa.12046.

[3] M. T. Barlow. A diffusion model for electricity prices. *Mathematical finance*, 12(4): 287–298, 2002.

[4] V. Bentkus. On the dependence of the Berry-Esseen bound on dimension. *J. Statist. Plann. Inference*, 113(2):385–402, 2003. ISSN 0378-3758. doi: 10.1016/S0378-3758(02)00094-0. URL https://doi.org/10.1016/S0378-3758(02)00094-0.

[5] V. Bentkus. A Lyapunov type bound in $\mathbf{R}^d$. *Teor. Veroyatn. Primen.*, 49(2):400–410, 2004. ISSN 0040-361X. doi: 10.1137/S0040585X97981123. URL https://doi.org/10.1137/S0040585X97981123.

[6] A. C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.*, 49:122–136, 1941. ISSN 0002-9947. doi: 10.2307/1990053. URL https://doi.org/10.2307/1990053.

[7] P. Breuer and P. Major. Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.*, 13(3):425–441, 1983. ISSN 0047-259X. doi: 10.1016/0047-259X(83)90019-2. URL https://doi.org/10.1016/0047-259X(83)90019-2.

[8] S. Campese, I. Nourdin, and D. Nualart. Continuous Breuer-Major theorem: tightness and nonstationarity. *Ann. Probab.*, 48(1):147–177, 2020. ISSN 0091-1798. doi: 10.1214/19-AOP1357. URL https://doi.org/10.1214/19-AOP1357.

[9] R. Carmona, M. Coulon, and D. Schwarz. Electricity price modeling and asset valuation: a multi-fuel structural approach. *Mathematics and Financial Economics*, 7: 167–202, 2013.

[10] W. Chang, K. Takatsu, K. Urban, and A. K. Kuchibhotla. The berry-esseen bound for high-dimensional self-normalized sums, 2025. URL https://arxiv.org/abs/2501.08979.

[11] V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximation of suprema of empirical processes. *Ann. Statist.*, 42(4):1564–1597, 2014. ISSN 0090-5364. doi: 10.1214/14-AOS1230. URL https://doi.org/10.1214/14-AOS1230.

[12] V. Chernozhukov, D. Chetverikov, and K. Kato. Central limit theorems and bootstrap in high dimensions. *Ann. Probab.*, 45(4):2309–2352, 2017. ISSN 0091-1798. doi: 10.1214/16-AOP1113. URL https://doi.org/10.1214/16-AOP1113.

[13] V. Chernozhukov, D. Chetverikov, and Y. Koike. Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.*, 33(3):2374–2425, 2023. ISSN 1050-5164. doi: 10.1214/22-aap1870. URL https://doi.org/10.1214/22-aap1870.

[14] G. Ciołek, D. Marushkevych, and M. Podolskij. On Dantzig and Lasso estimators of the drift in a high dimensional Ornstein-Uhlenbeck model. *Electron. J. Stat.*, 14(2):4395–4420, 2020. ISSN 1935-7524. doi: 10.1214/20-EJS1775. URL https://doi.org/10.1214/20-EJS1775.

[15] S. Coskun and R. Korn. Modeling the intraday electricity demand in germany. *Mathematical modeling, simulation and optimization for power engineering and management*, pages 3–23, 2021.

[16] G. de Maere d'Aertrycke and Y. Smeers. The valuation of power futures based on optimal dispatch. *The Journal of Energy Markets*, 3(3):27, 2010.

[17] H. Dehling, A. Rooch, and M. S. Taqqu. Non-parametric change-point tests for long-range dependent data. *Scand. J. Stat.*, 40(1):153–173, 2013. ISSN 0303-6898,1467-9469. doi: 10.1111/j.1467-9469.2012.00799.x. URL https://doi.org/10.1111/j.1467-9469.2012.00799.x.

[18] T. Deschatre, O. Féron, and P. Gruet. A survey of electricity spot and futures price models for risk management applications. *Energy Economics*, 102:105504, 2021.

[19] S. Douissi, K. Es-Sebaiy, F. Alshahrani, and F. G. Viens. AR(1) processes driven by second-chaos white noise: Berry-Esséen bounds for quadratic variation and parameter estimation. *Stochastic Process. Appl.*, 150:886–918, 2022. ISSN 0304-4149. doi: 10.1016/j.spa.2020.02.007. URL https://doi.org/10.1016/j.spa.2020.02.007.

[20] C.-G. Esseen. On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astr. Fys.*, 28A(9):19, 1942.

[21] C.-G. Esseen. Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.*, 77:1–125, 1945. ISSN 0001-5962. doi: 10.1007/BF02392223. URL https://doi.org/10.1007/BF02392223.

[22] C.-G. Esseen. A moment inequality with an application to the central limit theorem. *Scandinavian Actuarial Journal*, 1956(2):160–170, 1956.

[23] X. Fang and Y. Koike. High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.*, 31(4):1660–1686, 2021. ISSN 1050-5164. doi: 10.1214/20-aap1629. URL https://doi.org/10.1214/20-aap1629.

[24] F. Götze. On the rate of convergence in the multivariate central limit theorem. *Ann. Probab.*, 19(2):724–739, 1991. doi: 10.1214/aop/1176991339.

[25] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, 1994. ISBN 9780691218632. doi: doi:10.1515/9780691218632. URL https://doi.org/10.1515/9780691218632.

[26] S. Howison and M. Coulon. Stochastic behaviour of the electricity bid stack: from fundamental drivers to power prices. *The Journal of Energy Markets*, 2:29–69, 2009.

[27] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor. *An introduction to statistical learning: With applications in python*. Springer Nature, 2023.

[28] Y. T. Kim and H. S. Park. Optimal Berry-Esseen bound for an estimator of parameter in the Ornstein-Uhlenbeck process. *J. Korean Statist. Soc.*, 46(3):413–425, 2017. ISSN 1226-3192. doi: 10.1016/j.jkss.2017.01.002. URL https://doi.org/10.1016/j.jkss.2017.01.002.

[29] Y. Koike. Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data. *Ann. Statist.*, 47(3):1663–1687, 2019. ISSN 0090-5364. doi: 10.1214/18-AOS1731. URL https://doi.org/10.1214/18-AOS1731.

[30] Y. Koike. Supplement to "gaussian approximation of maxima of wiener functionals and its application to high-frequency data.", 2019.

[31] C. Lévy-Leduc, H. Boistard, E. Moulines, M. S. Taqqu, and V. A. Reisen. Asymptotic properties of U-processes under long-range dependence. *Ann. Statist.*, 39(3):1399–1426, 2011. ISSN 0090-5364,2168-8966. doi: 10.1214/10-AOS867. URL https://doi.org/10.1214/10-AOS867.

[32] H. Lütkepohl. *New introduction to multiple time series analysis*. Springer-Verlag, Berlin, 2005. ISBN 3-540-40172-5. doi: 10.1007/978-3-540-27752-1. URL https://doi.org/10.1007/978-3-540-27752-1.

[33] L. Maini and I. Nourdin. Spectral central limit theorem for additive functionals of isotropic and stationary Gaussian fields. *Ann. Probab.*, 52(2):737–763, 2024. ISSN 0091-1798. doi: 10.1214/23-aop1669. URL https://doi.org/10.1214/23-aop1669.

[34] D. Marinucci. The empirical process for bivariate sequences with long memory. *Stat. Inference Stoch. Process.*, 8(2):205–223, 2005. ISSN 1387-0874,1572-9311. doi: 10.1007/s11203-004-2790-9. URL https://doi.org/10.1007/s11203-004-2790-9.

[35] I. Nourdin and D. Nualart. The functional Breuer-Major theorem. *Probab. Theory Related Fields*, 176(1-2):203–218, 2020. ISSN 0178-8051. doi: 10.1007/s00440-019-00917-1. URL https://doi.org/10.1007/s00440-019-00917-1.

[36] I. Nourdin and G. Peccati. Stein's method on Wiener chaos. *Probab. Theory Related Fields*, 145(1-2):75–118, 2009. ISSN 0178-8051. doi: 10.1007/s00440-008-0162-x. URL https://doi.org/10.1007/s00440-008-0162-x.

[37] I. Nourdin and G. Peccati. *Normal approximations with Malliavin calculus*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2012. ISBN 978-1-107-01777-1. doi: 10.1017/CBO9781139084659. URL https://doi.org/10.1017/CBO9781139084659. From Stein's method to universality.

[38] I. Nourdin, G. Peccati, and A. Réveillac. Multivariate normal approximation using Stein's method and Malliavin calculus. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46 (1):45–58, 2010. ISSN 0246-0203. doi: 10.1214/08-AIHP308. URL https://doi.org/10.1214/08-AIHP308.

[39] I. Nourdin, G. Peccati, and M. Podolskij. Quantitative Breuer-Major theorems. *Stochastic Process. Appl.*, 121(4):793–812, 2011. ISSN 0304-4149. doi: 10.1016/j.spa.2010.12.006. URL https://doi.org/10.1016/j.spa.2010.12.006.

[40] I. Nourdin, G. Peccati, and Y. Swan. Entropy and the fourth moment phenomenon. *J. Funct. Anal.*, 266(5):3170–3207, 2014. ISSN 0022-1236. doi: 10.1016/j.jfa.2013.09.017. URL https://doi.org/10.1016/j.jfa.2013.09.017.

[41] I. Nourdin, G. Peccati, and X. Yang. Berry-Esseen bounds in the Breuer-Major CLT and Gebelein's inequality. *Electron. Commun. Probab.*, 24:Paper No. 34, 12, 2019. doi: 10.1214/19-ECP241. URL https://doi.org/10.1214/19-ECP241.

[42] I. Nourdin, G. Peccati, and X. Yang. Multivariate normal approximation on the Wiener space: new bounds in the convex distance. *J. Theoret. Probab.*, 35(3):2020–2037, 2022. ISSN 0894-9840. doi: 10.1007/s10959-021-01112-6. URL https://doi.org/10.1007/s10959-021-01112-6.

[43] D. Nualart. *Malliavin calculus and its applications*, volume 110 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2009. ISBN 978-0-8218-4779-4. doi: 10.1090/cbms/110. URL https://doi.org/10.1090/cbms/110.

[44] D. Nualart and E. Nualart. *Introduction to Malliavin calculus*. Cambridge University Press, 2018.

[45] D. Nualart and G. Peccati. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.*, 33(1):177–193, 2005. ISSN 0091-1798. doi: 10.1214/009117904000000621. URL https://doi.org/10.1214/009117904000000621.

[46] D. Nualart and A. Tilva. Continuous Breuer-Major theorem for vector valued fields. *Stoch. Anal. Appl.*, 38(4):668–685, 2020. ISSN 0736-2994. doi: 10.1080/07362994.2019.1711118. URL https://doi.org/10.1080/07362994.2019.1711118.

[47] D. Nualart and H. Zhou. Total variation estimates in the Breuer-Major theorem. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(2):740–777, 2021. ISSN 0246-0203. doi: 10.1214/20-aihp1094. URL https://doi.org/10.1214/20-aihp1094.

[48] C. Pirrong and M. Jermakyan. The price of power: The valuation of power and weather derivatives. *Journal of Banking & Finance*, 32(12):2520–2529, 2008.

[49] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602. Univ. California Press, Berkeley, CA, 1972.

[50] A. Wagner. Residual demand modeling and application to electricity pricing. *The Energy Journal*, 35(2):45–74, 2014.

[51] R. Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International journal of forecasting*, 30(4):1030–1081, 2014.

[52] F. Ziel and R. Steinert. Electricity price forecasting using sale and purchase curves: The x-model. *Energy Economics*, 59:435–454, 2016.

Paper A

Quantitative Bounds for High-Dimensional Random Vectors on Gaussian Spaces

Andreas Basse-O'Connor, Lota Copić, David Kramer-Bang

Abstract

We derive explicit dimension-dependent bounds for multivariate Gaussian approximations of random vectors admitting Wiener–Itô chaos expansions, with a particular focus on high-dimensional regimes. Our estimates control hyper-rectangular, convex and 1-Wasserstein distances via Malliavin–Stein approach and require only a natural contraction condition on the kernels governing the Wiener–Itô expansion. The developed methodology is applied to obtain a quantitative Breuer–Major theorem for vector-valued functions of multivariate stationary Gaussian sequences, and finite-sample Gaussian approximation bounds for the CLT component of parameter estimation for a stable VAR(p).

1 Introduction

Quantitative central limit theorems (CLTs) lie at the core of probability theory and play a central role in modern statistical inference. Classical results, such as the Berry–Esseen theorem [7, 24], yield sharp error bounds for sums of independent random variables (dependence on sample size n is $n^{-1/2}$). However, modern statistical applications such as deep neural networks (see [26, 8] and references therein), increasingly feature high-dimensional dependent data [13, 18] and nonlinear functionals of stationary Gaussian sequences [41], for which classical limit theorems require substantial extensions. For such applications, understanding explicitly how approximation error depends on both the dimension d and the sample size n has become essential.

For independent random vectors, Bentkus [5, 6] obtained the dimension-explicit Berry–Esseen bounds with rates of order $d^{1/4}n^{-1/2}$ in convex distance. Subsequently, Chernozhukov, Chetverikov, and Kato [16, 17] focused on the *hyper-rectangular distance*, obtaining rates of order $\log^{7/6}(d)n^{-1/6}$ for normalized sums of independent, centred d-dimensional random vectors. While this dependence on d is logarithmic, the rate remains suboptimal in the sample size. Under additional boundedness assumptions, Chernozhukov, Chetverikov, and Koike [19] improved the rate to $\log^{3/2}(d)n^{-1/2}\log(n)$, which is optimal in n up to logarithmic factors. Fang and Koike [25] developed a general framework for the hyper-rectangular distance and random vectors that admit Stein kernel. Specializing to normalized sums of independent and identically distributed (i.i.d) random vectors with long-concave densities, they obtained a rate of order $\log^{4/3}(dn)n^{-1/3}$. They further established logarithmic dimension dependence for vectors belonging to a Wiener chaos of fixed order, however the associated convergence rates in n remain suboptimal relative to the result of Berry–Esseen[7, 24].

In contrast, for dependent data arising as functionals of Gaussian processes, the fusion of Stein's method [51, 15] with Malliavin calculus, introduced by Nourdin and Peccati [38] and extended to the multivariate setting in [40], has led to a versatile framework for quantitative Gaussian approximations for functionals of Gaussian processes. Unlike classical Berry–Esseen theory, the Malliavin–Stein approach is well suited to analysing dependent structures by exploiting Wiener chaos decompositions. Notably, the approach underlies the seminal Fourth Moment Theorem [38, 47] and quantitative versions of the Breuer–Major CLT for nonlinear functionals of Gaussian processes [9, 41].

Existing Malliavin–Stein bounds are usually either not fully explicit (see [44, Thm 1.2] or [40, Thm 3.5]) or provided for a Wiener chaos of fixed order [44, Cor. 1.3], leaving a gap: dimension-explicit quantitative bounds for infinite chaos expansions remain absent from the literature.

A central motivating example for developing a dimension-explicit Malliavin–Stein theory is the *quantitative Breuer–Major theorem*. While it has been generalised and refined in various directions both quantitative [41, 43, 49], and qualitative [11, 12, 37, 35], none of these works yield dimension-explicit, vector-valued bounds for general nonlinear mappings $\Phi : \mathbb{R}^K \to \mathbb{R}^d$. The corresponding qualitative result in the continuous setting has been established by Nualart and Tiva [48], yet, to the best of the authors' knowledge, a quantitative counterpart is still missing. One of the main contributions of this paper is to fill this gap, providing multivariate Breuer–Major bounds with explicit rates in the input dimension K, output dimension d, and the sample size n. The same structure, namely partial sums of nonlinear functionals of a stationary Gaussian sequence with a given Hermite rank, appears in many statistical applications, especially those involving long-range dependent Gaussian inputs,including U-statistics and robust functionals [33, 22], sequential change-point tests [21], empirical and related processes for long-memory data [36, 52], and multivariate nonlinear functionals under long-range dependence [29, 2].

In this paper, we develop a framework for dimension-explicit quantitative Gaussian approximation of high-dimensional vectors whose components admit (possibly) infinite Wiener chaos expansions. Our results build upon the general Malliavin–Stein bounds established for the hyper-rectangular, convex, and 1-Wasserstein distances in [25, Thm. 1.1], [44, Thm. 1.2], and [40, Thm. 3.5], respectively. These three distances, which serve as our primary measures of Gaussian approximation throughout the paper, are recalled in Section 1.2.

The core technical component of our approach is a regularity condition on contraction of the kernels associated with the random vector $\boldsymbol{F}$ (see 1.1), balancing exponential growth and factorial decay:

$$||f_{i,p} \otimes_r f_{j,q}||_{\mathfrak{H}^{\otimes(p+q-2r)}} \leqslant \frac{\gamma e^{\alpha(p+q)}}{(p!q!)^\beta}, \quad \text{for all } i,j \in \{1,\dots,d\}, \text{ and appropriate } r,p,q. \tag{1.1}$$

In (1.1), the parameter $\beta \in [1/2, 1]$ controls the decay rate in the factorial term. When $\beta > 1/2$, the condition is automatically satisfied for any fixed α, whereas at the boundary

$\beta = 1/2$ we require α to be sufficiently negative, for explicit conditions, see Theorem 2.1. Under (1.1), our main contributions are as follows.

(1) Theorem 2.1 provides a general quantitative normal approximation result applicable to random vectors whose components admit Wiener–Itô chaos expansion. For hyper-rectangular distance, the bound exhibits sub-polynomial growth in d when $\beta > 1/2$, and polynomial growth d^k for $k \approx 0.368$ at threshold $\beta = 1/2$. We also derive related bounds for convex and Wasserstein distances with fixed polynomial rates $d^{65/24}$ and $d^{3/2}$ respectively. Whenever the use of the hyper-rectangular metric is appropriate, it therefore provides strictly sharper dimension dependence. The regularity parameters α and β affect only the hyper-rectangular bound: higher regularity of the vector $\boldsymbol{F}$ yields faster convergence.A detailed discussion of these effects is given in Remark 2.2 following Theorem 2.1.

(2) Theorem 2.3 and Corollary 2.4 establish the first, to our knowledge, dimension-explicit quantitative version of the multivariate Breuer–Major theorem for nonlinear mappings $\Phi : \mathbb{R}^K \to \mathbb{R}^d$. The resulting bounds are explicit in all relevant parameters: the input dimension K, the output dimension d, and the sample size n.

(3) In Theorem 3.4, we establish explicit convergence rates for the CLT component of the parameter estimator in a stable vector autoregressive process (VAR), with bounds explicit in d, p and n.

We also present a version of Theorem 2.1, where the vector entries admit finite chaos expansions in Theorem 3.1. In this case, for the hyper-rectangular distance, the dimension dependence is strictly better, namely the bounds scale as $(\log(d))^q$ with d, where q is the maximal chaos order. Specifically, Theorem 3.1 extends the recent work of Fang and Koike [25, Cor. 1.2] to allow heterogeneous chaos orders. For Theorem 3.1 and Theorem 3.4, there is no need for the assumption (1.1). Indeed, in the general case, Theorem 2.1, there is an infinite sum that appears in the proof (see (4.11) and (4.12) below) and the assumption (1.1) is needed to ensure the sum converges. As in Theorem 3.1 and Theorem 3.4 the components of the random vector in consideration have finite chaos expansion, the sum equivalents to (4.11) and (4.12) are finite and hence there is no need for the extra assumption.

1.1 Structure of the Paper

The paper is organised as follows. Subsection 1.2 introduces the notation and key objects used throughout the paper. Section 2 presents our main quantitative bound (Theorem 2.1) and its primary application: a multivariate quantitative Breuer–Major theorem for vector-valued functionals (Theorem 2.3). Section 3 establishes a finite-analogue Theorem 3.1, while Subsection 3.2 presents dimension-explicit convergence rates for the CLT component of the multivariate least squares estimator in a stable VAR(p) model. Section 4 provides the necessary technical background and develops a general framework (Theorem 4.1 and Corollary 4.2) from which Theorem 2.1 is derived. This section also contains the proofs of the remaining results. Auxiliary arguments of independent technical interest are deferred to the Appendix.

1.2 Preliminaries

We begin by fixing the notation and basic definitions used throughout the paper. Additional background material is collected in Section 4.1. Denote $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and let $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{R}$ denote the sets of natural, whole and real numbers, respectively. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathfrak{H}$ be a real separable Hilbert space, and let $\boldsymbol{X} = \{X(h) : h \in \mathfrak{H}\}$ denote an isonormal Gaussian process over $\mathfrak{H}$, i.e., a centred Gaussian family satisfying $\mathbb{E}[X(h)X(g)] = \langle h, g\rangle_{\mathfrak{H}}$ for all $h, g \in \mathfrak{H}$. Throughout this paper we assume $\mathcal{F} = \sigma(\boldsymbol{X})$ and write $L^2(\Omega)$ for the space of square-integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Every random element is assumed to be defined on this probability space. For a random variable X, we denote $\|X\|_{L^2}^2 = \mathbb{E}[|X|^2]$.

A $f : \mathbb{R}^K \to \mathbb{R}$ belongs to $L^2(\gamma^K, \mathbb{R})$, if $\mathbb{E}[|f(G)|^2] < \infty$. For $\varphi \in L^2(\gamma^K, \mathbb{R})$, its *Hermite rank* $m \in \mathbb{N}$ is defined as the smallest integer such that $\mathbb{E}[\varphi(\boldsymbol{G})p(\boldsymbol{G})] = 0$ fails to hold for all polynomials p of degree $\leqslant m-1$. The Hermite rank plays a crucial role in the Breuer–Major theorem, as it determines the minimal chaos order contributing to the limit.

The *q-th Wiener chaos* $\mathcal{H}_q$ associated with $\boldsymbol{X}$ is the closed linear subspace of $L^2(\Omega)$ generated by $\{H_q(X(h)) : h \in \mathfrak{H}, \|h\|_{\mathfrak{H}} = 1\}$, where H_q denotes the q-th Hermite polynomial (see Section 4.1 for definition). By convention, $\mathcal{H}_0 = \mathbb{R}$. The spaces $\mathcal{H}_p$ and $\mathcal{H}_q$ are orthogonal for $p \neq q$, and they span the $L^2(\Omega)$ space, see [39, Thm 2.2.4].

Denote by $\mathfrak{H}^{\odot q}$ the q-fold symmetric tensor product of $\mathfrak{H}$. For $f \in \mathfrak{H}^{\odot q}$, the *$q$-th multiple Wiener–Itô integral* of f with respect to $\boldsymbol{X}$ is denoted $I_q(f)$, with the convention $I_0(f_0) = c$ for constants $c \in \mathbb{R}$. Every $F \in L^2(\Omega)$ admits a unique *Wiener–Itô chaos expansion*

$$F = \mathbb{E}[F] + \sum_{q=1}^{\infty} I_q(f_q), \tag{1.2}$$

where $f_q \in \mathfrak{H}^{\odot q}$ are uniquely determined by F and the series converges in $L^2(\Omega)$; see [39, Cor. 2.7.8]. Define the space of Malliavin-differentiable functions $\mathbb{D}^{1,2} := \mathbb{D}^{1,2}(\Omega) = \{F \in L^2(\Omega) : \sum_{q=1}^{\infty} qq!\|f_q\|_{\mathfrak{H}^{\otimes q}}^2 < \infty\}$, see [39, Sec. 2.3 and 2.7]. For $f \in \mathfrak{H}^{\odot p}$ and $g \in \mathfrak{H}^{\odot q}$ with $p, q \geqslant 1$, the *r-th contraction* $f \otimes_r g \in \mathfrak{H}^{\otimes(p+q-2r)}$ is defined for $r = 0, 1, \ldots, p \wedge q$, with $f \otimes_0 \mathbb{1}_g = f \otimes g$ denoting the ordinary tensor product. The symmetrisation of $f \otimes_r g$ is denoted $f \widetilde{\otimes}_r g$. See [39, App. B.4] for further details.

For random vectors $\boldsymbol{F}, \boldsymbol{Z}$ with values in $\mathbb{R}^d$, we consider three probabilistic distances, convex, $d_{\mathscr{C}}$, 1-Wasserstein, $d_{\mathcal{W}}$, and hyper-rectangular, $d_{\mathcal{R}}$:

$$\begin{aligned}
d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) &:= \sup_{A \in \mathcal{R}} \left|\mathbb{P}(\boldsymbol{F} \in A) - \mathbb{P}(\boldsymbol{Z} \in A)\right|, \quad \mathcal{R} := \left\{ \bigtimes_{i=1}^{d} (a_i, b_i) : -\infty \leqslant a_i \leqslant b_i \leqslant \infty \right\}, \\
d_{\mathscr{C}}(\boldsymbol{F}, \boldsymbol{Z}) &:= \sup_{A \in \mathscr{C}} |\mathbb{P}(\boldsymbol{F} \in A) - \mathbb{P}(\boldsymbol{Z} \in A)|, \quad \mathscr{C} := \{A \subset \mathbb{R}^d : A \text{ convex}\}, \\
d_{\mathcal{W}}(\boldsymbol{F}, \boldsymbol{Z}) &:= \sup_{h \in \mathrm{Lip}(1)} |\mathbb{E}[h(\boldsymbol{F})] - \mathbb{E}[h(\boldsymbol{Z})]|, \quad \mathrm{Lip}(1) := \{h : \mathbb{R}^d \to \mathbb{R} \text{ is 1-Lipschitz}\}.
\end{aligned}$$

We say that a function h is 1-Lipschitz if $\mathrm{Lip}(1) := \{h : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz continuous with constant $\leqslant 1\}$. In general it holds that: (i) $d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant d_{\mathscr{C}}(\boldsymbol{F}, \boldsymbol{Z})$, as the supremum runs over all convex sets in $\mathbb{R}^d$ for $d_{\mathscr{C}}$ case, compared to only hyper-rectangulars in $\mathbb{R}^d$ for $d_{\mathcal{R}}$, and (ii) $d_{\mathscr{C}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant K\sqrt{d_{\mathcal{W}}(\boldsymbol{F}, \boldsymbol{Z})}$, for a positive universal constant K. See [44, Eq. (3)].

For $x \geqslant 0$, let $\log_+(x) = |\log x| \vee 1$. We write $\xrightarrow{\mathcal{D}}$ for convergence in distribution and $\mathbb{P}$- lim for convergence in probability. Vectors and matrices are denoted in boldface, e.g. $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{A} \in \mathbb{R}^{d\times d}$; matrix entries are written as $\boldsymbol{A}_{ij}$, and when the matrix carries a subscript, such as $\boldsymbol{\Sigma}_u$, we write entries as $(\boldsymbol{\Sigma}_u)_{ij}$. The notation $\mathrm{Cov}(\boldsymbol{F})$ refers to the covariance matrix of a random vector $\boldsymbol{F}$, while $\mathrm{Cov}(F_i, F_j)$ denotes the covariance between its components F_i and F_j. For a matrix $\boldsymbol{A} \in \mathbb{R}^{d\times d}$, we set $\overline{\sigma}(\boldsymbol{A}) = \max_{1\leqslant j\leqslant d} \boldsymbol{A}_{jj}$ and $\underline{\sigma}(\boldsymbol{A}) = \min_{1\leqslant j\leqslant d} \boldsymbol{A}_{jj}$, and denote by $\sigma_*(\boldsymbol{A})$ and $\sigma^*(\boldsymbol{A})$ its smallest and largest eigenvalues, respectively. The Kronecker product is written as $\boldsymbol{A} \otimes \boldsymbol{B}$ (see [34, App. A.11]). For a vector $\boldsymbol{x}$, the Euclidean norm is $\|\boldsymbol{x}\|_2$. For a matrix $\boldsymbol{A}$, the operator (spectral) norm is $\|\boldsymbol{A}\|_{\mathrm{op}} = \sup_{\|x\|_2=1} \|\boldsymbol{A}x\|_2$, the Hilbert–Schmidt norm is $\|\boldsymbol{A}\|^2_{\mathrm{H.S.}} = \sum_{i,j=1}^d \boldsymbol{A}^2_{ij}$, and the maximum norm is $\|\boldsymbol{A}\|_{\max} = \max_{1\leqslant i,j\leqslant d} \boldsymbol{A}_{ij}$.

Notation $\boldsymbol{Z} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means that $\boldsymbol{Z}$ is a d-dimensional Gaussian random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let $W(x)$ denote the Lambert W-function, then define constant α_0 used in the main result, as follows

$$\alpha_0 := (1/2)\log[W(e^{1/e-1})/(16e^{1+5/(2e)})] \approx -2.846. \tag{1.3}$$

We finish this subsection with a crucial function that describes d-dependence for hyper-rectangular distance in Theorem 2.1. Fix $\beta \in [1/2, 1]$ and $\alpha \in \mathbb{R}$. For $d, K \in \mathbb{N}$, define

$$\psi_{\alpha,\beta}(d) := \log_+^{1/(2\beta)}(d) e^{k_1 \log_+^{1/(2\beta)}(d)}, \text{ and} \tag{1.4}$$

$$\zeta(d,K) := \begin{cases} \log_+^{1/(2\beta)}(d) e^{c_2 K^{1/(2\beta)} \log_+^{1/(2\beta)}(d)}, & \beta \in (1/2, 1], \\ \log_+(d) d^{k_2}(1 - v_1(K))^{-1}, & \beta = 1/2, \end{cases} \tag{1.5}$$

where $v_1(K)$ is given explicitly in (2.7), and the constants k_1, k_2, c_2 are given as follows: $k_2 = 16\exp(2\alpha_0 + 1 + 9/(2e)) \approx 0.768$, $c_2 = \exp\{2\beta e^{1/(2e)}(2^{5/2}e^{\alpha+1/2+2/e})^{1/\beta}\}$, and

$$k_1 = \begin{cases} 16\exp\left(2\alpha + 1 + \frac{5}{2e}\right), & \beta = \frac{1}{2}, \\ \exp\left\{2\beta e^{1/(2e)}(32e^{\alpha+1/2+1/e})^{1/\beta}\right\}, & \beta \in (1/2, 1], \end{cases}$$

Remark 1.1. **(1) Growth of $\psi_{\alpha,\beta}$.** For $\beta \in (1/2, 1]$ and $\alpha \in \mathbb{R}$, the function $\psi_{\alpha,\beta}(d)$ is constant when $d = 1$. For $d \geqslant 2$ and any $p > 0$, $\varepsilon > 0$, there exists d_0 such that $\log_+^p(d) \leqslant \psi_{\alpha,\beta}(d) \leqslant d^{\varepsilon}$ for all $d \geqslant d_0$. Thus, $\psi_{\alpha,\beta}$ grows sub-polynomially: faster than any fixed power of $\log(d)$, but slower than any fixed power of d. When $\beta = 1/2$, $\psi_{\alpha,\beta}(d) = \log_+(d) d^{k_1}$ grows polynomially with d.
(2) **Growth of $\zeta_{\alpha,\beta}$.** Fix $K \in \mathbb{N}$. If $\beta \in (1/2, 1]$, then $\zeta_{\alpha,\beta}(d, K)$ has the same d–dependence as $\psi_{\alpha,\beta}(d)$ up to a change in the constant multiplying $\log_+^{1/(2\beta)}(d)$ in the exponent. When $\beta = 1/2$, $\zeta_{\alpha,\beta}$ increases polynomially with d.

2 Main Results

The following theorem provides a quantitative Gaussian approximation for a d-dimensional random vector $\boldsymbol{F}$ with components in $\mathbb{D}^{1,2}$. For clarity of exposition, we state the results in this section under the assumption that $\mathrm{Cov}(\boldsymbol{F}) = \boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is the (invertible) covariance matrix of the Gaussian vector $\boldsymbol{Z}$. This assumption is made purely for notational convenience: all bounds remain valid for a general invertible matrix $\boldsymbol{\Sigma}$, in which case an additional term appears that explicitly quantifies the discrepancy between $\mathrm{Cov}(\boldsymbol{F})$ and $\boldsymbol{\Sigma}$. The general analogue is available as Corollary 4.2.

Theorem 2.1. *Fix $d \geqslant 1$. Let $\boldsymbol{F} = (F_1, \dots, F_d)$ be a centred random vector in $\mathbb{R}^d$, where $F_i = \sum_{p=1}^{\infty} I_p(f_{i,p})$, for $f_{i,p} \in \mathfrak{H}^{\odot p}$, $i \in \{1, \dots d\}$, such that $F_i \in \mathbb{D}^{1,2}$. Assume $\boldsymbol{\Sigma} := \mathrm{Cov}(\boldsymbol{F})$ is invertible, and let $\boldsymbol{Z} \sim \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{\Sigma})$, and let $\boldsymbol{\Lambda} := \mathrm{Corr}(\boldsymbol{F})$. Assume for $\gamma, \alpha \in \mathbb{R}$ and $\beta \in [1/2, 1]$, that*

$$\|f_{j,p} \otimes_r f_{k,q}\|_{\mathfrak{H}^{\otimes(p+q-2r)}} \leqslant \frac{\gamma e^{\alpha p} e^{\alpha q}}{(p!q!)^{\beta}}, \qquad \text{for all } j, k \in \{1, \dots, d\}, \tag{2.1}$$

and all $p, q, r \in \mathbb{N}$, where $(p, q) \neq (1, 1)$, and $1 \leqslant r \leqslant p \wedge q - \mathbb{1}_{\{p=q\}}$. Additionally, if $\beta = 1/2$, for (2.2) assume that $\alpha < \alpha_0 \approx -2.846$, and for (2.3) that $\alpha < \log(1/2) - e^{1/(2e)} \approx -1.895$. Then there exists a finite constant $C_{\boldsymbol{\theta}} > 0$, depending on multivariate parameter $\boldsymbol{\theta} = (\alpha, \beta)$ such that

$$d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C_{\boldsymbol{\theta}} \psi_{\alpha,\beta}(d) \gamma \log_+(\gamma) \frac{\log_+(\sigma_*(\boldsymbol{\Lambda}))}{\sigma_*(\boldsymbol{\Lambda})}, \tag{2.2}$$

$$d_{\mathscr{C}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C_{\boldsymbol{\theta}} d^{65/24} \gamma \frac{1}{\sigma_*(\boldsymbol{\Lambda})^{3/2}}, \quad \text{and} \quad d_{\mathcal{W}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C_{\boldsymbol{\theta}} d^{3/2} \gamma \frac{\sigma^*(\boldsymbol{\Sigma})}{\sigma_*(\boldsymbol{\Sigma})}. \tag{2.3}$$

Remark 2.2. We draw some observations about Theorem 2.1:

(1) **Dimensional dependence.** The three bounds in Theorem 1.3 exhibit distinct behaviours with respect to the dimension d. The hyper-rectangular distance bound (2.2) is the only one whose rate depends on the regularity parameters α and β (see (1.4)). When $\beta \in (1/2, 1]$, the associated rate $\psi_{\alpha,\beta}(d)$ grows sub-polynomially in d, whereas at the boundary value $\beta = 1/2$ and for sufficiently negative $\alpha < \alpha_0$, it becomes polynomial of order d^{k_1} with $k_1 \leqslant 0.368$. Hence, higher regularity of the vector $\boldsymbol{F}$ directly improves the dimension dependence of the hyper-rectangular bound. This sensitivity is absent from the convex and 1-Wasserstein distances, whose bounds exhibit fixed polynomial dependence of orders $d^{65/24}$ and $d^{3/2}$, respectively. Overall, whenever applicable, the hyper-rectangular distance yields the sharpest dimension dependence among the three.

(2) **Sample size dependence via γ.** The parameter γ in the contraction bound (2.1) captures all dependence on the sample size n in the CLT setting. For instance, in the case of univariate independent random variables, the resulting bounds obtain the classical Berry–Esseen rates as follows: for the hyper-rectangular distance, $n^{-1/2} \log_+(n)$, which is optimal up to logarithmic factors, and for convex and 1-Wasserstein distance, the optimal rate $n^{-1/2}$. This reflects the expected trade-off, that is, the improved dimensional dependence comes at the cost of weaker n-dependence in the hyper-rectangular distance.

(3) **Role of σ_*.** The bounds in Theorem 2.1 depend on $\sigma_*(\boldsymbol{M})^{-1}$, for $\boldsymbol{M} \in \{\boldsymbol{\Lambda}, \boldsymbol{\Sigma}\}$, the smallest eigenvalue of $\boldsymbol{M}$. If $\sigma_*(\boldsymbol{M})$ is bounded away from 0 by an absolute constant, $\sigma_*(\boldsymbol{M})^{-1}$ reduces to a universal constant. The 1-Wasserstein bound in (2.3) also involves the maximal eigenvalue $\sigma^*(\boldsymbol{\Sigma})$. Consequently, the bound may still deteriorate with the dimension through the overall scale of $\boldsymbol{\Sigma}$, and controlling non-singularity alone does not prevent dimension-dependent constants.

(4) **Contraction assumption.** The explicit bounds (2.2)–(2.3) rely on the kernel contraction assumption (2.1), which yields dimension-dependent rates. More generally, Theorem 4.1 (stated below) shows that $d_{\mathcal{R}}(\boldsymbol{F},\boldsymbol{Z})$, $d_{\mathscr{C}}(\boldsymbol{F},\boldsymbol{Z})$ and $d_{\mathcal{W}}(\boldsymbol{F},\boldsymbol{Z})$ are finite provided only that the infinite sums $\Phi(d)$ and $\Psi(d)$ (defined in (4.11) and (4.12)) are finite. The contraction condition is a sufficient (but not necessary) way to ensure these sums converge uniformly in d. This is stated in our general results, Theorem 4.1 and its Corollary 4.2 right below it.

2.1 Quantitative Multivariate Breuer–Major Theorem

The seminal work of Breuer and Major [9] established a qualitative CLT for functionals $\Phi : \mathbb{R} \to \mathbb{R}$ of stationary Gaussian sequences, showing that for centred sequences $(G_k)_{k\in\mathbb{Z}}$ with correlation function ρ, $n^{-1/2}\sum_{k=1}^{n}\Phi(G_k) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2)$, as $n \to \infty$, under the summability assumption $\sum_{k\in\mathbb{Z}}|\rho(k)|^m < \infty$, where m denotes the Hermite rank of Φ. This result was subsequently extended to multivariate inputs by Arcones [1], allowing functionals $\Phi : \mathbb{R}^K \to \mathbb{R}$, and further to vector-valued outputs $\Phi : \mathbb{R}^K \to \mathbb{R}^d$ in the continuous setting by Nualart and Tiva [48].

Quantitative counterparts of the Breuer–Major theorem are now well understood in several settings: (i) scalar-valued functionals with $K = 1$ [43, 49], (ii) scalar-valued functionals with general $K \geqslant 1$ [41], although that result is not explicit in K or d, and (iii) vector-valued functionals $\Phi : \mathbb{R} \to \mathbb{R}^d$ [40, 44]. To the best of our knowledge, no quantitative bounds are currently available for general mappings $\Phi : \mathbb{R}^K \to \mathbb{R}^d$ with explicit dependence on the input dimension K, output dimension d, and sample size n, despite the relevance of such results for high-dimensional inference, where quantitative control of Gaussian approximation errors is essential.

The results in this section provide non-asymptotic, dimension-explicit bounds for the multivariate Breuer–Major theorem, tracking the dependence on the sample size n, the input dimension K, and the output dimension d. Throughout, let $(\boldsymbol{G}_k)_{k\in\mathbb{Z}}$ denote a centred, stationary Gaussian sequence in $\mathbb{R}^K$ with $\boldsymbol{G}_k \sim \mathcal{N}_K(\boldsymbol{0},\boldsymbol{I}_K)$ for all $k \in \mathbb{Z}$. Consider non-affine functions $\Phi = (\varphi_1,\ldots,\varphi_d) : \mathbb{R}^K \to \mathbb{R}^d$, where each component $\varphi_i \in L^2(\gamma^K,\mathbb{R}^K)$ admits the Hermite expansion $\varphi_i(\boldsymbol{x}) = \sum_{i=m_i}^{\infty}\varphi_{i,q}(\boldsymbol{x})$, with Hermite rank $m_i \geqslant 2$ and terms $\varphi_{i,q}(\boldsymbol{x})$ stated explicitly in (4.2). We write $m := \min_{1\leqslant i\leqslant d} m_i$. The normalised partial sums are defined by

$$\boldsymbol{S}_n = n^{-1/2}\sum_{k=1}^{n}\Phi(\boldsymbol{G}_k), \tag{2.4}$$

and the results below provides explicit quantitative Gaussian approximation bounds for the law of $\boldsymbol{S}_n$, uniform in n, K, and d.

We next state Theorem 2.3, a general Breuer–Major result formulated in terms of the covariance structure of the underlying Gaussian sequence. The theorem provides non-asymptotic bounds that depend on the finiteness of the quantities $\Delta_1(n,d,K)$ and $\Delta_2(n,d,K)$ defined therein. Corollary 2.4 sharpens this result by imposing additional structural assumptions (2.6) together with a polynomial decay condition on the covariances: $|\rho^{(i,j)}(k)| \leqslant c|k|^{-a}$ for some $a > 0$. This yields fully explicit bounds, in the terms of the dimensions d, K, and the sample size n.

Theorem 2.3. *Let $K, d \in \mathbb{N}$, and consider the stationary Gaussian sequence $(\boldsymbol{G}_k)_{k\in\mathbb{Z}} \subset \mathbb{R}^K$, the function $\Phi = (\varphi_1, \dots, \varphi_d) : \mathbb{R}^K \to \mathbb{R}^d$, and the normalised partial sums $\boldsymbol{S}_n$ for $n \in \mathbb{N}$, as in (2.4). Define covariance function $\rho^{(i,j)}(\ell) := \mathbb{E}[\boldsymbol{G}_1^{(i)}\boldsymbol{G}_{1+\ell}^{(j)}]$, where $i, j \in \{1, \dots, K\}$, $\ell \in \mathbb{Z}$, and the maximal covariance $\overline{\rho}(\ell) := \max_{1\leqslant i,j\leqslant K} |\rho^{(i,j)}(\ell)|$, for all $\ell \in \mathbb{Z}$. Assume the summability condition $\sum_{\ell\in\mathbb{Z}} \overline{\rho}(\ell)^m < \infty$, that $\mathbb{E}[\Phi(\boldsymbol{G}_k)] = 0$ for all $k \in \mathbb{Z}$, and that $\boldsymbol{\Sigma}_n := \mathrm{Cov}(\boldsymbol{S}_n)$ is invertible, and $\boldsymbol{\Lambda}_n := \mathrm{Corr}(\boldsymbol{S}_n)$. Furthermore, let $\boldsymbol{Z}_n \sim \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{\Sigma}_n)$. Define, for $p, q \geqslant m$ and $r \in \{1, \dots, p \wedge q - \mathbb{1}_{\{p=q\}}\}$,*

$$\gamma_{n,p,q,r} := \sqrt{\frac{2}{n}\sum_{k\in\mathbb{Z}} |\overline{\rho}(k)|^m \sum_{|t|<n} |\overline{\rho}(t)|^r \sum_{|s|<n} |\overline{\rho}(s)|^{p\wedge q-r}}. \tag{2.5}$$

Also set, for $p \in \mathbb{N}$, $\mathfrak{a}_p(K) := p(p!)^{-1/2}(4\sqrt{eK})^p \max_{1\leqslant j\leqslant d} \|\varphi_{j,p}(\boldsymbol{G}_1)\|_{L^2}$. Finally, define

$$\Delta_1(n, d, K) := \sum_{p,q\geqslant m} \mathfrak{a}_p(K)\mathfrak{a}_q(K) \log^{(p+q)/2-1}\left(2d^2 + e^{(p+q)/2-2}\right) \sum_{r=1}^{p\wedge q-\mathbb{1}_{\{p=q\}}} \gamma_{n,p,q,r}, \quad \text{and}$$

$$\Delta_2(n, d, K) := \sum_{p,q\geqslant m} \frac{\sqrt{p}(2\sqrt{K})^{p+q}}{\sqrt{q}} \max_{1\leqslant i,j\leqslant d} \|\varphi_{j,p}(\boldsymbol{G}_1)\|_{L^2} \|\varphi_{i,q}(\boldsymbol{G}_1)\|_{L^2} \sum_{r=1}^{p\wedge q-\mathbb{1}_{\{p=q\}}} \gamma_{n,p,q,r}.$$

Then there exists a finite constant $C > 0$ (independent of n, d, K), such that

$$d_{\mathcal{R}}(\boldsymbol{S}_n, \boldsymbol{Z}_n) \leqslant C \log_+(d)\Delta_1(n, d, K) \log_+\left(\Delta_1(n, d, K)\right) \frac{\log_+(\sigma_*(\boldsymbol{\Lambda}_n))}{\sigma_*(\boldsymbol{\Lambda}_n)},$$

$$d_{\mathscr{C}}(\boldsymbol{S}_n, \boldsymbol{Z}_n) \leqslant 402 d^{65/24} \Delta_2(n, d, K) \frac{1}{\sigma_*(\boldsymbol{\Lambda}_n)^{3/2}}, \quad \text{and}$$

$$d_{\mathcal{W}}(\boldsymbol{S}_n, \boldsymbol{Z}_n) \leqslant d^{3/2} \Delta_2(n, d, K) \frac{\sigma^*(\boldsymbol{\Sigma}_n)^{1/2}}{\sigma_*(\boldsymbol{\Sigma}_n)}.$$

Theorem 2.3 is a general result and note that the quantity is not necessarily finite. Indeed, there is no guarantee that the sums given in $\Delta_1(n, d, K)$ and $\Delta_2(n, d, K)$ are finite. However, given the assumption (2.6), then $\Delta_1(n, d, K)$ and $\Delta_2(n, d, K)$ are finite and together with an additional control on the covariance function ρ, we obtain an explicit result presented in Corollary 2.4, below.

Corollary 2.4. *In the setting as in Theorem 2.3, let $\alpha \in \mathbb{R}$, $c > 0$, and $\beta \in [1/2, 1]$. If $\beta = 1/2$, for hyper-rectangular distance bound assume that $\alpha < \alpha_0 - \log_+(K)/2$ and for the other two distances that $\alpha < -1/(2e) - \ln(2\sqrt{K})$. Assume that*

$$\mathbb{E}[\varphi_{i,q}(\boldsymbol{G}_1)^2]^{1/2} \leqslant \frac{ce^{\alpha q}}{(q!)^{\beta-1/2}}, \quad \text{for } i \in \{1, \dots, d\}, \text{ and all } q \in \mathbb{N}. \tag{2.6}$$

Assume additionally that $|\rho^{(i,j)}(k)| \leqslant c_1|k|^a$, for $a < -1/m$, $k \in \mathbb{Z}\backslash\{0\}$, and $c_1 > 0$, and denote the multivariate parameter $\boldsymbol{\theta} = (c, c_1, \alpha, \beta)$. For constants k_3, c_3, c_4, depending on $\boldsymbol{\theta}$, define

$$\Gamma_{a,m}(n) := \begin{cases} n^{-1/2}, & a < -1, \\ n^{-1/2}\log_+(n), & a = -1, \\ n^{a/2}, & a \in (-1, -1/(m-1)], \\ n^{(am+1)/2}, & a \in (-1/(m-1), -1/m), \end{cases}$$

$$\upsilon_1(K) = 4\sqrt{K}e^{1+\alpha+3/(4e)+k_3K^{\frac{1}{2\beta}}}, \text{ and } \upsilon_2(K) := \begin{cases} \exp\left(c_3K^{\frac{1}{2\beta-1}}\right), & \beta \in (1/2,1], \\ (1-c_4\sqrt{K})^{-2}, & \beta = 1/2. \end{cases} \tag{2.7}$$

Recall, $\zeta_{\alpha,\beta}(d,K)$, *from* (1.5)*. Then there exists a positive constant* $C_{\boldsymbol{\theta}}$ *depending only on* $\boldsymbol{\theta}$*, such that*

$$d_{\mathcal{R}}(\boldsymbol{S}_n,\boldsymbol{Z}_n) \leqslant C_{\boldsymbol{\theta}}\zeta(d,K)\upsilon_1(K)^{1/(\beta-1/2)}\exp\left(\frac{3}{2}\upsilon_1(K)^{1/(\beta-1/2)}\right)\Gamma_{a,m}(n)\log_+(n)$$
$$\times\left(\log_+\log_+(n)\mathbb{1}_{\{a=-1\}}+\mathbb{1}_{\{a\neq-1\}}\right)\frac{\log_+(\sigma_*(\boldsymbol{\Lambda}_n))}{\sigma_*(\boldsymbol{\Lambda}_n)},$$
$$d_{\mathscr{C}}(\boldsymbol{S}_n,\boldsymbol{Z}_n) \leqslant C_{\boldsymbol{\theta}}d^{65/24}\upsilon_2(K)\Gamma_{a,m}(n)\frac{1}{\sigma_*(\boldsymbol{\Lambda}_n)^{3/2}}, \text{ and} \tag{2.8}$$
$$d_{\mathcal{W}}(\boldsymbol{S}_n,\boldsymbol{Z}_n) \leqslant C_{\boldsymbol{\theta}}d^{3/2}\upsilon_2(K)\Gamma_{a,m}(n)\frac{\sigma^*(\boldsymbol{\Sigma}_n)^{1/2}}{\sigma_*(\boldsymbol{\Sigma}_n)}. \tag{2.9}$$

In the boundary case $\beta = 1/2$, Assumption (2.6) imposes the dimension-dependent restriction it holds that $\alpha < \alpha_0 - \log_+(K)/2$, which is decreasing in K. Thus, larger input dimension K forces a smaller admissible α, i.e. tighter growth control of the Hermite components $\varphi_{i,q}$ for all $i \in \{1,\ldots,d\}$, and all $q \in \mathbb{N}$ (equivalently, stronger regularity assumption on Φ).

At the same time, in the hyper-rectangular distance the quantitative bound deteriorates with K through the explicit factor $\zeta(d,K)\,\upsilon_1(K)^{1/(\beta-1/2)}\exp\{(3/2)\upsilon_1(K)^{1/(\beta-1/2)}\}$. Consequently, increasing K allows Φ to depend on a richer Gaussian input but requires stronger regularity and creates a larger constant, whereas smaller K yields a sharper bound and permits weaker regularity. Recall that $\zeta_{\alpha,\beta}(d,K)$, described in (1.5) grows sub-polynomially in d for $\beta \in (1/2,1]$, and polynomially in the boundary case $\beta = 1/2$. The order of the polynomial growth $k_2 \approx 0.768$, which is less than polynomial growths order 65/24 for the convex distance, and for 3/2 for 1-Wasserstein distance in (1.18).

As a benchmark, consider the $K = 1$ and let $(G_k)_{k\in\mathbb{Z}}$ be fractional Gaussian noise, $G_k = B^H_{k+1} - B^H_k$, associated with fractional Brownian motion of Hurst index $H \in (0,1)$. Then $\rho(k) = \mathrm{Cov}(G_0,G_k) \sim H(2H-1)k^{2H-2}$ as $k \to \infty$, so Corollary 2.4 applies with $a = 2H-2$. Specialising (2.8)–(2.9) to this case yields the same n-rates as in [41, Ex. 2.6] for the convex distance and as in [40, Thm. 4.1] for the 1–Wasserstein distance, up to an additional factor $\log_+(n)$ at the boundary $H = 1/2$ (i.e. for standard Brownian increments).

Remark 2.5. Assume $K = 1$. Every square integrable $\varphi : \mathbb{R} \to \mathbb{R}$ admits a Hermite expansion $\varphi(x) = \sum_{q=0}^{\infty} a_qH_q(x)/(q!)$, where H_q denotes q-th Hermite polynomial and a_q, the so-called Hermite coefficients. For a small discussion on Hermite polynomials and expansion, see Section 4.1. For $K = 1$, Corollary 2.4 recovers that the condition (2.6) corresponds to the usual assumption on the Hermite coefficients, as in work of [3]. Indeed, when $K = 1$, then (2.6) becomes

$$\mathbb{E}[\varphi_{i,q}(\boldsymbol{G}_1)^2]^{1/2} \leqslant \frac{e^{\alpha q}}{(q!)^{\beta-1/2}}, \qquad \text{where } \varphi_{i,q}(x) = a_{i,q}H_q(x), \text{ for } x \in \mathbb{R}.$$

Therefore from (4.46) it follows that $\mathbb{E}[\varphi_{i,q}(\boldsymbol{G}_1)^2]^{1/2} = \sqrt{q!}|a_q|$ which then yields that the (2.6) for $K = 1$ is equivalent to the conditions on the Hermite coefficients $|a_q| \leqslant e^{\alpha q}/(q!)^{\beta}$, as in [3, Thm 2.1].

3 Related Results and Examples

In this section we provide a finite analogue of Theorem 2.1 and as its application we find a rate of convergence of the CLT component in parameter estimation of VAR(p).

3.1 Finite Chaos Expansion Case

The next result provides a dimension-dependent Gaussian approximation in the special case where each component of $\boldsymbol{F}$ admits a *finite* Wiener chaos expansion. Unlike the infinite-sum case treated in Theorem 2.1, the result below does not follow from the general Theorem 4.1. Namely, proving the statement directly allows us to exploit the finiteness of the chaos structure and thereby obtain strictly sharper dependence on the dimension d for the hyper-rectangular distance. As before, we present the result under the simplifying assumption $\mathrm{Cov}(\boldsymbol{F}) = \boldsymbol{\Sigma}$, but as in Theorem 2.1, the conclusion remains valid for any invertible target covariance matrix. In that general case, an additional correction term appears in the bound, capturing the discrepancy between the two covariances.

Theorem 3.1. *Fix integer $d \geqslant 1$ as well as $q_1, \ldots, q_d \geqslant 1$. Let $\boldsymbol{F} = (F_1, \ldots, F_d)$ be a d-dimensional random vector with values in $\mathbb{R}^d$, where $F_i = \sum_{k=1}^{q_i} I_k(f_{i,k})$, for $f_{i,k} \in \mathfrak{H}^{\odot k}$, $k \in \{1, \ldots, q_i\}$ and $i \in \{1, \ldots d\}$ such that $F_i \in \mathbb{D}^{1,2}$. Suppose also $\mathrm{Cov}(\boldsymbol{F}) = \boldsymbol{\Sigma}$ is invertible, and let $\boldsymbol{Z} \sim (\boldsymbol{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Lambda} := \mathrm{Corr}(\boldsymbol{Z})$. Finally, define $q := \max_{1 \leqslant i \leqslant d} q_i$ and let*

$$\Delta_{\boldsymbol{F}} := \max\Big\{ \|f_{i,k} \otimes_r f_{j,\ell}\|_{\mathfrak{H}^{\otimes(k+\ell-2r)}} : 1 \leqslant r \leqslant k \wedge \ell - \mathbb{1}_{\{k=\ell\}},\ 2 \leqslant k \leqslant q_i,\ 2 \leqslant \ell \leqslant q_j,\ 1 \leqslant i, j \leqslant d \Big\}$$

Then there exists a positive constant C_q, depending only on q, such that

$$\begin{aligned} &d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C_q \log^q(d) \log_+ \log(d) \Delta_{\boldsymbol{F}} \log_+(\Delta_{\boldsymbol{F}}) \frac{\log_+(\sigma_*(\boldsymbol{\Lambda}))}{\sigma_*(\boldsymbol{\Lambda})}, \\ &d_{\mathscr{C}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C_q d^{65/24} \Delta_{\boldsymbol{F}} \frac{1}{\sigma_*(\boldsymbol{\Lambda})^{3/2}} \quad \text{and} \quad d_{\mathcal{W}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C_q d^{3/2} \Delta_{\boldsymbol{F}} \frac{\sigma^*(\boldsymbol{\Sigma})^{1/2}}{\sigma_*(\boldsymbol{\Sigma})}. \end{aligned} \tag{3.1}$$

Remark 3.2. (1) **Dimensional dependence.** The d-dependence in the hyper-rectangular distance bound is improved, and simply given as $\log^q(d)$ compared to the infinite-chaos-expansion case. On the other hand, the dimensional dependence does not change for convex and 1-Wasserstein distance.

(2) **Comparison with Fang–Koike.** In the case when all components of $\boldsymbol{F}$ have the same chaos order, i.e. $q_i = q$ for all $i = 1, \ldots, d$, Theorem 3.1 recovers the bound from [25, Cor. 1.2] with an extra logarithmic factor $\log_+ \log_+(d)$, which we believe should also be present in [25, Cor. 1.2]. Theorem 3.1 thus extends [25, Cor. 1.2] in two directions: (i) each component F_i may belong to a different fixed chaos order, and (ii) each component may admit a finite chaos expansion.

There is a natural connection between contractions and fourth cumulants. For readers preferring cumulant-based expressions, $\Delta_{\boldsymbol{F}}$ in (3.1) admits the alternative representation. Let $\kappa_4(\cdot)$ denote the fourth cumulant defined as $\kappa_4(X) := \mathbb{E}[X^4] - 3(\mathbb{E}[X^2])^2$, for a centred real-valued random variable X. The proof of the following statement is provided in Section 4.3.

Corollary 3.3. *Assume the same setting as in Theorem 3.1, however define*

$$\Delta_{\boldsymbol{F}} = \max_{1\leqslant i\leqslant d}\max_{1\leqslant k\leqslant q_i}(\kappa_4(I_k(f_{i,k})))^{1/2} + \max_{1\leqslant i,j\leqslant d}\max_{\substack{1\leqslant k\leqslant q_i\\ 1\leqslant \ell\leqslant q_j}}\left(\mathbb{1}_{\{k\neq\ell\}}\mathbb{E}[I_k(f_{i,k})^2]\kappa_4(I_\ell(f_{j,\ell}))^{1/2}\right)^{1/2}.$$

Then, there exists a positive constant C_q, depending only on q, such that (3.1) holds.

3.2 VAR(p) Estimation

Vector autoregressive (VAR) models form one of the central tools for analysing multivariate time series in econometrics, finance, macroeconomics, and many applied sciences [27, 34]. They provide a flexible linear framework for capturing dynamic interactions among multiple variables and underpin widely used procedures such as Impulse Response analysis, forecasting, and Granger-Causality testing (see [34]). Accurate parameter estimation is therefore crucial: without well-estimated coefficients, a VAR cannot faithfully represent the underlying data-generating mechanism, and its forecasts and inference become unreliable. Since inference in VAR models typically relies on the asymptotic normality of the parameter estimators, establishing quantitative CLT rates is essential for assessing accuracy in finite samples and in high-dimensional regimes. In particular, explicit error bounds are required to construct reliable confidence intervals and hypothesis tests. In this section, we derive explicit convergence rates for the *CLT component* of the OLS estimator (defined formally below).

A stable VAR(p) is a process $(\boldsymbol{y}_k)_{k\in\mathbb{Z}} \subseteq \mathbb{R}^d$ defined by the following recursion

$$\boldsymbol{y}_k = \boldsymbol{A}_1\boldsymbol{y}_{k-1} + \cdots + \boldsymbol{A}_p\boldsymbol{y}_{k-p} + \boldsymbol{u}_k \qquad \text{for } k\in\mathbb{Z}, \text{ and } p\geqslant 1, \tag{3.2}$$

with coefficient matrices $\boldsymbol{A}_i \in \mathbb{R}^{d\times d}$, for $i \in \{1,\dots,p\}$ and an innovation process $(\boldsymbol{u}_k)_{k\in\mathbb{Z}} \subset \mathbb{R}^d$. In this text, we assume that $(\boldsymbol{u}_k)_{k\in\mathbb{Z}}$ is a Gaussian white noise, i.e. $\boldsymbol{u}_k \sim \mathcal{N}_d(\boldsymbol{0},\boldsymbol{\Sigma}_u)$, with $\mathbb{E}[\boldsymbol{u}_k\boldsymbol{u}_k^\top] = \boldsymbol{\Sigma}_u$, $\mathbb{E}[\boldsymbol{u}_k] = 0$ and $\mathbb{E}[\boldsymbol{u}_k\boldsymbol{u}_\ell] = 0$ for $k \neq \ell$. Assume the covariance $\boldsymbol{\Sigma}_u$ to be invertible and denote $\boldsymbol{\Sigma}_y := \mathbb{E}[\boldsymbol{y}_0\boldsymbol{y}_0^\top] \in \mathbb{R}^{d\times d}$. Note that a stable VAR(p) process is also stationary, see [34, Prop. 2.1] and for general theory on VAR(p), see [34] or [27]. Stationarity implies that $\mathrm{Cov}(\boldsymbol{y}_k) = \mathbb{E}[\boldsymbol{y}_k\boldsymbol{y}_k^\top] = \boldsymbol{\Sigma}_y$ for all $k\in\mathbb{Z}$.

Now, fix positive integers T and n, such that $T-n \geqslant p$ (i.e. assuring that we have enough observations). Define $\boldsymbol{y}_t = (y_t^{(1)},\dots,y_t^{(d)})^\top \in \mathbb{R}^d$ to be a d-dimensional time series available for $t\in\{1,\dots,T\}$, generated by a stable VAR(p). In other words (3.2) holds for $t\in\{1,\dots,T\}$. One of the central tasks in statistics is the estimation of the underlying coefficient matrices $\boldsymbol{B} := (\boldsymbol{A}_1,\dots,\boldsymbol{A}_p) \in \mathbb{R}^{d\times dp}$, where the (multivariate) least squares estimator $\widehat{\boldsymbol{B}}$ is the most widely used method in such a multivariate setting.

Denote $\boldsymbol{\beta} := \mathrm{vec}(\boldsymbol{B}) \in \mathbb{R}^{d^2p}$ and $\widehat{\boldsymbol{\beta}} := \mathrm{vec}(\widehat{\boldsymbol{B}}) \in \mathbb{R}^{d^2p}$ to be vector-forms of the matrix of coefficients $\boldsymbol{B}$ and its estimator $\widehat{\boldsymbol{B}}$. Further, for $n\in\mathbb{N}$, denote the stacked p-lagged vectors as $\boldsymbol{\mathcal{Y}}_t := (\boldsymbol{y}_t,\dots,\boldsymbol{y}_{t-p+1})^\top \in \mathbb{R}^{dp}$, for $t = \{1,\dots,n\}$, and a matrix consisting of the p-lagged vectors $\boldsymbol{\mathcal{Y}} := (\boldsymbol{\mathcal{Y}}_1,\dots,\boldsymbol{\mathcal{Y}}_n) \in \mathbb{R}^{dp\times n}$. Let $\boldsymbol{\varpi} := (\boldsymbol{u}_1,\dots,\boldsymbol{u}_n)^\top \in \mathbb{R}^{dn}$, denote the stacked innovations. Note that $\boldsymbol{\varpi}$ has covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varpi}} = \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}_u$ and denote $\boldsymbol{\Sigma}_{\mathcal{Y}} := \mathbb{E}[\boldsymbol{\mathcal{Y}}_0\boldsymbol{\mathcal{Y}}_0^\top] = \mathrm{Cov}(\boldsymbol{\mathcal{Y}}_k)$ to be the covariance matrix of matrix $\boldsymbol{\mathcal{Y}}_k$ for all $k\in\mathbb{Z}$. Then [34, Eq. (3.2.9)] gives a useful representation of $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\beta}}$, i.e. the vector form of the coefficient matrix $\boldsymbol{B}$ and its estimator $\widehat{\boldsymbol{B}}$, as follows

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\left(\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^\top\right)^{-1}\boldsymbol{\mathcal{Y}}\otimes\boldsymbol{I}_d\right)\boldsymbol{\varpi}. \tag{3.3}$$

It is well known that a least squares estimator $\widehat{\boldsymbol{B}}$ is consistent, i.e. $\mathbb{P}\text{-}\lim_{T\to\infty}\widehat{\boldsymbol{B}} = \boldsymbol{B}$, and that it satisfies a CLT. Indeed, [34, Prop. 3.1], yields that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}_{d^2p}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\mathcal{Y}}^{-1} \otimes \boldsymbol{\Sigma}_u\right), \qquad \text{as } n \to \infty, \tag{3.4}$$

where $\boldsymbol{\Sigma}_{\mathcal{Y}} = \mathbb{P}\text{-}\lim_{n\to\infty} n^{-1}\sum_{k=1}^{n} \boldsymbol{\mathcal{Y}}_{p+k}\boldsymbol{\mathcal{Y}}_{p+k}^{\top} = \mathbb{P}\text{-}\lim_{n\to\infty}(\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^{\top}/n)$ exists and is invertible. The existence and invertibility of $\boldsymbol{\Sigma}_{\mathcal{Y}}$ follow from Gaussianity of the innovations, namely, for Gaussian white noise innovation process, VAR(p) is ergodic for second moments, see for example [27, Prop. 10.2] together with [34, Def. 3.1] and discussion just below the definition.

To the best of the author's knowledge, the rate of convergence in (3.4) is currently unknown, except for $p = 1$ and $d = 1$ see [30, 20, 23]. In the following section we show the rate of convergence for the CLT part of (3.4). Multiplying (3.3) by $\sqrt{n}$ and using standard properties of Kronecker products, we obtain

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\left(\frac{1}{n}\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^{\top}\right)^{-1} \otimes \boldsymbol{I}_d\right)\frac{1}{\sqrt{n}}\left(\boldsymbol{\mathcal{Y}} \otimes \boldsymbol{I}_d\right)\boldsymbol{\varpi} = \left(\left(\frac{1}{n}\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^{\top}\right)^{-1} \otimes \boldsymbol{I}_d\right)\boldsymbol{S}_n, \tag{3.5}$$

where $\boldsymbol{S}_n$ is defined as $\boldsymbol{S}_n := n^{-1/2}\left(\boldsymbol{\mathcal{Y}} \otimes \boldsymbol{I}_d\right)\boldsymbol{\varpi}$. Following the classical approach to proving asymptotic normality, we separate in (3.5) the law of large numbers part $(n^{-1}\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^{\top})^{-1}$ from the CLT part of $\boldsymbol{S}_n$. Additionally, [34, Prop. C.2(4)] yields that $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ has the same asymptotic distribution as $(\boldsymbol{\Sigma}_{\mathcal{Y}}^{-1} \otimes \boldsymbol{I}_d)\boldsymbol{S}_n$ and hence, [34, Lem. 3.1], yields that

$$\boldsymbol{S}_n = n^{-1/2}(\boldsymbol{\mathcal{Y}} \otimes \boldsymbol{I}_d)\boldsymbol{\varpi} \xrightarrow{\mathcal{D}} \mathcal{N}_{d^2p}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathcal{Y}} \otimes \boldsymbol{\Sigma}_u), \qquad \text{as} \qquad n \to \infty. \tag{3.6}$$

Using our methodology, specifically Theorem 3.1, we present the rate of convergence of (3.6) in Theorem 3.4 below.

Theorem 3.4. *Assume the stable* VAR(p)*-setting of* (3.2)*. Let* $\boldsymbol{S}_n = n^{-1/2}(\boldsymbol{\mathcal{Y}} \otimes \boldsymbol{I}_d)\boldsymbol{\varpi}$ *as in* (3.5) *and let* $\boldsymbol{Z} \sim \mathcal{N}_{d^2p}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathcal{Y}} \otimes \boldsymbol{\Sigma}_u)$*, and* $\boldsymbol{\Lambda} := \mathrm{Corr}(\boldsymbol{Z})$*. For* $k \in \mathbb{Z}$*, define* $\boldsymbol{G}_k := (\boldsymbol{u}_k, \boldsymbol{\mathcal{Y}}_{k-1})^{\top} \in \mathbb{R}^{d(p+1)}$*, for which* $\boldsymbol{G}_k \sim \mathcal{N}_{d(p+1)}(\mathbf{0}, \boldsymbol{\Sigma}_G)$*, where* $(\boldsymbol{\Sigma}_G)_{i,j} = \mathrm{Cov}(\boldsymbol{G}_k^i, \boldsymbol{G}_k^j)$*. Further, let* $\boldsymbol{\Gamma}_G(\ell) := \mathrm{Cov}(\boldsymbol{G}_k, \boldsymbol{G}_{k-\ell})$ *for* $\ell > 0$ *and* $\boldsymbol{\Gamma}_G(0) := \boldsymbol{\Sigma}_G$*. Then, there exists a universal constant* $C > 0$*, such that*

$$\begin{aligned} d_{\mathcal{R}}(\boldsymbol{S}_n, \boldsymbol{Z}) &\leqslant C\log_+^2(d^2p)\log_+\log_+(d^2p)\frac{\log_+(n)}{\sqrt{n}}\frac{\log_+(\sigma_*(\boldsymbol{\Lambda}))}{\sigma_*(\boldsymbol{\Lambda})}\|\boldsymbol{\Sigma}_u\|_{\max}^{1/2} \\ &\qquad \times\left(\sum_{\ell=0}^{\infty}\|\boldsymbol{\Gamma}_G(\ell)\|_{\max}\right)^{3/2}\log_+\|\boldsymbol{\Sigma}_u\|_{\max}\log_+\left(\sum_{\ell=0}^{\infty}\|\boldsymbol{\Gamma}_G(\ell)\|_{\max}\right), \qquad (3.7) \\ d_{\mathscr{C}}(\boldsymbol{S}_n, \boldsymbol{Z}) &\leqslant C(d^2p)^{65/24}n^{-1/2}\frac{1}{\sigma_*(\boldsymbol{\Lambda})^{3/2}}\|\boldsymbol{\Sigma}_u\|_{\max}^{1/2}\left(\sum_{\ell=0}^{\infty}\|\boldsymbol{\Gamma}_G(\ell)\|_{\max}\right)^{3/2}, \qquad \textit{and} \\ d_{\mathcal{W}}(\boldsymbol{S}_n, \boldsymbol{Z}) &\leqslant C(d^2p)^{3/2}n^{-1/2}\frac{\sigma^*(\boldsymbol{\Sigma}_{\mathcal{Y}} \otimes \boldsymbol{\Sigma}_u)^{1/2}}{\sigma_*(\boldsymbol{\Sigma}_{\mathcal{Y}} \otimes \boldsymbol{\Sigma}_u)}\|\boldsymbol{\Sigma}_u\|_{\max}^{1/2}\left(\sum_{\ell=0}^{\infty}\|\boldsymbol{\Gamma}_G(\ell)\|_{\max}\right)^{3/2}. \end{aligned}$$

Remark 3.5. (1) Recall (3.3). The quantity $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, is a self-normalised statistic, and obtaining quantitative rates for self-normalised statistics in high-dimensional settings, is difficult and largely open. See, for example, the recent contributions for the special cases of

scalar self-normalised dependent sums [28] and high-dimensional maxima of self-normalised coordinates [14]. A full treatment of the self-normalisation problem is beyond the scope of the present paper, as our primary goal here is to demonstrate how our methodology, in particular, Theorem 3.1, can be applied in a high-dimensional time-series setting. Hence, establishing a quantitative rate for the CLT component $\boldsymbol{S}_n$ of (3.7) provides a proof of concept for our general results.
(2) Related work on the error bounds for stable VAR(p) estimators exists in the univariate case, however the high-dimensional case, is currently an open question: In the continuous-time scalar setting, Kim and Park [30] obtain optimal Berry–Esseen bound of order $T^{-1/2}$ for the normalised drift estimator $\sqrt{T}(\widehat{\theta}-\theta)$ in a one-dimensional Ornstein–Uhlenbeck process (the continuous-time analogue of a stable VAR(1) with Gaussian innovations). Douissi et al. [23] derive optimal rates of order $n^{-1/2}$ for suitably normalised parameter estimation for univariate AR(1) process (univariate VAR(1) process). For the high-dimensional continuous-time rates, Ciołek et al. [20] obtain oracle inequalities and error bounds for Dantzig and Lasso estimators of the drift in a multivariate Ornstein–Uhlenbeck process, under sparsity assumptions on the drift matrix, but do not study Berry–Esseen type bounds for the estimator.

Remark 3.6 (VAR(p), $p>1$, presented as VAR(1)). It is quite straightforward and useful to represent a VAR(p) process, $p>1$, as a VAR(1) process. For example, see [34, Sec. 2.1.1 specifically Eq. (2.1.8)] or [27, Sec. 10.1]. Namely, for every $k\in\mathbb{Z}$ let $\boldsymbol{A}\in\mathbb{R}^{dp\times dp}$ and $\boldsymbol{U}_k\in\mathbb{R}^{dp}$ be given as follows

$$\boldsymbol{A}:=\begin{pmatrix}\boldsymbol{A}_1 & \boldsymbol{A}_2 & \dots & \boldsymbol{A}_{p-1} & \boldsymbol{A}_p\\ \boldsymbol{I}_d & \boldsymbol{0} & \dots & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_d & \dots & \boldsymbol{0} & \boldsymbol{0}\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{I}_d & \boldsymbol{0}\end{pmatrix},\quad \boldsymbol{U}_k:=\begin{pmatrix}\boldsymbol{u}_k\\ \boldsymbol{0}\\ \boldsymbol{0}\\ \vdots\\ \boldsymbol{0}\end{pmatrix}.$$

Then, (3.2) can be written equivalently as

$$\boldsymbol{\mathcal{Y}}_k=\boldsymbol{A}\boldsymbol{\mathcal{Y}}_{k-1}+\boldsymbol{U}_k,\tag{3.8}$$

which is exactly a dp-dimensional VAR(1) process $(\boldsymbol{\mathcal{Y}}_k)$. Let $\boldsymbol{J}=(\boldsymbol{I}_d:\boldsymbol{0}:\cdots:\boldsymbol{0})\in\mathbb{R}^{d\times dp}$, then one can go between the two settings as follows: $\boldsymbol{y}_k=\boldsymbol{J}\boldsymbol{\mathcal{Y}}_k$ and $\boldsymbol{\Sigma}_y=\boldsymbol{J}\boldsymbol{\Sigma}_{\mathcal{Y}}\boldsymbol{J}^\top$, where $\boldsymbol{\Sigma}_{\mathcal{Y}}:=\operatorname{Cov}(\mathcal{Y}_k)$.

Remark 3.7. Let $\boldsymbol{A}\in\mathbb{R}^{dp\times dp}$ and $\boldsymbol{\Sigma}_U\in\mathbb{R}^{dp\times dp}$ be as in Remark 3.6, and denote $\boldsymbol{\Sigma}_U:=\operatorname{Cov}(\boldsymbol{U}_k)=\operatorname{Diag}(\boldsymbol{\Sigma}_u,\boldsymbol{0},\dots,\boldsymbol{0})$ for $k\in\mathbb{Z}$. For each integer $\ell\geqslant 0$, Lemma 3.8 yields explicit structure of $\boldsymbol{\Gamma}_G(\ell)$ and specifically,

$$\|\boldsymbol{\Gamma}_G(\ell)\|_{\max}=\max\Big\{\|\boldsymbol{A}^\ell\boldsymbol{\Sigma}_U\boldsymbol{J}^\top\|_{\max},\Big\|\sum_{i=0}^{\infty}\boldsymbol{A}^{\ell+i}\boldsymbol{\Sigma}_U(\boldsymbol{A}^i)^\top\Big\|_{\max}\Big\}.$$

In general, if $\rho(\boldsymbol{A})<1$, then the series $\sum_{\ell=0}^{\infty}\|\boldsymbol{\Gamma}_G(\ell)\|_{\max}$ converges. Indeed, for any r with $\rho(\boldsymbol{A})<r<1$ there exists a finite constant $C>0$ (depending on r and $\boldsymbol{A}$) such that $\|\boldsymbol{\Gamma}_G(\ell)\|_{\max}\leqslant C'r^\ell$ for every $\ell\geqslant 0$, where one may take $C'=C\|\boldsymbol{\Sigma}_U\|_{\mathrm{op}}\|\boldsymbol{J}\|_{\mathrm{op}}\big(1+(1-r^2)^{-1}C\big)$. Consequently, as a geometric series converges for $r<1$, $\sum_{\ell\geqslant 0}\|\boldsymbol{\Gamma}_G(\ell)\|_{\max}\leqslant C'\sum_{\ell\geqslant 0}r^\ell<\infty$ is finite, and so are the bounds in Theorem 3.4.

Lemma 3.8. *Assume the* VAR(p) *setting* (3.2).
(1) If $p = 1$, it holds for all $\ell > 0$, that

$$\boldsymbol{\Gamma}_G(\ell) = \begin{pmatrix} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times d} \\ \boldsymbol{A}_1^{\ell-1}\boldsymbol{\Sigma}_u & \boldsymbol{A}_1^{\ell}\boldsymbol{\Sigma}_y \end{pmatrix}, \text{ where } \boldsymbol{\Sigma}_y = \sum_{i=0}^{\infty} \boldsymbol{A}_1^i \boldsymbol{\Sigma}_u (\boldsymbol{A}_1^i)^\top, \text{ and}$$

$$\boldsymbol{\Gamma}_G(0) = \boldsymbol{\Sigma}_G = \begin{pmatrix} \boldsymbol{\Sigma}_u & \mathbf{0}_{d\times d} \\ \mathbf{0}_{d\times d} & \boldsymbol{\Sigma}_{\mathcal{Y}} \end{pmatrix}.$$

(2) If $p \in \mathbb{N}\backslash\{1\}$, then $\boldsymbol{\Gamma}_G(0) = \boldsymbol{\Sigma}_G \in \mathbb{R}^{d(p+1)\times d(p+1)}$, and for all integers $\ell > 0$,

$$\boldsymbol{\Gamma}_G(\ell) = \begin{pmatrix} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times dp} \\ \boldsymbol{A}^{\ell-1}\boldsymbol{\Sigma}_U \boldsymbol{J}^\top & \boldsymbol{\Gamma}_{\mathcal{Y}}(\ell) \end{pmatrix}, \text{ where}$$

$$\boldsymbol{\Gamma}_{\mathcal{Y}}(\ell) = \sum_{i=0}^{\infty} \boldsymbol{A}^{\ell+i} \boldsymbol{\Sigma}_U (\boldsymbol{A}^i)^\top, \text{ and } \boldsymbol{A}^{\ell-1}\boldsymbol{\Sigma}_U \boldsymbol{J}^\top \in \mathbb{R}^{dp\times d}.$$

The strength of Lemma 3.8 is that it gives a representation of lag-process covariances which is independent of the process $\boldsymbol{y}_k$. It is explicit gives, only in terms of the coefficient matrices $\boldsymbol{A}_i$, and the covariance matrix of the innovation process, $\boldsymbol{\Sigma}_U$.

Example 3.9. Assume $p = 1$ and $\|\boldsymbol{A}_1\|_{\mathrm{op}} < 1$. Let $\boldsymbol{y}_k$ be a stable VAR(1) process with i.i.d. innovations, in other words, $\boldsymbol{y}_k = \boldsymbol{A}_1 \boldsymbol{y}_{k-1} + \boldsymbol{u}_k$, where $\boldsymbol{\Sigma}_u = \boldsymbol{I}_d$. Now, Lemma 3.8, yields that $\|\boldsymbol{\Gamma}_G(\ell)\|_{\max} = \max\{\|\boldsymbol{A}_1^{\ell-1}\|_{\max}, \|\boldsymbol{A}_1^{\ell}\sum_{i=0}^{\infty}\boldsymbol{A}_1^i(\boldsymbol{A}_1^i)^\top\|_{\max}\}$. Hence we wish to give useful bounds for these quantities. Note that for $\ell > 0$, $\|\boldsymbol{A}_1^{\ell-1}\|_{\max} \leqslant \|\boldsymbol{A}_1^{\ell-1}\|_{\mathrm{op}}$, where we used sub-multiplicativity of the operator norm. Furthermore, using the sub-multiplicativity and -additivity of the operator norm, and the fact $\|\boldsymbol{B}\|_{\max} \leqslant \|\boldsymbol{B}\|_{\mathrm{op}}$ for all $\boldsymbol{B} \in \mathbb{R}^{d\times d}$,

$$\|\boldsymbol{A}_1^{\ell}\sum_{i=0}^{\infty}\boldsymbol{A}_1^i\boldsymbol{\Sigma}_u(\boldsymbol{A}_1^i)^\top\|_{\max} \leqslant \|\boldsymbol{A}_1\|_{\mathrm{op}}^{\ell}\sum_{i=0}^{\infty}\|\boldsymbol{A}_1\|_{\mathrm{op}}^{2i} = \|\boldsymbol{A}_1\|_{\mathrm{op}}^{\ell}\left(1 - \|\boldsymbol{A}_1\|_{\mathrm{op}}^2\right)^{-1}.$$

As $\|A_1\|_{\mathrm{op}} < 1$, then $\|\boldsymbol{\Gamma}_G(\ell)\|_{\max} \leqslant \|\boldsymbol{A}_1\|_{\mathrm{op}}^{\ell-1}$ when $\|A_1\|_{\mathrm{op}} < (\sqrt{5}-1)/2$, and $\|\boldsymbol{\Gamma}_G(\ell)\|_{\max} \leqslant \|\boldsymbol{A}_1\|_{\mathrm{op}}^{\ell}\left(1 - \|\boldsymbol{A}_1\|_{\mathrm{op}}^2\right)^{-1}$ otherwise, as $\gamma < 1 - \gamma^2$ when $\gamma < (\sqrt{5}-1)/2$. Next, we calculate

$$\sum_{\ell=0}^{\infty}\|\boldsymbol{\Gamma}_G(\ell)\|_{\max} \leqslant \sum_{\ell=0}^{\infty}\|\boldsymbol{A}_1\|_{\mathrm{op}}^{\ell-1} \leqslant \frac{1}{\|\boldsymbol{A}_1\|_{\mathrm{op}} - \|\boldsymbol{A}_1\|_{\mathrm{op}}^2}, \text{ and}$$

$$\sum_{\ell=0}^{\infty}\|\boldsymbol{\Gamma}_G(\ell)\|_{\max} \leqslant \sum_{\ell=0}^{\infty}\|\boldsymbol{A}_1\|_{\mathrm{op}}^{\ell}\left(1 - \|\boldsymbol{A}_1\|_{\mathrm{op}}^2\right)^{-1} \leqslant \frac{1}{(1 - \|\boldsymbol{A}_1\|_{\mathrm{op}})(1 + \|\boldsymbol{A}_1\|_{\mathrm{op}}^2)},$$

and denote

$$\mathfrak{A}(\boldsymbol{A}_1) := \begin{cases} (\|\boldsymbol{A}_1\|_{\mathrm{op}} - \|\boldsymbol{A}_1\|_{\mathrm{op}}^2)^{-1}, & \text{if } \|A_1\|_{\mathrm{op}} < (\sqrt{5}-1)/2, \\ \{(1 - \|\boldsymbol{A}_1\|_{\mathrm{op}})(1 + \|\boldsymbol{A}_1\|_{\mathrm{op}}^2)\}^{-1}, & \text{otherwise.} \end{cases}$$

Plugging this into the bound (3.7), yields that

$$d_{\mathcal{R}}(\boldsymbol{S}_n, \boldsymbol{Z}) \leqslant C\log_+^2(d)\log_+\log_+(d)\frac{\log_+(n)}{\sqrt{n}}\frac{\log_+(\sigma_*(\boldsymbol{\Lambda}))}{\sigma_*(\boldsymbol{\Lambda})}\mathfrak{A}(\boldsymbol{A}_1)^{3/2}\log_+(\mathfrak{A}(\boldsymbol{A}_1)),$$

$$d_{\mathscr{C}}(\boldsymbol{S}_n, \boldsymbol{Z}) \leqslant C(d)^{65/24}n^{-1/2}\frac{1}{\sigma_*(\boldsymbol{\Lambda})^{3/2}}\mathfrak{A}(\boldsymbol{A}_1)^{3/2}, \text{ and}$$

$$d_{\mathcal{W}}(\boldsymbol{S}_n, \boldsymbol{Z}) \leqslant C(d^2)^{3/2}n^{-1/2}\frac{\sigma^*(\boldsymbol{\Sigma}_{\mathcal{Y}})^{1/2}}{\sigma_*(\boldsymbol{\Sigma}_{\mathcal{Y}})}\mathfrak{A}(\boldsymbol{A}_1)^{3/2}.$$

Example 3.9 is restricted to the VAR(1) case ($p = 1$) because the condition $\|\boldsymbol{A}_1\|_{\mathrm{op}} < 1$ provides a convenient sufficient criterion for stability, even though the standard condition only requires the spectral radius of $\boldsymbol{A}_1$ to be strictly less than 1. Moreover, as noted in [4, Sec. 2.1], such an operator-norm condition is restrictive and does not generalise in a natural way beyond VAR(1). Nevertheless, the example illustrates how the explicit structure of $\boldsymbol{\Gamma}_G$ in Lemma 3.8 leads to more transparent bounds than in the general case.

4 Background and Technical Results

This section collects the technical arguments used in the proofs of the results from Sections 2 and 3. In Subsection 4.1, we recall the Malliavin calculus and Stein–kernel tools needed in the proofs (building on Section 1.2). Subsection 4.2 summarises the proof strategy. In Subsection 4.3, we establish the general framework given by Theorem 4.1 and Corollary 4.2. Theorem 2.1 is then obtained as a specialisation. The remaining subsections contain the proofs of the remainder of the results: Theorems 2.1 and 3.1 in Subsections 4.4 and 4.6, the quantitative Breuer–Major results in Subsection 2.1, and the VAR(p) estimation results in Subsection 4.7.

4.1 Background

Recall from Section 1.2 that, $\log_+(x) = |\log(x)| \vee 1$ for $x \in (0, \infty)$. Note that for this function it holds that for all $x, y > 0$

$$\begin{aligned}\log_+(xy) &= |\log(xy)| \vee 1 = |\log(x) + \log(y)| \vee 1 \leqslant (|\log(x)| + |\log(y)|) \vee 1 \\ &\leqslant \log_+(x) + \log_+(y) \leqslant 2\log_+(x)\log_+(y),\end{aligned} \tag{4.1}$$

as $\log_+(x) \geqslant 1$ for all $x > 0$.

Let γ denote the standard Gaussian measure γ on $\mathbb{R}$, given by $\gamma(\mathrm{d}x) = (2\pi)^{-1}e^{-x^2/2}\mathrm{d}x$. Accordingly, let $L^p(\gamma; \mathbb{R}^d)$ be the set of measurable functions $\Phi : \mathbb{R} \to \mathbb{R}^d$ such that $\mathbb{E}_\gamma[|\Phi(x)|^p] < \infty$. For $q \in \mathbb{N}_0$, the *q-th Hermite polynomial* $H_q : \mathbb{R} \to \mathbb{R}$ is defined by $H_q(x) = (-1)^q e^{x^2/2}\mathrm{d}^q(e^{-x^2/2})/(\mathrm{d}x^q)$. For illustrative purposes, this yields that for $x \in \mathbb{R}$, $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2-1$,... The family $\{(q!)^{-1/2}H_q : q \geqslant 0\}$ is an orthonormal basis of $L^2(\gamma)$. Hence, every $\varphi \in L^2(\gamma)$, admits a *Hermite expansion*

$$\varphi(x) = \sum_{q=0}^{\infty} \frac{a_q}{q!} H_q(x), \quad \text{where } a_q = \mathbb{E}[\varphi(Z)H_q(Z)], \text{ and } Z \sim \mathcal{N}(0, 1),$$

with $\sum_{q\in\mathbb{N}_0} q! a_q^2 < \infty$. The *Hermite rank* of φ is $\mathrm{rank}(\varphi) := \min\{q \geqslant 1 : a_q \neq 0\}$, with the $\mathrm{rank}(\varphi) = \infty$ if φ is constant.

For the multivariate case, let γ^K be the standard Gaussian measure on $\mathbb{R}^K$ and write $\boldsymbol{x} = (x_1, \ldots, x_K)$. For a multi-index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K) \in \mathbb{N}_0^K$, set $|\boldsymbol{\alpha}| = \sum_{j=1}^K \alpha_j$, and $\boldsymbol{\alpha}! = \prod_{j=1}^K \alpha_j!$. Under γ^K, the multivariate Hermite polynomials factorise as

$$H_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \prod_{j=1}^K H_{\alpha_j}(x_j).$$

Then any $\varphi \in L^2(\gamma^K)$ admits the expansion

$$\varphi(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}\in\Lambda} a_{\boldsymbol{\alpha}} H_{\boldsymbol{\alpha}}(\boldsymbol{x}), \quad \text{where } a_{\boldsymbol{\alpha}} = \frac{1}{\boldsymbol{\alpha}!}\mathbb{E}\left[\varphi(\boldsymbol{G})H_{\boldsymbol{\alpha}}(\boldsymbol{G})\right], \text{ for } \boldsymbol{G} \sim \mathcal{N}_K(\boldsymbol{0}, \boldsymbol{I}_K). \tag{4.2}$$

If φ has Hermite rank m, this can be grouped by total degree

$$\varphi(\boldsymbol{x}) = \sum_{q=m}^{\infty} \varphi_q(\boldsymbol{x}), \quad \text{where } \varphi_q(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}\in\Lambda:|\boldsymbol{\alpha}|=q} a_{\boldsymbol{\alpha}} \prod_{j=1}^{K} H_{\alpha_j}(x_j), \quad \boldsymbol{x} \in \mathbb{R}^K. \tag{4.3}$$

The Malliavin–Stein method provides a convenient framework to compare the law of smooth functionals of a Gaussian field with a Gaussian law. We recall the basic objects used throughout the paper; see [39, 45, 46] for background, and [39] for a detailed treatment of the Malliavin–Stein method.

Let $\mathfrak{H}$ be a real separable Hilbert space with inner product $\langle\cdot,\cdot\rangle_{\mathfrak{H}}$ and norm $\|\cdot\|_{\mathfrak{H}}$. Let $\boldsymbol{X} = \{X(h) : h \in \mathfrak{H}\}$ be an isonormal Gaussian process over $\mathfrak{H}$ on $(\Omega, \mathcal{F}, \mathbb{P})$. For $q \in \mathbb{N}_0$, denote by $\mathcal{H}_q$ the Wiener–Itô chaos of order q, and set $\mathcal{P}_q := \bigoplus_{i=0}^{q} \mathcal{H}_i$.

The multiple Wiener–Itô integral $I_q : \mathfrak{H}^{\odot q} \to \mathcal{H}_q$ satisfies the isometry

$$\mathbb{E}[I_p(f)I_q(g)] = \mathbf{1}_{\{p=q\}}\, q!\,\langle f, g\rangle_{\mathfrak{H}^{\otimes q}}, \qquad p, q \geqslant 1, \quad f \in \mathfrak{H}^{\odot p}, \quad g \in \mathfrak{H}^{\odot q}, \tag{4.4}$$

and Hermite polynomials are related to multiple integrals by

$$H_q(X(h)) = I_q(h^{\otimes q}), \qquad h \in \mathfrak{H}, \ \|h\|_{\mathfrak{H}} = 1.$$

Let $\mathcal{S}$ denote the class of smooth cylindrical random variables of the form $F = f(X(h_1), \ldots, X(h_d))$, where $h_i \in \mathfrak{H}$, for all $i \in \{1, \ldots, d\}$, and $f \in \mathcal{C}^\infty(\mathbb{R}^d)$ has partial derivatives of polynomial growth. For $F \in \mathcal{S}$, the Malliavin derivative is defined as $DF = \sum_{i=1}^{d} \partial_i f(X(h_1), \ldots, X(h_d))h_i$. The space $\mathbb{D}^{1,2}$ is the closure of $\mathcal{S}$ under

$$\|F\|_{1,2}^2 := \mathbb{E}[F^2] + \mathbb{E}[\|DF\|_{\mathfrak{H}}^2],$$

and higher-order derivatives $D^kF \in L^2(\Omega; \mathfrak{H}^{\otimes k})$ are defined iteratively, with $\mathbb{D}^{k,2}$ the corresponding completion. The chain rule holds: if $\boldsymbol{F} = (F_1, \ldots, F_d)$ with $F_i \in \mathbb{D}^{1,2}$ and $\phi : \mathbb{R}^d \to \mathbb{R}$ is C^1 with bounded partial derivatives, then $D(\phi(\boldsymbol{F})) = \sum_{i=1}^{d} \partial_i \phi(\boldsymbol{F})\, DF_i$,see [45, Prop. 1.2.3].

The divergence operator δ is the adjoint of D, it acts on elements $u \in L^2(\Omega; \mathfrak{H})$ in its domain $\mathrm{Dom}(\delta)$, characterized by $\mathbb{E}[\langle DF, u\rangle_{\mathfrak{H}}] = \mathbb{E}[F\delta(u)]$, for all $F \in \mathbb{D}^{1,2}$. Moreover, for $f \in \mathfrak{H}^{\odot q}$, $\delta^q(f) = I_q(f)$, and $I_q(f) \in \mathbb{D}^{\infty,2}$ with $DI_q(f) = q\, I_{q-1}(f)$.

Let $F \in L^2(\Omega)$ with its Wiener–Itô chaos expansion (1.2). The Ornstein–Uhlenbeck generator L and its pseudo-inverse L^{-1} are given by

$$LF = -\sum_{q=1}^{\infty} qI_q(f_q), \quad \text{and} \quad L^{-1}F = -\sum_{q=1}^{\infty} \frac{1}{q} I_q(f_q), \tag{4.5}$$

whenever $\sum_{q\geqslant 1} q^2\mathbb{E}[I_q(f_q)^2] < \infty$. The pseudo-inverse L^{-1} is, however, defined for all $F \in L^2(\Omega)$ [39, Defs 2.8.7–2.8.10]. Let $F \in L^2(\Omega)$, then $F \in \mathrm{Dom}(L)$ if and only if $F \in \mathbb{D}^{1,2}$ and $DF \in \mathrm{Dom}(\delta)$, in which case $\delta DF = -LF$ [39, Prop. 2.8.8].

A central objects in our proofs is the Stein kernel. A measurable $d \times d$ matrix-valued function $\boldsymbol{x} \mapsto \boldsymbol{\tau}^{\boldsymbol{F}}(\boldsymbol{x}) = (\boldsymbol{\tau}^{\boldsymbol{F}}_{i,j}(\boldsymbol{x}))_{i,j \in \{1,\ldots,d\}}$ on $\mathbb{R}^d$ is the Stein's kernel for the law of $\boldsymbol{F}$ if $\mathbb{E}[|\boldsymbol{\tau}^{\boldsymbol{F}}_{i,j}(\boldsymbol{F})|] < \infty$ for any $i, j \in \{1, \ldots, d\}$ and

$$\sum_{j=1}^{d} \mathbb{E}[\partial_j f(\boldsymbol{F}) F_j] = \sum_{i,j=1}^{d} \mathbb{E}[\partial_{i,j} f(\boldsymbol{F}) \boldsymbol{\tau}^{\boldsymbol{F}}_{i,j}(\boldsymbol{F})],$$

for all functions $f : \mathbb{R}^d \to \mathbb{R}$ where $f \in C^\infty$ with bounded partial derivatives of all orders [42, Def. 2.7]. Stein's kernels often exist in the Malliavin calculus setting. Indeed, if $\boldsymbol{F}$ is a centred random vector with $F_k \in \mathbb{D}^{1,2}$ for $k \in \{1, \ldots, d\}$, then [42, Prop. 3.7] implies that $\boldsymbol{F}$ has a Stein's kernel given by

$$\boldsymbol{\tau}^{\boldsymbol{F}}_{i,j}(\boldsymbol{x}) = \mathbb{E}[\langle -DL^{-1} F_i, DF_j \rangle_{\mathfrak{H}} | \boldsymbol{F} = \boldsymbol{x}], \quad \text{for all } i, j \in \{1, \ldots, d\}, \text{ and } \boldsymbol{x} \in \mathbb{R}^d. \tag{4.6}$$

Finally, we recall a useful notion, the sub-chaos property. For $q \in \mathbb{N}$ and $M \geqslant 0$, a random variable Y is called *sub-q-th chaos relative to scale M* if $\mathbb{E}[\exp((|Y|/M)^{2/q})] \leqslant 2$, and it is sub-$q$-th chaos relative to 0 if and only if $Y = 0$ a.s. Moreover, for each $q \in \mathbb{N}$ there exists a constant $C_q > 0$, depending only on q, such that every $Y \in \mathcal{P}_q$ is sub-q-th chaos relative to scale $C_q \|Y\|_2$; see [32, Def. A.1 and Prop. A.1]. We also use the standard tensor-product identity: if $\mathfrak{H}_1, \mathfrak{H}_2$ are Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$, then for $f_1, g_1 \in \mathfrak{H}_1$ and $f_2, g_2 \in \mathfrak{H}_2$,

$$\langle f_1 \otimes f_2,\ g_1 \otimes g_2 \rangle = \langle f_1, g_1 \rangle_1 \langle f_2, g_2 \rangle_2. \tag{4.7}$$

4.2 Methodology of Proofs

Our proofs are based on the Malliavin–Stein method, which yields non-asymptotic bounds in probabilistic distances via explicit control of Stein kernels. For the Theorem 2.1, we start from the general multivariate bounds of [25, Thm. 1.1], [44, Thm. 2.1], and [40, Thm. 3.5], respectively. These are also presented below in (4.8), (4.9) and (4.10). In the setting of Theorem 2.1, the objects of interest involve multiple Wiener–Itô integrals, for which Stein kernels admit explicit representations, see [42, Prop. 3.7] and (4.6). We then reduce the resulting kernel terms using the combinatorial ideas from [39, Lem. 6.2.1], combined with insights from [32, Lem. A.1 and Prop. A.2] about sub-Gaussian random variables, [31, Lem. 2.2]. Finally, Stirling's inequality (4.18) is the main ingredient when showing the summability of key series, namely (4.11) and (4.12).

For the quantitative Breuer–Major results and the parameter estimation problem in stable VAR(p) models, the proofs reduces to (i) identifying the relevant multiple Wiener–Itô integral structure, (ii) writing down the associated kernels explicitly, and (iii) calculating the contraction norms arising from (2.1) and invoking Theorem 4.1 and Theorem 3.1, respectively.

The main inequalities used in the proof of Theorem 2.1 are: (4.8), (4.9) and (4.10) stated below. If $\boldsymbol{F}$ admits a Stein kernel $\boldsymbol{\tau}^{\boldsymbol{F}}$ and $\boldsymbol{Z} \sim \mathcal{N}_d(0, \boldsymbol{\Sigma})$ with $\sigma_* := \sigma_*(\boldsymbol{\Sigma}) > 0$, then [25, Thm. 1.1] yields

$$d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant \frac{C \Delta_{\boldsymbol{F}} \log_+(d)}{\sigma_*} \log_+\left(\frac{\sigma \Delta_{\boldsymbol{F}}}{\overline{\sigma} \sigma_*}\right), \qquad \Delta_{\boldsymbol{F}} := \mathbb{E}\left[\max_{1 \leqslant i,j \leqslant d} \left|\boldsymbol{\Sigma}_{ij} - \boldsymbol{\tau}^{\boldsymbol{F}}_{i,j}(\boldsymbol{F})\right|\right], \tag{4.8}$$

for all $d \geqslant 3$, where $\overline{\sigma}$ and $\underline{\sigma}$ are both associated with $\mathbf{\Sigma}$.

For a vector $\boldsymbol{F} = (\delta(u_1), \dots, \delta(u_d))$ in $\mathbb{R}^d$ of centred random variables such that $u_i \in \mathrm{Dom}(\delta)$ for $i \in \{1, \dots, d\}$, and for $\boldsymbol{Z} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{\Sigma})$ assuming $\mathbf{\Sigma}$ is invertible, for $M_{\boldsymbol{F}}(i,j) := \langle DF_i, u_j \rangle_{\mathfrak{H}}$, it holds that

$$d_{\mathscr{C}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant 402 \left(\|\mathbf{\Sigma}^{-1}\|_{\mathrm{op}}^{3/2} + 1 \right) d^{41/24} \sqrt{\mathbb{E}\left[\|M_{\boldsymbol{F}} - \mathbf{\Sigma}\|_{\mathrm{H.S.}}^2 \right]}, \quad \text{and} \tag{4.9}$$

$$d_{\mathcal{W}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant \sqrt{d} \|\mathbf{\Sigma}^{-1}\|_{\mathrm{op}} \|\mathbf{\Sigma}\|_{\mathrm{op}}^{1/2} \sqrt{\mathbb{E}\left[\|M_{\boldsymbol{F}} - \mathbf{\Sigma}\|_{\mathrm{H.S.}}^2 \right]}. \tag{4.10}$$

For (4.9), see [44, Thm 1.2], and for (4.10) see [44, Sec. 1.3], or the statement in the original paper [40, Thm 3.5].

4.3 General Main Theorem and Corollary

In this subsection, we present two generalisations of Theorem 2.1. Theorem 4.1 is our most general result, and it generalises Theorem 2.1 in two directions: (i) it no longer requires the covariance matrices of the target vector $\boldsymbol{F}$ and the Gaussian reference vector $\boldsymbol{Z}$ to coincide, and (ii) as anticipated in Remark 2.2(4), there are no assumptions on contraction bound, instead it is enough that the two sums below (4.11) and (4.12) are finite.

When the contraction bound (2.1) is imposed, the sums (4.11) and (4.12) are finite, and Theorem 4.1 specialises to Corollary 4.2. This corollary still strictly extends Theorem 2.1, since it continues to allow for non-matching covariance structures between $\boldsymbol{F}$ and $\boldsymbol{Z}$.

For the rest of the section, fix integer $d \geqslant 1$, and define the following

$$\Phi(d) := \sum_{\substack{p,q \geqslant 1 \\ (p,q) \neq (1,1)}} q(4\sqrt{e})^{p+q} \log^{(p+q)/2-1}(2d^2 + e^{(p+q)/2-2}) \Xi(p,q,d), \qquad \text{and} \tag{4.11}$$

$$\Psi(d) := \sum_{\substack{p,q \geqslant 1 \\ (p,q) \neq (1,1)}} \frac{\sqrt{p}(p!q!)^{1/2}}{\sqrt{q}} 2^{p+q} \Xi(p,q,d), \tag{4.12}$$

where $\Xi(p,q,d) := \sum_{r=1}^{p \wedge q - \mathbb{1}_{\{p=q\}}} \max_{1 \leqslant i,j \leqslant d} \|f_{i,p} \otimes_r f_{j,q}\|_{\mathfrak{H}^{\otimes(p+q-2r)}}$, for $f_{i,p} \in \mathfrak{H}^{\odot p}$ defined uniquely by $F_i = \sum_{i=0}^{\infty} I_p(f_{i,p})$ for all $i = 1, \dots, d$, $k \in \{1, \dots, d\}$, and $p \in \mathbb{N}$.

Theorem 4.1. *Fix integer $d \geqslant 1$. Let $\boldsymbol{F} = (F_1, \dots, F_d)$ be a d-dimensional vector such that $F_i = \sum_{p=1}^{\infty} I_p(f_{i,p})$, for $f_{i,p} \in \mathfrak{H}^{\odot p}$, $i \in \{1, \dots d\}$. Assume $\mathbf{\Sigma}$ is invertible, and let $\boldsymbol{Z} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{\Sigma})$, and $\mathbf{\Lambda} := \mathrm{Corr}(\boldsymbol{Z})$. For $\Phi(d)$ defined in (4.11), and $\Psi(d)$ in (4.12) define*

$$\Delta_{\boldsymbol{F}} = \Phi(d) + \max_{1 \leqslant i,j \leqslant d} \left| \mathbf{\Lambda}_{ij} - \frac{\mathbb{E}[F_i F_j]}{(\mathbf{\Sigma}_{ii} \mathbf{\Sigma}_{jj})^{1/2}} \right|.$$

Then, there exists a constant $C > 0$ independent of d, such that

$$d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C \log(d) \Delta_{\boldsymbol{F}} \log_+(\Delta_{\boldsymbol{F}}) \frac{\log_+(\sigma_*(\mathbf{\Lambda}))}{\sigma_*(\mathbf{\Lambda})},$$

$$d_{\mathscr{C}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant 402 d^{41/24} \left(\sqrt{\sum_{i,j=1}^{d} \left(\mathbf{\Lambda}_{ij} - \frac{\mathbb{E}[F_i F_j]}{(\mathbf{\Sigma}_{ii} \mathbf{\Sigma}_{jj})^{1/2}} \right)^2} + d\Psi(d) \right) \frac{1}{\sigma_*(\mathbf{\Lambda})^{3/2}}, \qquad \textit{and}$$

$$d_{\mathcal{W}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant \sqrt{d} \left(\|\mathbf{\Sigma} - \mathrm{Cov}(\boldsymbol{F})\|_{\mathrm{H.S.}} + d\Psi(d) \right) \frac{\sigma^*(\mathbf{\Sigma})^{1/2}}{\sigma_*(\mathbf{\Sigma})}.$$

The following corollary shows that under assumption (2.1), $\Phi(d)$ and $\Psi(d)$ are finite, and we get an explicit dependence on d.

Corollary 4.2. *In the setting as in Theorem 4.1 , let $\gamma, \alpha \in \mathbb{R}$ and $\beta \in [1/2, 1]$. Assume*

$$\|f_{i,p} \otimes_r f_{j,q}\|_{\mathfrak{H}^{\otimes(p+q-2r)}} \leqslant \frac{\gamma e^{\alpha p} e^{\alpha q}}{(p!q!)^{\beta}} \qquad \text{for all } i, j \in \{1, \dots, d\}, \tag{4.13}$$

and all integers p, q and r, where $p, q \geqslant 1$, not simultaneously 1, and $1 \leqslant r \leqslant p \wedge q - \mathbb{1}_{\{p=q\}}$. If $\beta = 1/2$, for (4.15) assume additionally that $\alpha < \alpha_0$ (recall (1.3)), and for (4.16) and (4.17) that $\alpha < \log(1/2) - e^{1/(2e)}$. Define

$$\Delta_{\boldsymbol{F}} = \gamma \frac{e^{k \log_+^{1/2\beta}(d)}}{\log_+(d)} + \max_{1 \leqslant i,j \leqslant d} \left| \boldsymbol{\Lambda}_{ij} - \frac{\mathbb{E}[F_i F_j]}{(\boldsymbol{\Sigma}_{ii}\boldsymbol{\Sigma}_{jj})^{1/2}} \right|. \tag{4.14}$$

Then there exists a constant $C > 0$ depending only on α and β, such that

$$d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C \log(d) \frac{\Delta_{\boldsymbol{F}}}{\sigma_*(\boldsymbol{\Lambda})} \log_+ \left(\frac{\Delta_{\boldsymbol{F}}}{\sigma_*(\boldsymbol{\Lambda})} \right), \tag{4.15}$$

$$d_{\mathscr{C}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant 402 d^{41/24} \left(\sqrt{\sum_{i,j=1}^{d} \left(\boldsymbol{\Lambda}_{ij} - \frac{\mathbb{E}[F_i F_j]}{(\boldsymbol{\Sigma}_{ii}\boldsymbol{\Sigma}_{jj})^{1/2}} \right)^2} + dC \right) \frac{1}{\sigma_*(\boldsymbol{\Lambda})^{3/2}} \quad \text{and} \tag{4.16}$$

$$d_{\mathcal{W}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant \sqrt{d} \left(\|\boldsymbol{\Sigma} - \operatorname{Cov}(\boldsymbol{F})\|_{\text{H.S.}} + dC \right) \frac{\sigma^*(\boldsymbol{\Sigma})^{1/2}}{\sigma_*(\boldsymbol{\Sigma})}. \tag{4.17}$$

In order to prove these results, we present a couple of technical results. The first is built on [39, Lem. 6.2.1].

Lemma 4.3. *Let $F = I_p(f)$ and $G = I_q(g)$ with $f \in \mathfrak{H}^{\odot p}$ and $g \in \mathfrak{H}^{\odot q}$. Further, let $\alpha = \mathbb{E}[FG]$. Then it holds that*

$$\mathbb{E}\left[\left(\mathbb{E}[FG] - \frac{1}{p} \langle DF, DG \rangle_{\mathfrak{H}} \right)^2 \right] \leqslant \frac{qp!q!}{p} 2^{2p+2q} \sum_{r=1}^{b(p,q)} \|f \otimes_r g\|^2_{\mathfrak{H}^{\otimes(p+q-2r)}},$$

for $p, q \geqslant 1$, where $b(p,q) = p \wedge q - \mathbb{1}_{\{p=q\}}$.

The following Lemma shows that [32, Prop. A.2] holds for $q = 1$ as well.

Lemma 4.4. *Fix integer $q \geqslant 1$. For each $k \in \{1, \dots, m\}$, let Y_k be a sub-q-th chaos random variable relative to scale $M_k \geqslant 0$. Then we have*

$$\mathbb{E}\Big[\max_{1 \leqslant k \leqslant m} |Y_k| \Big] \leqslant \Big(\max_{1 \leqslant k \leqslant m} M_k \Big) \log^{q/2}(2m + e^{q/2-1}).$$

And finally, a useful inequality used extensively throughout the proofs, that builds on the repetitive application of the Stirling's inequality [50, Eqs (1) and (2)]

$$\sqrt{2\pi n} \left(\frac{n}{e} \right)^n < n! < e^{1/12} \sqrt{2\pi n} \left(\frac{n}{e} \right)^n, \qquad \text{for all } n \geqslant 1. \tag{4.18}$$

Lemma 4.5. *Let $\varpi > 1/2$, and $u(d)$ some positive function. Then it follows that*

$$\sum_{q\geqslant 0} \frac{u(d)^q}{(q!)^{\varpi}} \leqslant 2^{2\varpi} \left(\frac{\pi e^{1/2}}{2}\right)^{\varpi/2} e^{\varpi e^{1/(2e)} u(d)^{1/\varpi}}.$$

Remark 4.6. Let $\boldsymbol{F}$ and $\boldsymbol{Z}$ be d-dimensional random vectors, with $\boldsymbol{\Sigma}$, the covariance matrix of $\boldsymbol{Z}$. Note that $d_{\mathcal{R}}$ is invariant under multiplication with a diagonal matrix, meaning

$$d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) = d_{\mathcal{R}}(\boldsymbol{DF}, \boldsymbol{DZ}),$$

for any strictly positive diagonal matrix $\boldsymbol{D}$. Specifically, one can take the correlation matrix $\boldsymbol{\Lambda}$ corresponding to $\boldsymbol{\Sigma}$, as it is well known that $\boldsymbol{\Lambda} = \boldsymbol{D\Sigma D}$, for a diagonal matrix with positive entries $\boldsymbol{D}$. For $\boldsymbol{\Lambda}$, we know that $\underline{\sigma}(\boldsymbol{\Lambda}) = \overline{\sigma}(\boldsymbol{\Lambda}) = 1$ and that the smallest eigenvalue $\sigma_*(\boldsymbol{\Lambda})$ is certainly less than or equal to 1, as $\boldsymbol{\Lambda}$ is a correlation matrix.

Assume now that $d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant f(\boldsymbol{F}, \boldsymbol{Z})$, for a particular positive function f and let $\mathcal{D} := \{\boldsymbol{D} \in \mathbb{R}^d : \boldsymbol{D}_{i,i} > 0, \text{ and } \boldsymbol{D}_{i,j} = 0 \text{ for } i \neq j, i,j \in \{1,\dots,d\}\}$. As infimum is preserved under soft inequalities, it then follows $d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) = \inf_{\boldsymbol{D}\in\mathcal{D}} d_{\mathcal{R}}(\boldsymbol{DF}, \boldsymbol{DZ}) \leqslant \inf_{\boldsymbol{D}\in\mathcal{D}} f(\boldsymbol{DF}, \boldsymbol{DZ})$, and specifically that $d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant f(\boldsymbol{DF}, \boldsymbol{DZ})$.

Now we are ready to prove Theorem 4.1 and Corollary 4.2.

Proof of Theorem 4.1. We split the proof in three parts, one for each distance considered. **Part 1.** We begin with hyper-rectangular distance. As $\boldsymbol{F}$ is centred with elements in $\mathbb{D}^{1,2}$, by [42, Prop. 3.7] it has a Stein's kernel given by

$$\boldsymbol{\tau}^{\boldsymbol{F}}_{j,k}(\boldsymbol{x}) = \mathbb{E}[\langle -DL^{-1}F_j, DF_k\rangle_{\mathfrak{H}} \mid \boldsymbol{F} = \boldsymbol{x}], \quad \text{for all } j,k \in \{1,\dots,d\}, \text{ and } \boldsymbol{x} \in \mathbb{R}^d.$$

Hence, [25, Thm 1.1] yields that for all $d \geqslant 3$

$$\begin{aligned} d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) &\leqslant C\frac{\Delta_{\boldsymbol{F}}}{\sigma_*} \log(d) \log_+\left(\frac{\underline{\sigma}\Delta_{\boldsymbol{F}}}{\overline{\sigma}\sigma_*}\right), \quad \text{where} \\ \Delta_{\boldsymbol{F}} &:= \mathbb{E}\Big[\max_{1\leqslant j,k\leqslant d} |\boldsymbol{\Sigma}_{jk} - \tau^{\boldsymbol{F}_{j,k}(\boldsymbol{F})}|\Big]. \end{aligned} \tag{4.19}$$

If $d = 1$ or $d = 2$, then [3, Lem. 5.4] yields that

$$d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C\frac{\Delta_{\boldsymbol{F}}}{\sigma_* \wedge 1} \log(d) \log_+\left(\frac{(\underline{\sigma} \wedge 1)\Delta_{\boldsymbol{F}}}{(\overline{\sigma} \vee 1)(\sigma_* \wedge 1)}\right).$$

Hence, the main goal of the rest of the proof is finding a good bound for $\Delta_{\boldsymbol{F}}$. Begin by plugging in $\boldsymbol{\tau}^{\boldsymbol{F}}_{j,k}(\boldsymbol{F})$ into (4.19). Using Jensen's inequality for conditional expectation, the fact that $\max_i \mathbb{E}[X_i|\mathcal{F}] \leqslant \mathbb{E}[\max_i X_i|\mathcal{F}]$, and the law of total expectation yields

$$\begin{aligned} \Delta_{\boldsymbol{F}} :&= \mathbb{E}\Big[\max_{1\leqslant j,k\leqslant d} |\boldsymbol{\Sigma}_{jk} - \mathbb{E}[\langle -DL^{-1}F_j, DF_k\rangle_{\mathfrak{H}} \mid \boldsymbol{F}]|\Big] \\ &= \mathbb{E}\Big[\max_{1\leqslant j,k\leqslant d} |\mathbb{E}[\boldsymbol{\Sigma}_{jk} - \langle -DL^{-1}F_j, DF_k\rangle_{\mathfrak{H}} \mid \boldsymbol{F}]|\Big] \\ &\leqslant \mathbb{E}\Big[\max_{1\leqslant j,k\leqslant d} \mathbb{E}[|\boldsymbol{\Sigma}_{jk} - \langle -DL^{-1}F_j, DF_k\rangle_{\mathfrak{H}}| \mid \boldsymbol{F}]\Big] \\ &= \mathbb{E}\Big[\max_{1\leqslant j,k\leqslant d} |\boldsymbol{\Sigma}_{jk} - \langle -DL^{-1}F_j, DF_k\rangle_{\mathfrak{H}}|\Big]. \end{aligned} \tag{4.20}$$

Continuing from the obtained bound, adding and subtracting $\mathbb{E}[F_jF_k]$, using triangle inequality, the fact that max is sub-additive, we get

$$\begin{aligned}\Delta_{\boldsymbol{F}} :&= \mathbb{E}\bigg[\max_{1\leqslant j,k\leqslant d}|\boldsymbol{\Sigma}_{jk} - \langle -DL^{-1}F_j, DF_k\rangle_{\mathfrak{H}}|\bigg] \\ &\leqslant \max_{1\leqslant j,k\leqslant d}|\boldsymbol{\Sigma}_{jk} - \mathbb{E}[F_jF_k]| + \mathbb{E}\left[\max_{1\leqslant j,k\leqslant d}|\mathbb{E}[F_jF_k] - \langle -DL^{-1}F_j, DF_k\rangle_{\mathfrak{H}}|\right] \\ &= \max_{1\leqslant j,k\leqslant d}|\boldsymbol{\Sigma}_{jk} - \mathbb{E}[F_jF_k]| + \mathbb{E}\bigg[\max_{1\leqslant j,k\leqslant d}|\Delta_{j,k}|\bigg] \\ &= \max_{1\leqslant j,k\leqslant d}|\boldsymbol{\Sigma}_{jk} - \mathbb{E}[F_jF_k]| + \widetilde{\Delta}_{\boldsymbol{F}},\end{aligned} \tag{4.21}$$

where $\Delta_{j,k} := \mathbb{E}[F_jF_k] - \langle -DL^{-1}F_j, DF_k\rangle_{\mathfrak{H}}$, for $j,k \in \{1,\dots,d\}$ and

$$\widetilde{\Delta}_{\boldsymbol{F}} := \mathbb{E}[\max_{1\leqslant j,k\leqslant d}|\Delta_{j,k}|]. \tag{4.22}$$

Hence, we want to bound $\widetilde{\Delta}_{\boldsymbol{F}}$. By (4.5), together with the linearity of the derivative operator and the inner product, it follows that for all $j,k \in \{1,\dots,d\}$,

$$\begin{aligned}\Delta_{j,k} &= \mathbb{E}[F_jF_k] - \Big\langle -DL^{-1}\sum_{p\geqslant 1} I_p(f_{j,p}), D\sum_{q\geqslant 1} I_q(f_{k,q})\Big\rangle_{\mathfrak{H}} \\ &= \mathbb{E}[F_jF_k] - \Big\langle D\sum_{p\geqslant 1} p^{-1}I_p(f_{j,p}), D\sum_{q\geqslant 1} I_q(f_{k,q})\Big\rangle_{\mathfrak{H}} \\ &= \sum_{p\geqslant 1}\sum_{q\geqslant 1}\big(\mathbb{E}[I_p(f_{j,p})I_q(f_{k,q})] - p^{-1}\langle DI_p(f_{j,p}), DI_q(f_{k,q})\rangle_{\mathfrak{H}}\big) = \sum_{p\geqslant 1}\sum_{q\geqslant 1}\alpha(j,k,p,q),\end{aligned}$$

where $\alpha(j,k,p,q) := \mathbb{E}[I_p(f_{j,p})I_q(f_{k,q})] - p^{-1}\langle DI_p(f_{j,p}), DI_q(f_{k,q})\rangle_{\mathfrak{H}}$. Rewriting $\alpha(j,k,p,q)$ with the use of [39, Prop. 2.7.4] and the product formula [39, Thm 2.7.10] one gets

$$\begin{aligned}\alpha(j,k,p,q) &= \mathbb{E}[I_p(f_{j,p})I_q(f_{k,q})] - q\langle I_{p-1}(f_{j,p}), I_{q-1}(f_{k,q})\rangle_{\mathfrak{H}} \\ &= \mathbb{E}[I_p(f_{j,p})I_q(f_{k,q})] - q\sum_{r=1}^{p\wedge q}(r-1)!\binom{p-1}{r-1}\binom{q-1}{r-1}I_{p+q-2r}(f\widetilde{\otimes}_r g).\end{aligned}$$

Note that when $p = q = 1$, we get

$$\alpha(j,k,1,1) = \langle f_{j,1}, f_{k,1}\rangle_{\mathfrak{H}} - I_0(f_{j,1}\otimes_1 f_{k,1}) = 0.$$

Hence, we can assume that p and q are not simultaneously 1, as $\Delta_{j,k} = \sum_{\substack{p,q\geqslant 1\\(p,q)\neq(1,1)}} \alpha(j,k,p,q)$.

For all non-zero $\alpha(j,k,p,q)$, it holds that $\alpha(j,k,p,q) \in \mathcal{P}_{\varpi}(\mathfrak{H})$, and hence, according to [32, Prop. A.1], $\alpha(j,k,p,q)$ is a sub-ϖ-chaos random variable relative to the scale $M_{\varpi}||\alpha(j,k,p,q)||_2$, where we know (from [3, Proof of Thm 5.1]) that $M_{\varpi} = (4e/\varpi)^{\varpi/2}$. Lemma 4.4 then yields that

$$\mathbb{E}\left[\max_{1\leqslant j,k\leqslant d}|\alpha(j,k,p,q)|\right] \leqslant \log^{\varpi/2}(2d^2 + e^{\varpi/2})\max_{1\leqslant j,k\leqslant d} M_{\varpi}\|\alpha(j,k,p,q)\|_2. \tag{4.23}$$

To bound $\max_{1\leqslant j,k\leqslant d}\|\alpha(j,k,p,q)\|_2$ we use Lemma 4.3 for $p,q \in \mathbb{N}$, $(p,q) \neq (1,1)$, and the sub-additivity of the square root, such that

$$\max_{1\leqslant j,k\leqslant d}\|\alpha(j,k,p,q)\|_2 \leqslant \left(\frac{qp!q!}{p}\right)^{1/2} 2^{p+q}\Xi(p,q,d) \quad \text{with} \tag{4.24}$$

$$\Xi(p,q,d) = \sum_{r=1}^{b(p,q)} \max_{1\leqslant j,k,\leqslant d} \|f_{j,p}\otimes_r f_{k,q}\|_{\mathfrak{H}^{\otimes(p+q-2r)}}, \quad \text{where } b(p,q) = p\wedge q - \mathbb{1}_{\{p=q\}}.$$

Plugging $M_{p+q-2} = (4e/(p+q-2))^{(p+q-2)/2}$ and (4.24) into (4.23), we get for $p,q\geqslant 1$, when $(p,q)\neq(1,1)$, that

$$\begin{aligned}\mathbb{E}\left[\max_{1\leqslant j,k\leqslant d}|\alpha(j,k,p,q)|\right] &\leqslant \frac{1}{4e}\frac{\sqrt{qp!q!}}{\sqrt{p}}\frac{\sqrt{pq}(4e)^{(p+q)/2}}{(p!q!)^{1/2}}\log^{(p+q-2)/2}(2d^2-1+e^{\frac{p+q}{2}-2})\Xi(p,q,d)\\ &\leqslant \frac{1}{4e}q(4e^{1/2})^{p+q}\log^{(p+q-2)/2}(2d^2-1+e^{(p+q)/2-2})\Xi(p,q,d),\end{aligned}$$

where we used that $(p+q-2)^{(p+q-2)/2} = (p+q-2)^{(p-1)/2}(p+q-2)^{(q-1)/2}\geqslant p^{(p-1)/2}q^{(q-1)/2}$ for $p,q\geqslant 1$ and that $p^{-(p-1)/2} = \sqrt{p}p^{-p/2}\leqslant\sqrt{p}(p!)^{-1/2}$, as $p^p\geqslant p!$. Recall (4.21) and (4.22). Then using the triangular inequality and sub-additivity of max, it follows that

$$\begin{aligned}\widetilde{\Delta}_{\boldsymbol{F}} &= \mathbb{E}\left[\max_{1\leqslant j,k\leqslant d}|\Delta_{j,k}|\right] = \mathbb{E}\left[\max_{1\leqslant j,k\leqslant d}\Big|\sum_{\substack{p,q\geqslant 1\\(p,q)\neq(1,1)}}\alpha(j,k,p,q)\Big|\right]\\ &\leqslant \mathbb{E}\left[\max_{1\leqslant j,k\leqslant d}\sum_{\substack{p,q\geqslant 1\\(p,q)\neq(1,1)}}|\alpha(j,k,p,q)|\right]\leqslant\sum_{\substack{p,q\geqslant 1\\(p,q)\neq(1,1)}}\mathbb{E}\left[\max_{1\leqslant j,k\leqslant d}|\alpha(j,kp,q)|\right]\\ &\leqslant\frac{1}{4e}\sum_{\substack{p,q\geqslant 1\\(p,q)\neq(1,1)}}q(4e^{1/2})^{p+q}\log^{(p+q-2)/2}(2d^2-1+e^{(p+q)/2-2})\Xi(p,q,d) =: \Phi(d).\end{aligned}$$

Under the assumption that $\Phi(d)$ is finite, it follows from (4.21) that $\Delta_{\boldsymbol{F}}\leqslant c\Phi(d)+\max_{1\leqslant j,k\leqslant d}|\boldsymbol{\Sigma}_{jk}-\mathbb{E}[F_jF_k]|$. We can then conclude from (4.19) that there exist positive constants C_1 and C_2 such that for $d\geqslant 3$

$$\begin{aligned}d_{\mathcal{R}}(\boldsymbol{F},\boldsymbol{Z}) &\leqslant C_1\frac{\overline{\Delta}_{\boldsymbol{F}}(\boldsymbol{F},\boldsymbol{\Sigma})}{\sigma_*}\log(d)\log_+\left(\frac{\underline{\sigma}\overline{\Delta}_{\boldsymbol{F}}(\boldsymbol{F},\boldsymbol{\Sigma})}{\overline{\sigma}\sigma_*}\right), \quad \text{where}\\ \overline{\Delta}_{\boldsymbol{F}}(\boldsymbol{F},\boldsymbol{\Sigma}) &:= \Phi(d)+\max_{1\leqslant j,k\leqslant d}|\boldsymbol{\Sigma}_{jk}-\mathbb{E}[F_jF_k]|.\end{aligned}\tag{4.25}$$

On the other hand, for $d\in\{1,2\}$, one gets

$$d_{\mathcal{R}}(\boldsymbol{F},\boldsymbol{Z})\leqslant C_2\frac{\overline{\Delta}_{\boldsymbol{F}}(\boldsymbol{F},\boldsymbol{\Sigma})}{\sigma_*\wedge 1}\log(d)\log_+\left(\frac{(\underline{\sigma}\wedge 1)\overline{\Delta}_{\boldsymbol{F}}(\boldsymbol{F},\boldsymbol{\Sigma})}{(\overline{\sigma}\vee 1)(\sigma_*\wedge 1)}\right).$$

We have shown that (for all $d\geqslant 1$) $d_{\mathcal{R}}(\boldsymbol{F},\boldsymbol{Z})\leqslant f(\boldsymbol{F},\boldsymbol{Z})$, for a specific positive f. Choose $\boldsymbol{D} = \mathrm{Diag}(\boldsymbol{\Sigma}_{11}^{-1},\ldots,\boldsymbol{\Sigma}_{dd}^{-1})$. Then, due to Remark 4.6, it follows that for correlation matrix $\boldsymbol{\Lambda} = \boldsymbol{D}\boldsymbol{\Lambda}\boldsymbol{D}$ it holds that

$$d_{\mathcal{R}}(\boldsymbol{F},\boldsymbol{Z})\leqslant f(\boldsymbol{D}\boldsymbol{F},\boldsymbol{D}\boldsymbol{Z}).$$

From (4.25) it is clear that we have to calculate $\overline{\Delta}_{\boldsymbol{F}}(\boldsymbol{D}\boldsymbol{F},\boldsymbol{\Lambda})$. To that end, as $\boldsymbol{\Lambda}_{jk} = (\boldsymbol{\Sigma}_{jj}^{1/2}\boldsymbol{\Sigma}_{kk}^{1/2})^{-1}\boldsymbol{\Sigma}_{jk}$, it holds that $|\boldsymbol{\Lambda}_{jk}-\mathbb{E}[(\boldsymbol{D}\boldsymbol{F})_j(\boldsymbol{D}\boldsymbol{F})_k]| = |\boldsymbol{\Lambda}_{jk}-\left(\boldsymbol{\Sigma}_{jj}^{1/2}\boldsymbol{\Sigma}_{kk}^{1/2}\right)^{-1}\mathbb{E}[F_jF_k]|$, which leads to

$$\widetilde{\Delta}_{\boldsymbol{F}} := \overline{\Delta}_{\boldsymbol{F}}(\boldsymbol{D}\boldsymbol{F},\boldsymbol{\Lambda}) = \Phi(d)+\max_{1\leqslant j,k\leqslant d}\left|\boldsymbol{\Lambda}_{jk}-\frac{\mathbb{E}[F_jF_k]}{(\boldsymbol{\Sigma}_{jj}\boldsymbol{\Sigma}_{kk})^{1/2}}\right|$$

and to a more elegant statement of the above bounds, namely, for all $d \geqslant 1$ we have shown that

$$d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) = d_{\mathcal{R}}(\boldsymbol{DF}, \boldsymbol{DZ}) \leqslant C \frac{\widetilde{\Delta}_{\boldsymbol{F}}}{\sigma_*(\boldsymbol{\Lambda})} \log(d) \log_+ \left(\frac{\widetilde{\Delta}_{\boldsymbol{F}}}{\sigma_*(\boldsymbol{\Lambda})} \right), \quad \text{where}$$

$$\widetilde{\Delta}_{\boldsymbol{F}} = \Phi(d) + \max_{1 \leqslant j,k \leqslant d} \left| \boldsymbol{\Lambda}_{jk} - \frac{\mathbb{E}[F_j F_k]}{(\boldsymbol{\Sigma}_{jj} \boldsymbol{\Sigma}_{kk})^{1/2}} \right|.$$

Part 2. In order to prove the convex distance case, we utilise [44, Thm 1.2], also given in (4.9), which is given for vector $\boldsymbol{F}$ with components of the form $\boldsymbol{F}_i = \delta(u_i)$, such that $u_i \in \operatorname{Dom}(\delta)$. In our case, we have that $F_k = \sum_{q \geqslant 1} I_q(f_{k,q})$ for $f_{k,q} \in \mathfrak{H}^{\odot q}$ and all $k \in \{1, \ldots, d\}$. As discussed in [44] one can always take $u_k = -D(L^{-1}F_k)$ for all $k \in \{1, \ldots, d\}$, such that $\delta(u_k) = -\delta(DL^{-1}F_k) = F_k$, and hence (4.9) holds in our setting.

Now it remains to deal with $\sqrt{\mathbb{E}[\|M_{\boldsymbol{F}} - \boldsymbol{\Sigma}\|_{\text{H.S.}}^2]}$. Recall from (4.9), that $M_{\boldsymbol{F}}(i,j) = \langle DF_i, u_j \rangle_{\mathfrak{H}}$ for $i, j \in \{1, \ldots, d\}$. By the sub-additivity of the Hilbert-Schmidt and L^2-norm it follows that

$$\begin{aligned}
&\sqrt{\mathbb{E}[\|M_{\boldsymbol{F}} - \boldsymbol{\Sigma}\|_{\text{H.S.}}^2]} \leqslant \sqrt{\mathbb{E}[(\|M_{\boldsymbol{F}} - \operatorname{Cov}(\boldsymbol{F})\|_{\text{H.S.}} + \|\operatorname{Cov}(\boldsymbol{F}) - \boldsymbol{\Sigma}\|_{\text{H.S.}})^2]} \\
&= \big\| \|M_{\boldsymbol{F}} - \operatorname{Cov}(\boldsymbol{F})\|_{\text{H.S.}} + \|\operatorname{Cov}(\boldsymbol{F}) - \boldsymbol{\Sigma}\|_{\text{H.S.}} \big\|_{L^2} \\
&\leqslant \sqrt{\mathbb{E}[\|M_{\boldsymbol{F}} - \operatorname{Cov}(\boldsymbol{F})\|_{\text{H.S.}}^2]} + \|\operatorname{Cov}(\boldsymbol{F}) - \boldsymbol{\Sigma}\|_{\text{H.S.}}.
\end{aligned}$$

Bounding the $\sum_{i,j=1}^d$ in the Hilbert-Schmidt norm in $\|M_{\boldsymbol{F}} - \operatorname{Cov}(\boldsymbol{F})\|_{\text{H.S.}}$ by the maximum over $i, j \in \{1, \ldots, d\}$ and using the sub-additivity of the square root,one then obtains

$$\begin{aligned}
\sqrt{\mathbb{E}[\|M_{\boldsymbol{F}} - \boldsymbol{\Sigma}\|_{\text{H.S.}}^2]} &\leqslant \|\operatorname{Cov}(\boldsymbol{F}) - \boldsymbol{\Sigma}\|_{\text{H.S.}} + \left(d^2 \max_{1 \leqslant i,j \leqslant d} \|\operatorname{Cov}(\boldsymbol{F})_{ij} - M_{\boldsymbol{F}}(i,j)\|_2^2 \right)^{1/2} \\
&\leqslant \|\operatorname{Cov}(\boldsymbol{F}) - \boldsymbol{\Sigma}\|_{\text{H.S.}} + d \max_{1 \leqslant i,j \leqslant d} \|\operatorname{Cov}(\boldsymbol{F})_{ij} - M_{\boldsymbol{F}}(i,j)\|_2
\end{aligned}$$

The definition of $M_{\boldsymbol{F}}$ and (4.5) now yield that

$$\begin{aligned}
&\operatorname{Cov}(\boldsymbol{F})_{ij} - M_{\boldsymbol{F}}(i,j) = \mathbb{E}[F_i F_j] - \langle DF_i, -DL^{-1}F_j \rangle_{\mathfrak{H}} \\
&= \mathbb{E}[F_i F_j] - \langle D \sum_{p \geqslant 1} I_p(f_{i,p}), D \sum_{q \geqslant 1} \frac{1}{q} I_q(f_{j,q}) \rangle_{\mathfrak{H}} \\
&= \sum_{p,q \geqslant 1} \left(\mathbb{E}[I_p(f_{i,p}) I_q(f_{j,q})] - \frac{1}{q} \langle DI_p(f_{i,p}), DI_q(f_{j,q}) \rangle_{\mathfrak{H}} \right) =: \sum_{p,q \geqslant 1} \xi(i,j,p,q),
\end{aligned}$$

where it was also used that $\mathbb{E}\big[\sum_{p \geqslant 1} I_p(f_{i,p}) \sum_{q \geqslant 1} I_q(f_{j,q})\big] = \sum_{p,q \geqslant 1} \mathbb{E}\big[I_p(f_{i,p}) I_q(f_{j,q})\big]$ by linearity of the expectation. Now, as before, use Lemma 4.3 to obtain that

$$\max_{1 \leqslant i,j \leqslant d} \|\xi(i,j,p,q)\|_2 \leqslant \frac{\sqrt{q}(p!q!)^{1/2}}{\sqrt{p}} 2^{p+q} \max_{1 \leqslant i,j \leqslant d} \sum_{r=1}^{b(p,q)} \|f_{i,p} \otimes_r f_{j,q}\|_{\mathfrak{H}^{\odot(p+q-2r)}},$$

where $b(p,q) = p \wedge q - \mathbb{1}_{\{p=q\}}$, and where we used sub-additivity of the square root. Hence, as

$$\sqrt{\mathbb{E}[\|M_{\boldsymbol{F}} - \boldsymbol{\Sigma}\|_{\text{H.S.}}^2]} \leqslant \|\operatorname{Cov}(\boldsymbol{F}) - \boldsymbol{\Sigma}\|_{\text{H.S.}} + d \sum_{p,q} \max_{1 \leqslant i,j \leqslant d} \|\xi(i,j,p,q)\|_2,$$

we have obtained exactly that

$$\sqrt{\mathbb{E}[\|M_{\boldsymbol{F}}-\boldsymbol{\Sigma}\|^2_{\text{H.S.}}]} \leqslant \|\text{Cov}(\boldsymbol{F})-\boldsymbol{\Sigma}\|_{\text{H.S.}} + d\Psi(d), \tag{4.26}$$

where $\Psi(d)$ is defined in (4.12). Plugging (4.26) into (4.9), yields

$$d_{\mathscr{C}}(\boldsymbol{F},\boldsymbol{Z}) \leqslant 402(\|\boldsymbol{\Sigma}^{-1}\|_{\text{op}}^{3/2}+1)d^{41/24}\left(\|\text{Cov}(\boldsymbol{F})-\boldsymbol{\Sigma}\|_{\text{H.S.}}+d\Psi(d)\right).$$

As convex distance is invariant under affine transformations, similar to Remark 4.6, for Λ denoting correlation matrix related to $\boldsymbol{\Sigma}$, $\boldsymbol{D}$ denoting a diagonal matrix that has $\{\boldsymbol{\Sigma}_{ii}^{-1/2} : i \in \{1,\dots,d\}\}$ on its diagonal, it holds that

$$\begin{aligned} d_{\mathscr{C}}(\boldsymbol{F},\boldsymbol{Z}) = d_{\mathscr{C}}(\boldsymbol{DF},\boldsymbol{DZ}) &\leqslant 402(\|\boldsymbol{\Lambda}^{-1}\|_{\text{op}}^{3/2}+1)d^{41/24}\left(\|\text{Cov}(\boldsymbol{DF})-\boldsymbol{\Lambda}\|_{\text{H.S.}}+d\Psi(d)\right)\\ &\leqslant 402 d^{41/24}\left(\|\text{Cov}(\boldsymbol{DF})-\boldsymbol{\Lambda}\|_{\text{H.S.}}+d\Psi(d)\right)\left(\frac{1}{\sigma_*(\boldsymbol{\Lambda})}\right)^{3/2}, \end{aligned}$$

where we have used that $\|\boldsymbol{\Lambda}^{-1}\|_{\text{op}} = \sigma_*(\boldsymbol{\Lambda})^{-1}$, by the definition of the operator norm, and that then there exists a constant $c \geqslant 1$ such that $c/(\sigma_*(\boldsymbol{\Lambda}))^{3/2} \geqslant 1$ and therefore $c/(\sigma_*(\boldsymbol{\Lambda}))^{3/2}+1 \leqslant 2c/(\sigma_*(\boldsymbol{\Lambda}))^{3/2}$.

Part 3. Finally, for 1-Wasserstein distance recall (4.10). To obtain the wanted result, one has to bound $\mathbb{E}[\|M_{\boldsymbol{F}}-\boldsymbol{\Sigma}\|^2_{\text{H.S.}}]^{1/2}$, hence the result follows immediately from (4.26):

$$d_{\mathcal{W}}(\boldsymbol{F},\boldsymbol{Z}) \leqslant \sqrt{d}\left(\|\text{Cov}(\boldsymbol{F})-\boldsymbol{\Sigma}\|_{\text{H.S.}}+d\Psi(d)\right)\|\boldsymbol{\Sigma}^{-1}\|_{\text{op}}\|\boldsymbol{\Sigma}\|_{\text{op}}^{1/2}$$

Similar observation as in Part 2., namely from the definition of the operator norm it follows that $\|\boldsymbol{\Sigma}^{-1}\|_{\text{op}} = \sigma_*(\boldsymbol{\Sigma})^{-1}$, and $\|\boldsymbol{\Sigma}\|_{\text{op}} = \sigma^*(\boldsymbol{\Sigma})$. This concludes the proof.

□

Proof of Corollary 4.2. Begin by showing the claim for the hyper-rectangular distance (4.15). And then for convex and 1-Wasserstein. Theorem 4.1 yields that

$$\begin{aligned} d_{\mathcal{R}}(\boldsymbol{F},\boldsymbol{Z}) &\leqslant C\frac{\Delta_{\boldsymbol{F}}}{\sigma_*(\boldsymbol{\Lambda})}\log(d)\log_+\left(\frac{\Delta_{\boldsymbol{F}}}{\sigma_*}\right), \quad \text{where}\\ \Delta_{\boldsymbol{F}} &= \Phi(d) + \max_{1\leqslant j,k\leqslant d}\left|\boldsymbol{\Lambda}_{jk} - \frac{\mathbb{E}[F_jF_k]}{(\boldsymbol{\Sigma}_{jj}\boldsymbol{\Sigma}_{kk})^{1/2}}\right|, \end{aligned} \tag{4.27}$$

and where $\Phi(d)$ is defined in (4.11), and $C > 0$, independent of d. Hence we want to show that $\Phi(d)$ is finite under the Assumption (4.13). In fact we will show that

(1) If $\beta \in (1/2, 1]$, then

$$\Phi(d) \leqslant \gamma\frac{c}{\log(d)}e^{k\log^{1/2\beta}(d)}, \tag{4.28}$$

for a positive constant c and $k = \exp\left(2\beta e^{1/2e}(2^{5/2}e^{\alpha+1/2+1/e})^{1/\beta}\right)$.

(2) Assume $\beta = 1/2$ then for $\alpha < \alpha_0$ it holds that

$$\Phi(d) \leqslant \gamma c\frac{d^k}{\log(d)}, \tag{4.29}$$

for a positive constants c and $k = 16\exp(2\alpha + 1 + 5/(2e))$.

Equations (4.28) and (4.29) yield for $\beta \in [1/2, 1]$, that $\Phi(d) \leqslant \gamma c e^{k \log^{1/2\beta}(d)} / \log_+(d)$. Hence, plugging this into the definition of $\Delta_{\boldsymbol{F}}$ in (4.27), we obtain the wanted result.

Let us begin by looking at $\Xi(p,q,d) = \sum_{r=1}^{p\wedge q - \mathbb{1}_{\{p=q\}}} \max_{1\leqslant j,k,\leqslant d} \|f_{j,p} \otimes_r f_{k,q}\|_{\mathfrak{H}^{\otimes(p+q-2r)}}$, for $f_{k,p} \in \mathfrak{H}^{\odot p}$ for all $k \in \{1, \dots, d\}$ and $p \in \mathbb{N}$. Under the assumptions on the contractions (4.13), it holds that

$$\Xi(p,q,d) \leqslant \gamma(p \wedge q - \mathbb{1}_{\{p=q\}}) \frac{e^{\alpha p} e^{\alpha q}}{(p!q!)^\beta} \leqslant \gamma p \frac{e^{\alpha p} e^{\alpha q}}{(p!q!)^\beta}$$

and hence plugging into (4.11) we get

$$\Phi(d) \leqslant \gamma \sum_{\substack{p,q \geqslant 1 \\ (p,q) \neq (1,1)}} pq \frac{(4e^{1/2+\alpha})^{p+q}}{(p!q!)^\beta} \log^{(p+q)/2-1}(2d^2 + e^{(p+q)/2-2}). \tag{4.30}$$

As $2d^2$ and $e^{(p+q)/2-1}$ are both greater or equal to 1, when $d \geqslant 1$ and $p+q > 2$, then due to $a + b \leqslant 2ab$ for $a, b \geqslant 1$ we have

$$\begin{aligned} \log(2d^2 + e^{(p+q)/2-2}) \leqslant \log(4d^2 e^{(p+q)/2-2}) &\leqslant \log(e^2 d^2 e^{(p+q)/2-2}) \\ &= 2\log(d) + \frac{p+q}{2}, \end{aligned} \tag{4.31}$$

where in the second to last inequality we used that $2 < e$ and that log is an increasing function.

In (4.30), the summands are positive, and hence we can include the case $(p,q) = (1,1)$ in the sum, as it makes it bigger. Then, plugging (4.31) into (4.30), one obtains that $\Phi(d) \leqslant \gamma \Phi_1(d) + \gamma \Phi_2$, where

$$\begin{aligned} \Phi_1(d) &:= \sum_{p,q\geqslant 1} pq \frac{(4e^{1/2+\alpha})^{p+q}}{(p!q!)^\beta} (2\log(d))^{(p+q)/2-1}, \quad \text{and} \\ \Phi_2 &:= \sum_{p,q\geqslant 1} pq \frac{(4e^{1/2+\alpha})^{p+q}}{(p!q!)^\beta} \left(\frac{p+q}{2}\right)^{(p+q)/2-1}. \end{aligned} \tag{4.32}$$

Note that $[(p+q)/2]^{(p+q-2)/2}$ grows much faster than $[2\log(d)]^{(p+q-2)/2}$, hence if Φ_2 converges, so does $\Phi_1(d)$. Also remark that $\Phi_1(d)$ depends on the dimension d, while Φ_2 does not. For $p, q \geqslant 1$ we have

$$(2\log(d))^{(p+q)/2-1} = \frac{2^{(p+q)/2}}{2\log_+(d)} \log^{p/2}(d) \log^{q/2}(d), \qquad \text{and} \tag{4.33}$$

$$\left(\frac{p+q}{2}\right)^{(p+q-2)/2} \leqslant p^{(p+q-2)/2} + q^{(p+q-2)/2} = p^{p/2} p^{q/2} 2 + q^{p/2} q^{q/2}. \tag{4.34}$$

Plug (4.33) and (4.34) into $\Phi_1(d)$ and Φ_2 given in (4.32). Due to the symmetry in p and q

in obtained expressions, and the double sums in (4.32), one gets the following expressions

$$\Phi_1(d) = \frac{1}{2\log_+(d)} \left(\sum_{p \geqslant 1} p \frac{(2^{5/2} e^{1/2+\alpha})^p}{(p!)^\beta} \log^{p/2}(d) \right)^2, \quad \text{and} \tag{4.35}$$

$$\begin{aligned} \Phi_2 &= \sum_{p \geqslant 1} \frac{p(4e^{1/2+\alpha})^p}{(p!)^\beta} \sum_{q \geqslant 1} \frac{q(4e^{1/2+\alpha})^q}{(q!)^\beta} \left(p^{p/2} p^{q/2} + q^{p/2} q^{q/2} \right) \\ &= 2 \sum_{p \geqslant 1} \frac{p(4e^{1/2+\alpha})^p}{(p!)^\beta} p^{p/2} \sum_{q \geqslant 1} \frac{q(4e^{1/2+\alpha})^q}{(q!)^\beta} p^{q/2}. \end{aligned} \tag{4.36}$$

Applying Stirling's inequality (4.18) to bound p^p in (4.36) with $p!$, i.e. $p^p \leqslant (p! e^p)/\sqrt{2\pi p}$, we get

$$\Phi_2 \leqslant \frac{2}{(2\pi)^{1/4}} \sum_{p \geqslant 1} \frac{p^{3/4}(4e^{1+\alpha})^p}{(p!)^{\beta-1/2}} \sum_{q \geqslant 1} \frac{q(4e^{1/2+\alpha}\sqrt{p})^q}{(q!)^\beta}. \tag{4.37}$$

Next we want to prove equations (4.28) and (4.29).

Part 1. Assume that $1/2 < \beta \leqslant 1$. First, focus on $\Phi_1(d)$. For $\alpha > 0$ it holds that $q^\alpha = \left(e^{\alpha \log q/q}\right)^q$. Note that $(\log q)/q$ obtains maximum at $q = e$ with value $1/e$, hence

$$q^\alpha = \left(e^{\alpha \log q/q}\right)^q \leqslant \left(e^{\alpha/e}\right)^q. \tag{4.38}$$

By applying (4.38) with $\alpha = 1$ together with Lemma 4.5 for $\varpi = \beta$, using that $2\varpi = 2\beta > 1$, and combining terms with identical powers, we obtain

$$\Phi_1(d) \leqslant \frac{\gamma}{2\log_+(d)} \left(\sum_{p \geqslant 0} \frac{(2^{5/2} e^{\alpha+1/2+\frac{1}{e}} \log^{1/2}(d))^p}{(p!)^\beta} \right)^2 \leqslant \frac{\gamma c_2}{\log_+(d)} e^{k_1 \log^{1/(2\beta)}(d)}, \tag{4.39}$$

where $c_2 = c_2(\beta) = 2^{4\beta-1}(\pi e^{1/3}/2)^\beta$ and $k_1 = k_1(\alpha, \beta) = \exp\left(2\beta e^{1/2e}(2^{5/2} e^{\alpha+1/2+1/e})^{1/\beta}\right)$.

It remains to check that Φ_2, in (4.37) converges. Applying (4.38) with $\alpha = 3/4$ in the sum over p and $\alpha = 1$ in the sum over q, together with Lemma 4.5 for $\varpi = \beta$, we arrive at the following expression. In doing so, we use that $2\beta \in (1, 2]$, which implies $p^{1/(2\beta)} \leqslant p$ for $p \geqslant 1$, and that adding the term for $p = 0$ increases the sum since all terms are positive:

$$\begin{aligned} \Phi_2 &\leqslant \frac{2\gamma}{(2\pi)^{1/4}} \sum_{p \geqslant 1} \frac{(4e^{1+\alpha+3/(4e)})^p}{(p!)^{\beta-1/2}} \sum_{q \geqslant 1} \frac{(4e^{1/2+1/e+\alpha} p^{1/2})^q}{(q!)^\beta} \leqslant c_3\gamma \sum_{p \geqslant 1} \frac{(4e^{1+\alpha+3/(4e)})^p}{(p!)^{\beta-1/2}} e^{k_2 p^{1/(2\beta)}} \\ &\leqslant c_3\gamma \sum_{p \geqslant 1} \frac{(4e^{1+\alpha+3/(4e)+k_2})^p}{(p!)^{\beta-1/2}} = c_3\gamma \sum_{p \geqslant 0} \frac{(4e^{1+\alpha+3/(4e)+k_2})^p}{(p!)^{\beta-1/2}} < \infty, \end{aligned} \tag{4.40}$$

where $c_3 = c_3(\beta) = 2^{2\beta+5/4}/\pi^{1/4}(\pi e^{1/3}/2)^{\beta/2}$ and $k_2 = \exp\left(\beta e^{1/(2e)}(4e^{1/2+1/e+\alpha})^{1/\beta}\right)$. The series in (4.40) converges since $\beta - 1/2 > 0$. Indeed, the ratio test yields $\sum_{p \geqslant 0} |a_p| < \infty$ when $a_{p+1}/a_p \to 0$ as $p \to \infty$. As for $\sum_{p \geqslant 0} a_p$ with $a_p = (Ae^B)^p (p!)^{-\varepsilon}$ and $\varepsilon > 0$, where $A = 4$ and $B = 1 + \alpha + 3/(4e) + k_2$, it holds that,

$$\lim_{p \to \infty} \frac{a_{p+1}}{a_p} = \lim_{p \to \infty} \frac{Ae^B}{(p+1)^\varepsilon} = 0,$$

it implies the claimed finiteness in (4.40).

Recall that $k_1 = \exp\big(2\beta e^{1/(2e)}(2^{5/2}e^{\alpha+1/2+1/e})^{1/\beta}\big)$, where $\beta \in (1/2, 1]$ and $\alpha \in \mathbb{R}$. Note that $k_1 \to 1$ as $\alpha \to -\infty$, which implies $k_1 \geqslant 1$. Consequently,

$$\frac{\exp\big(k_1 \log^{1/(2\beta)} d\big)}{\log_+ d} \geqslant \frac{e^{\log^{1/2} d}}{\log(d)} > 1, \quad d > 1,$$

where the first inequality holds for $d \geqslant e$ (since $1/(2\beta) \geqslant 1/2$) and the second follows because $e^{\sqrt{x}} > x$ for all $x > 0$. For $1 < d < e$, the ratio is trivially large, and for $d = 1$, it equals 1. Thus, the ratio is bounded below by 1 for all $d \geqslant 1$.

Combining (4.39) and (4.40), and using that $\exp(k_1 \log_+^{1/(2\beta)}(d))/\log_+(d) \geqslant 1$ for $d \geqslant 1$, we obtain that for $\beta \in (1/2, 1]$ and finite constants c_4 and $c_5 := c_2 + c_4$,

$$\Phi(d) \leqslant \Phi_1(d) + \Phi_2 \leqslant \frac{\gamma c_2}{\log_+(d)} e^{k_1 \log^{1/2\beta}(d)} + \gamma c_4 \leqslant \gamma \frac{c_5}{\log_+(d)} e^{k_1 \log^{1/2\beta}(d)},$$

which proves (4.28).

Part 2. Let $\beta = 1/2$. Proceeding from (4.35) and (4.36), we first compute $\Phi_1(d)$ to extract the explicit dimension dependence, and then verify convergence of Φ_2. As in Part 1., using (4.93) we calculate $\Phi_1(d)$ in (4.35) as follows:

$$\Phi_1(d) \leqslant \frac{\gamma}{2\log_+(d)} \left(\sum_{p \geqslant 0} \frac{(2^{5/2}e^{\alpha+1/2+1/e} \log^{1/2}(d))^p}{(p!)^{1/2}} \right)^2 \leqslant \frac{\gamma c_6}{\log_+(d)} d^{k_3}, \tag{4.41}$$

where $c_6 = \sqrt{2\pi e^{1/3}}$ and $k_3 = 2^4 \exp(2\alpha + 1 + 5/(2e))$.

It remains to show that Φ_2 in (4.36) converges. Calculate, using (4.93) in the second sum of (4.36) as follows:

$$\begin{aligned} \Phi_2 &\leqslant \frac{2\gamma}{(2\pi)^{1/4}} \sum_{p \geqslant 1} (4e^{1+3/(4e)+\alpha})^p \sum_{q \geqslant 1} \frac{\big(16e^{1+2/e+2\alpha}p\big)^{q/2}}{(q!)^{1/2}} \\ &\leqslant c_7\gamma \sum_{p \geqslant 1} (4e^{1+3/(4e)+\alpha})^p e^{8e^{1+5/(2e)+2\alpha}p} = c_7\gamma \sum_{p \geqslant 1} (4e^{1+3/(4e)+\alpha+8e^{1+5/(2e)+2\alpha}})^p \\ &\leqslant c_7\gamma \sum_{p \geqslant 0} (4e^{1+3/(4e)+\alpha+8e^{1+5/(2e)+2\alpha}})^p, \end{aligned} \tag{4.42}$$

where $c_7 = 2^{3/4}e^{1/12}$. Note that $\sum_{p \geqslant 0} (4\exp(1 + 3/(4e) + \alpha + 8e^{1+5/(2e)+2\alpha}))^p$ is a geometric series and it converges if $x := 4\exp(1 + 3/(4e) + \alpha + 8e^{1+5/(2e)+2\alpha}) < 1$. To that end, let $d_1 = 4e^{1+3/(4e)}$ and $d_2 = 8e^{1+5/(2e)}$. Then (4.42) converges for those α such that $d_1 e^{\alpha} \exp(d_2 e^{2\alpha}) < 1$. Hence we find α such that,

$$d_1^2 e^{2\alpha} e^{2d_2 e^{2\alpha}} < 1 \quad \Longleftrightarrow \quad 2d_2 e^{2\alpha} e^{2d_2 e^{2\alpha}} < \frac{2d_2}{d_1^2} = e^{1/e-1}.$$

Using the Lambert W-function we obtain that $2d_2 e^{2\alpha} < W(e^{1/e-1})$ and hence, $\alpha < 1/2 \log(W(e^{1/e-1})/(2d_2)) \approx -2.846$. For such α, the sum $\sum_{p \geqslant 0} x^p$ converges. Since $d^{k_3} \geqslant c \log_+(d)$ for all $d \geqslant 1$ and some constant $c = c(k_3) > 0$, we have for finite constants $c_8 > 0$ and $c_9 = c_2 + c_4 > 0$ that

$$\Phi(d) \leqslant \gamma c_6 \frac{d^{k_3}}{\log_+(d)} + \gamma c_8 \leqslant \gamma c_9 \frac{d^{k_3}}{\log_+(d)}, \tag{4.43}$$

which establishes (4.29).

Convex and 1-Wasserstein distance. Proving (4.16) and (4.17) proceeds by using similar a toolbox to the proof for hyper-rectangular distance. Recall statements in Theorem 4.1 for $d_{\mathscr{C}}$ and $d_{\mathcal{W}}$: we have to ensure that $\Psi(d)$, defined in (4.12), is finite under the assumption (4.13). Apply assumption (4.13) to (4.12) to obtain

$$\Psi(d) \leqslant \gamma \sum_{\substack{p,q\geqslant 1\\(p,q)\neq(1,1)}} \sqrt{pq}\frac{e^{\alpha(p+q)}2^{p+q}}{(p!q!)^{\beta-1/2}} \leqslant \gamma \sum_{p,q\geqslant 1} \sqrt{pq}\frac{e^{\alpha(p+q)}2^{p+q}}{(p!q!)^{\beta-1/2}} =: \gamma\Psi, \tag{4.44}$$

where we omitted the condition $(p,q)\neq(1,1)$ in the summation since removing it adds a positive term to the sum, thereby increasing the bound. To conclude the proof, it suffices to show that $\Psi<\infty$. For the remainder of the proof, we split it into two cases, when $\beta\in(1/2,1]$ and when $\beta=1/2$.

Assume first that $\beta=1/2$. Applying (4.38) to (4.44) implies $\Psi<\infty$. Indeed, we observe that

$$\sum_{p,q\geqslant 1}\sqrt{pq}e^{\alpha(p+q)}2^{p+q} \leqslant \sum_{p,q\geqslant 1} e^{(p+q)/(2e)}e^{\alpha(p+q)}2^{p+q} = \left(\sum_{p\geqslant 1}\left(2e^{\alpha+1/(2e)}\right)^p\right)^2.$$

This expression is finite because the inner sum is a geometric series that converges provided $2e^{\alpha+1/(2e)}<1$, or equivalently, $\alpha<\log(1/2)-1/(2e)\approx -0.877$.

Assume $\beta\in(1/2,1]$ and let $\varepsilon:=\beta-1/2>0$. Then by ratio test, Ψ in (4.44) is finite. Indeed

$$\sum_{p,q\geqslant 1}\frac{e^{(\alpha+1/(2e))(p+q)}2^{p+q}}{(p!q!)^{\varepsilon}} = \left(\sum_{p\geqslant 1}\frac{(2e^{\alpha+1/(2e)})^p}{(p!)^{\varepsilon}}\right)^2<\infty,$$

as for $a_p=(2e^{\alpha+1/(2e)})^p(p!)^{-\varepsilon}$, $a_{p+1}/a_p\to 0$ for $p\to\infty$. This concludes the proof. □

4.4 Proof of Theorem 2.1

Theorem 2.1 can be obtained from Corollary 4.2 as follows.

Proof of Theorem 2.1. Applying the additional assumption, that $\boldsymbol{F}$ and $\boldsymbol{Z}$ have the same covariance matrix, to Corollary 4.2 yields the result. Indeed, using that $\boldsymbol{\Sigma}_{jk}=\mathbb{E}[F_jF_k]$, it follows by definition of $\Delta_{\boldsymbol{F}}$ in (4.14), that

$$\Delta_{\boldsymbol{F}} = \frac{\gamma}{\log_+(d)}e^{k\log_+^{1/2\beta}(d)}. \tag{4.45}$$

Then, for $d\geqslant 1$, plugging (4.45) into (4.15), and using that $\log_+(xy)=\log_+(x)+\log_+(y)\leqslant 2\log_+(x)\log_+(y)$ for all $x,y>0$, one gets

$$\begin{aligned}
d_{\mathcal{R}}(\boldsymbol{F},\boldsymbol{Z}) &\leqslant C\gamma\frac{\exp\left(k\log^{1/(2\beta)}(d)\right)}{\log(d)\sigma_*}\log(d)\log_+\left(\frac{\overline{\sigma}}{\overline{\sigma}\sigma_*}\right)\log_+\left(\gamma\frac{\exp\left(k\log^{1/(2\beta)}(d)\right)}{\log(d)}\right)\\
&\leqslant C_1\gamma\log_+(\gamma)\frac{\log_+\left(\frac{\overline{\sigma}}{\overline{\sigma}\sigma_*}\right)}{\sigma_*}\exp\left(k\log^{1/(2\beta)}(d)\right)\left(k\log^{1/(2\beta)}(d)-\log_+\log(d)\right)\\
&\leqslant C_2\gamma\log_+(\gamma)\frac{\log_+\left(\frac{\overline{\sigma}}{\overline{\sigma}\sigma_*}\right)}{\sigma_*}\exp\left(k\log^{1/(2\beta)}(d)\right)\log^{1/(2\beta)}(d),
\end{aligned}$$

where $C_1 = 4C$ and $C_2 = kC_1$.

Convex and 1-Wasserstein distance. Under the assumption that $\boldsymbol{\Sigma}_{ij} = \mathbb{E}[F_i F_j]$ for all $i, j \in \{1, \ldots, d\}$ (i.e., the covariances of $\boldsymbol{F}$ and $\boldsymbol{Z}$ coincide), the result follows immediately from (4.16) and (4.17), since the terms involving the difference of the covariance matrices vanish. ☐

4.5 Proof of Quantitative Breuer–Major

Proof of Theorem 2.3. Recall that

$$\boldsymbol{S}_n = (S_n^i, \ldots, S_n^d) := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \Phi(\boldsymbol{G}_k) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (\varphi_1(\boldsymbol{G}_k), \ldots, \varphi_d(\boldsymbol{G}_k)).$$

The idea is to write $S_n^{(i)}$ for all $i \in \{1, \ldots, d\}$ in a multiple integral form and then use Theorem 4.1. Recall the multivariate Hermite expansion (4.2) for functions in $L^2(\gamma^K, \mathbb{R}^K)$. Since $\mathbb{E}[|\varphi_i(\boldsymbol{G}_1)|^2] < \infty$, the function φ_i admits a Hermite expansion. By (4.3), we can write

$$S_n^{(i)} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \varphi_i(\boldsymbol{G}_k) = n^{-1/2} \sum_{k=1}^{n} \sum_{q=m_i}^{\infty} \varphi_{i,q}(\boldsymbol{G}_k), \quad \text{where}$$
$$\varphi_{i,q}(\boldsymbol{x}) := \sum_{\boldsymbol{\alpha} \in \Lambda : |\boldsymbol{\alpha}| = q} a_{i,\boldsymbol{\alpha}} \prod_{j=1}^{K} H_{\alpha^{(j)}}(x^{(j)}) \quad \text{and} \quad a_{i,\boldsymbol{\alpha}} := \frac{1}{\boldsymbol{\alpha}!} \mathbb{E}\left[\varphi_i(\boldsymbol{G}_1) \prod_{j=1}^{K} H_{\alpha^{(j)}}(\boldsymbol{G}_1^{(j)})\right],$$

with $\Lambda = \{\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K) : \alpha_j \in \mathbb{N}_0\}$, $|\boldsymbol{\alpha}| = \sum_{j=1}^{K} \alpha_j$, and $\boldsymbol{\alpha}! = \prod_{j=1}^{K} \alpha_j!$.

Recall, $(\boldsymbol{G}_k)_{k \in \mathbb{Z}}$ is a K-dimensional centred stationary Gaussian sequence. Therefore, by [39, Rem. 2.1.5] there exists a separable Hilbert space $\mathfrak{H}$ and an isonormal Gaussian process $\{X(h) : h \in \mathfrak{H}\}$ such that $\boldsymbol{G}_k$ can be represented as $\boldsymbol{G}_k \stackrel{\mathcal{D}}{=} (X(e_{k,1}), \ldots, X(e_{k,K}))$ for a sequence of vectors $(e_{k,\ell})_{k \in \mathbb{Z}, \ell \in \{1,\ldots,K\}} \subset \mathfrak{H}$ satisfying $\langle e_{k,\ell}, e_{k',\ell'} \rangle_{\mathfrak{H}} = \mathrm{Cov}(\boldsymbol{G}_k^{(\ell)}, \boldsymbol{G}_{k'}^{(\ell')}) = \rho^{(\ell,\ell')}(k - k')$ for all $k, k' \in \mathbb{Z}$ and $\ell, \ell' \in \{1, \ldots, K\}$. Hence, using [39, Thm 2.7.7], it holds that

$$\prod_{j=1}^{K} H_{\alpha^j}(\boldsymbol{G}_k^{(j)}) \stackrel{\mathcal{D}}{=} \prod_{j=1}^{K} H_{\alpha^j}(X(e_{k,j})) = I_q\left(\frac{\boldsymbol{\alpha}!}{q!} \sum_{\sigma \in S_q^{\boldsymbol{\alpha}}} \sigma(e_{k,1}^{\otimes \alpha^{(1)}} \otimes \cdots \otimes e_{k,K}^{\otimes \alpha^{(K)}})\right),$$

where $S_q^{\boldsymbol{\alpha}}$ denotes the q-permutations on the multiset $\{1, \ldots, 1, \ldots, K, \ldots, K\}$ where j appears with multiplicity $\alpha^{(j)}$ for all $j \in \{1, \ldots, K\}$. Hence, due to linearity of the multiple integrals,

$$\begin{aligned} \varphi_{i,q}(\boldsymbol{G}_k) &= \sum_{\boldsymbol{\alpha} \in \Lambda : |\boldsymbol{\alpha}| = q} a_{i,\boldsymbol{\alpha}} I_q\left(\frac{\boldsymbol{\alpha}!}{q!} \sum_{\sigma \in S_q^{\boldsymbol{\alpha}}} \sigma(e_{k,1}^{\otimes \alpha^{(1)}} \otimes \cdots \otimes e_{k,K}^{\otimes \alpha^{(K)}})\right) \\ &= I_q\left(\sum_{\boldsymbol{\alpha} \in \Lambda : |\boldsymbol{\alpha}| = q} a_{i,\boldsymbol{\alpha}} \frac{\boldsymbol{\alpha}!}{q!} \sum_{\sigma \in S_q^{\boldsymbol{\alpha}}} \sigma(e_{k,1}^{\otimes \alpha^{(1)}} \otimes \cdots \otimes e_{k,K}^{\otimes \alpha^{(K)}})\right). \end{aligned}$$

Thus, we obtain [41, Eq. (4.43)], for $i \in \{1, \ldots, d\}$

$$S_n^{(i)} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{q=m_i}^{\infty} \varphi_{i,q}(\boldsymbol{G}_k) = \sum_{q=m_i}^{\infty} I_q\Bigg(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{\substack{\boldsymbol{\alpha}\in\Lambda \\ |\boldsymbol{\alpha}|=q}} a_{i,\boldsymbol{\alpha}} \frac{\boldsymbol{\alpha}!}{q!} \sum_{\sigma \in S_q^{\boldsymbol{\alpha}}} \sigma\big(e_{k,1}^{\otimes\alpha^{(1)}} \otimes \cdots \otimes e_{k,K}^{\otimes\alpha^{(K)}}\big)\Bigg)$$

$$= \sum_{q=m_i}^{\infty} I_q(f_n(i,q)), \text{ where } f_n(i,q) := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \sum_{t\in\{1,\ldots,K\}^q} b_{i,t} e_{k,t^{(1)}} \otimes \cdots \otimes e_{k,t^{(q)}},$$

with $f_n(i,q) \in \mathfrak{H}^{\odot q}$, and where the coefficients $b_{i,t} = (q!)^{-1} a_{\boldsymbol{\alpha}(t)}(\boldsymbol{\alpha}(t))!$, for $\boldsymbol{\alpha}(t) = \sum_{\ell=1}^{q} (\mathbb{1}_{\{t^{(\ell)}=1\}}, \ldots, \mathbb{1}_{\{t^{(\ell)}=K\}}) \in \mathbb{N}_0^K$. Let $[K]$ denote the set $\{1, \ldots, K\}$. Using the isometry property of multiple integrals (4.4), and the fact that $\rho^{(i,j)}(0) = \mathbb{1}_{\{i=j\}}$ for $i, j \in \{1, \ldots, K\}$, one obtains

$$\begin{aligned}
\mathbb{E}[\varphi_{i,q}(\boldsymbol{G}_1)^2] &= \mathbb{E}\Big[I_q\Big(\sum_{t\in[K]^q} b_{i,t} e_{1,t^{(1)}} \otimes \cdots \otimes e_{1,t^{(q)}} \Big)^2 \Big] \\
&= q! \sum_{s,t\in[K]^q} \big\langle b_{i,t} e_{1,t^{(1)}} \otimes \cdots \otimes e_{1,s^{(q)}}, b_{i,s} e_{1,s^{(1)}} \otimes \cdots \otimes e_{1,s^{(q)}} \big\rangle_{\mathfrak{H}^{\otimes q}} \qquad (4.46) \\
&= q! \sum_{s,t\in[K]^q} b_{i,t} b_{i,s} \prod_{k=1}^{q} \langle e_{1,t^{(k)}}, e_{1,s^{(k)}} \rangle_{\mathfrak{H}} = q! \sum_{t\in[K]^q} b_{i,t}^2.
\end{aligned}$$

Hence it follows that $\mathbb{E}[\varphi_i(\boldsymbol{G}_1)^2] = \sum_{q=m_i}^{\infty} q! \sum_{t\in\{1,\ldots,K\}^q} b_{i,t}^2$.

In order to use Theorem 4.1 for random vectors $\boldsymbol{S}_n$ and $\boldsymbol{Z}_n$, it remains to bound $\|f_n(i,q) \otimes_r f_n(j,p)\|_{\mathfrak{H}^{\otimes(p+q-2r)}}$ for every $i, j \in \{1, \ldots, d\}$, $p, q \in \mathbb{N}$ not simultaneously 1, and $r \in \{1, \ldots, p \wedge q - \mathbb{1}_{\{p=q\}}\}$. Some of the following steps are inspired by mainly [41, Sec. 4]. To that end, calculate

$$\begin{aligned}
f_n(i,q) \otimes_r f_n(j,p) &= \frac{1}{n} \sum_{t\in[K]^q} \sum_{s\in[K]^p} b_{t,i} b_{s,j} \sum_{k,\ell=1}^{n} \prod_{z=1}^{r} \langle e_{k,t^{(z)}}, e_{\ell,s^{(z)}} \rangle_{\mathfrak{H}} e_{k,t^{(r+1)}} \otimes \cdots \otimes e_{\ell,s^{(p)}} \\
&= \frac{1}{n} \sum_{t\in[K]^q} \sum_{s\in[K]^p} b_{t,i} b_{s,j} \sum_{k,\ell=1}^{n} \prod_{z=1}^{r} \rho^{(t^{(z)},s^{(z)})}(k-\ell) e_{k,t^{(r+1)}} \otimes \cdots \otimes e_{\ell,s^{(p)}},
\end{aligned}$$

and the $\mathfrak{H}^{\otimes(p+q-2r)}$-norm of $f_n(i,q) \otimes_r f_n(j,p)$, as follows:

$$\begin{aligned}
\|f_n(i,q) \otimes_r f_n(j,p)\|_{\mathfrak{H}^{\otimes(p+q-2r)}}^2 &= \frac{1}{n^2} \sum_{t,t'\in[K]^q} \sum_{s,s'\in[K]^p} |b_{t,i} b_{t',i} b_{s,j} b_{s',j}| \\
&\quad \times \sum_{k,l,k',\ell'=1}^{n} \prod_{z=1}^{r} \rho^{(t^{(z)},s^{(z)})}(k-\ell) \rho^{(t'^{(z)},s'^{(z)})}(k'-\ell') \\
&\quad \times \prod_{m=r+1}^{q} \rho^{(t^{(m)},t'^{(m)})}(k-k') \prod_{h=r+1}^{p} \rho^{(s^{(h)},s'^{(h)})}(\ell-\ell') \qquad (4.47) \\
&\leqslant \frac{1}{n^2} \sum_{t,t'\in[K]^q} \sum_{s,s'\in[K]^p} |b_{t,i} b_{t',i} b_{s,j} b_{s',j}| \\
&\quad \times \sum_{k,l,k',\ell'=1}^{n} |\overline{\rho}(k-\ell)^r \overline{\rho}(k'-\ell')^r \overline{\rho}(k-k')^{q-r} \overline{\rho}(\ell-\ell')^{p-r}|.
\end{aligned}$$

Using Jensen's inequality for finite sums and the convex function $f(x) = x^2$, together with (4.46), it follows that

$$\sum_{t\in[K]^q} b_{t,i} \sum_{t'\in[K]^q} b_{t',i} = \Big(\sum_{t\in[K]^q} b_{t,i}\Big)^2 = K^{2q}\Big(\frac{\sum_{t\in[K]^q} b_{t,i}}{K^q}\Big)^2 \leqslant K^q \sum_{t\in[K]^q} b_{t,i}^2$$
$$= \frac{K^q}{q!}\mathbb{E}[\varphi_{i,q}(\boldsymbol{G}_1)^2].$$

Hence, we conclude that the first sum in (4.47) satisfies

$$\sum_{t,t'\in[K]^q}\sum_{s,s'\in[K]^p} |b_{t,i}b_{t',i}b_{s,j}b_{s',j}| \leqslant \frac{K^qK^p}{p!q!}\mathbb{E}[\varphi_{i,q}(\boldsymbol{G}_1)^2]\mathbb{E}[\varphi_{j,p}(\boldsymbol{G}_1)^2]. \tag{4.48}$$

For the second part of the (4.47), recall that $r \in \{1,\dots,p\wedge q - \mathbb{1}_{\{p=q\}}\}$. As in [41, Proof of 4.44], note that $\overline{\rho}(k'-\ell')^r\overline{\rho}(k-k')^{q-r} \leqslant \overline{\rho}(k'-\ell')^q + \overline{\rho}(k-k')^q$. Hence,

$$\begin{aligned}
&\sum_{k,\ell,k',\ell'=1}^{n} |\overline{\rho}(k-\ell)^r\overline{\rho}(\ell-\ell')^{p-r}(\overline{\rho}(k'-\ell')^q+\overline{\rho}(k-k')^q)| \\
&\leqslant \sum_{k,\ell,\ell'=1}^{n}\Big|\overline{\rho}(k-\ell)^r\overline{\rho}(\ell-\ell')^{p-r}\Big(\sum_{\alpha=1-\ell'}^{n-\ell'}\overline{\rho}(\alpha)^q + \sum_{\alpha=1-k}^{n-k}\overline{\rho}(\alpha)^q\Big)\Big| \\
&\leqslant \sum_{k,\ell,\ell'=1}^{n}|\overline{\rho}(k-\ell)|^r|\overline{\rho}(\ell-\ell')|^{p-r}\Big(\sum_{|\alpha|<n}|\overline{\rho}(\alpha)|^q + \sum_{|\alpha|<n}|\overline{\rho}(\alpha)|^q\Big) \\
&\leqslant 2\sum_{|\alpha|<n}|\overline{\rho}(\alpha)|^q\sum_{k,\ell,\ell'=1}^{n}|\overline{\rho}(k-\ell)|^r|\overline{\rho}(\ell-\ell')|^{p-r}.
\end{aligned} \tag{4.49}$$

As $q \geqslant m$ and $|\overline{\rho}(x)| \leqslant 1$, we use that $|\overline{\rho}(\alpha)|^q \leqslant |\overline{\rho}(\alpha)|^m$ to further obtain that

$$\begin{aligned}
&\sum_{k,\ell,k',\ell'=1}^{n} |\overline{\rho}(k-\ell)^r\overline{\rho}(\ell-\ell')^{p-r}(\overline{\rho}(k'-\ell')^q+\overline{\rho}(k-k')^q)| \\
&\leqslant 2\sum_{|\alpha|<n}|\overline{\rho}(\alpha)|^m\sum_{\ell=1}^{n}\sum_{t=1-\ell}^{n-\ell}|\overline{\rho}(t)|^r\sum_{s=1-\ell}^{n-\ell}|\overline{\rho}(s)|^{p-r} \\
&\leqslant 2n\sum_{|\alpha|<n}|\overline{\rho}(\alpha)|^q\sum_{|t|<n}|\overline{\rho}(t)|^r\sum_{|s|<n}|\overline{\rho}(s)|^{p-r} \\
&\leqslant 2n\sum_{k\in\mathbb{Z}}|\overline{\rho}(k)|^m\sum_{|t|<n}|\overline{\rho}(t)|^r\sum_{|s|<n}|\overline{\rho}(s)|^{p-r}.
\end{aligned} \tag{4.50}$$

Recall that the starting point for the argument (4.49)-(4.50) is the inequality $\overline{\rho}(k'-\ell')^r\overline{\rho}(k-k')^{q-r} \leqslant \overline{\rho}(k'-\ell')^q + \overline{\rho}(k-k')^q$, hence use the same arguments (4.49)-(4.50) for $\overline{\rho}(k-\ell)^r\overline{\rho}(\ell-\ell')^{p-r} \leqslant \overline{\rho}(k-\ell)^p + \overline{\rho}(\ell-\ell')^p$ and obtain

$$\begin{aligned}
&\sum_{k,\ell,k',\ell'=1}^{n} |\overline{\rho}(k-\ell)^r\overline{\rho}(\ell-\ell')^{p-r}(\overline{\rho}(k'-\ell')^q+\overline{\rho}(k-k')^q)| \\
&\qquad\qquad \leqslant 2n\sum_{k\in\mathbb{Z}}|\overline{\rho}(k)|^m\sum_{|t|<n}|\overline{\rho}(t)|^r\sum_{|s|<n}|\overline{\rho}(s)|^{q-r}.
\end{aligned}$$

Combining the two displays above, it follows that

$$\sum_{k,\ell,k',\ell'=1}^{n} |\overline{\rho}(k-\ell)^r \overline{\rho}(\ell-\ell')^{p-r}(\overline{\rho}(k'-\ell')^q + \overline{\rho}(k-k')^q)|$$
$$\leqslant 2n \sum_{k\in\mathbb{Z}} |\overline{\rho}(k)|^m \sum_{|t|<n} |\overline{\rho}(t)|^r \sum_{|s|<n} |\overline{\rho}(s)|^{p\wedge q-r}. \quad (4.51)$$

Finally combining (4.51) with (4.48) we get that

$$\|f_n(i,q)\otimes_r f_n(j,p)\|_{\mathfrak{H}^{\otimes(p+q-2r)}} \leqslant \frac{K^{q/2}K^{p/2}}{\sqrt{p!q!}} \|\varphi_{i,q}(\boldsymbol{G}_1)\|_{L^2}\|\varphi_{j,p}(\boldsymbol{G}_1)\|_{L^2}\gamma_{n,p,q,r}, \quad (4.52)$$

where

$$\gamma_{n,p,q,r} := \sqrt{\frac{2}{n}\sum_{k\in\mathbb{Z}} |\overline{\rho}(k)|^m \sum_{|t|<n} |\overline{\rho}(t)|^r \sum_{|s|<n} |\overline{\rho}(s)|^{p\wedge q-r}}.$$

Using the inequality in (4.52), together with (4.11) and (4.12), implies

$$\Xi(p,q,d) \leqslant \frac{K^{q/2}K^{p/2}}{\sqrt{p!q!}} \max_{1\leqslant i,j\leqslant d} \|\varphi_{i,q}(\boldsymbol{G}_1)\|_{L^2}\|\varphi_{j,p}(\boldsymbol{G}_1)\|_{L^2} \sum_{r=1}^{p\wedge q-\mathbb{1}_{\{p=q\}}} \gamma_{n,p,q,r}.$$

Hence, Theorem 4.1 gives that

$$d_{\mathcal{R}}(\boldsymbol{S}_n,\boldsymbol{Z}_n) \leqslant C\log(d)\Delta_1(n,d,K)\log_+(\Delta_1(n,d,K))\frac{\log_+(\sigma_*)}{\sigma_*},$$
$$d_{\mathscr{C}}(\boldsymbol{S}_n,\boldsymbol{Z}_n) \leqslant 402\left(\frac{1}{\sigma_*}\right)^{3/2} d^{65/24}\Delta_2(n,d,K), \quad \text{and}$$
$$d_{\mathcal{W}}(\boldsymbol{S}_n,\boldsymbol{Z}_n) \leqslant \|\boldsymbol{\Sigma}^{-1}\|_{\mathrm{op}}\|\boldsymbol{\Sigma}\|_{\mathrm{op}}^{1/2} d^{3/2}\Delta_2(n,d,K),$$

where (4.11) and (4.12) give that

$$\Delta_1(n,d,K) := \sum_{\substack{p,q\geqslant 1\\(p,q)\neq(1,1)}} pq\frac{(4\sqrt{eK})^{p+q}}{\sqrt{p!q!}} \log^{(p+q)/2-1}(2d^2+e^{(p+q)/2-2})$$
$$\times \max_{1\leqslant i,j\leqslant d} \mathbb{E}[\varphi_{i,q}(\boldsymbol{G}_1)^2]^{1/2}\mathbb{E}[\varphi_{j,p}(\boldsymbol{G}_1)^2]^{1/2} \sum_{r=1}^{p\wedge q-\mathbb{1}_{\{p=q\}}} \gamma_{n,p,q,r}, \quad \text{and}$$
$$\Delta_2(n,d,K) := \sum_{\substack{p,q\geqslant 1\\(p,q)\neq(1,1)}} \frac{\sqrt{p}(2\sqrt{K})^{p+q}}{\sqrt{q}} \max_{1\leqslant i,j\leqslant d} \|\varphi_{i,q}(\boldsymbol{G}_1)\|_{L^2}\|\varphi_{j,p}(\boldsymbol{G}_1)\|_{L^2} \sum_{r=1}^{p\wedge q-\mathbb{1}_{\{p=q\}}} \gamma_{n,p,q,r}.$$

□

Proof of Corollary 2.4. We begin by observations for the hyper-rectangular distance. Plugging in the bound (2.6) into $\Delta_1(n,d,K)$, we get

$$\Delta_1(n,d,K) \leqslant \sum_{p,q\geqslant m} pq\frac{(4\sqrt{eK})^{p+q}}{\sqrt{p!q!}} \log^{(p+q)/2-1}(2d^2+e^{\frac{p+q}{2}-2})\frac{e^{\alpha(p+q)}}{(p!q!)^{\beta-1/2}} \sum_{r=1}^{p\wedge q-\mathbb{1}_{\{p=q\}}} \gamma_{n,p,q,r}$$
$$\leqslant \sum_{p,q\geqslant m} pq\frac{(4e^{1/2+\alpha})^{p+q}\sqrt{K}^{p+q}}{(p!q!)^{\beta}} \log^{(p+q)/2-1}(2d^2+e^{\frac{p+q}{2}-2}) \sum_{r=1}^{p\wedge q-\mathbb{1}_{\{p=q\}}} \gamma_{n,p,q,r},$$

where $\gamma_{n,p,q,r}$ is defined in (2.5). Note that $\sum_{k\in\mathbb{Z}}|k|^{am}$ is finite since $a<-1/m$. Furthermore, in general, using that $n^{b+1}\geqslant 1$ for $n\geqslant 1$ it holds that

$$\sum_{|t|<n}|t|^b \leqslant \begin{cases} \sum_{t\in\mathbb{Z}} t^b < C_2, & b<-1,\\ 2\int_1^{n-1}\frac{1}{t}\mathrm{d}t \leqslant C_3\log(n), & b=-1,\\ 2+2\int_2^{n-1} t^b\mathrm{d}t \leqslant 2+2\frac{n^{b+1}-1}{b+1} \leqslant C_4 n^{b+1}, & b>-1. \end{cases}$$

Combining these observations into $\gamma_{n,p,q,r}$ one obtains

$$\begin{aligned}\gamma^2_{n,p,q,r} &= \frac{2}{n}\sum_{k\in\mathbb{Z}}|k|^{am}\sum_{|t|<n}|t|^{ar}\sum_{|s|<n}|s|^{a(p\wedge q-r)}\\ &\leqslant \begin{cases} C_5 n^{-1}, & ar<-1,\ a(p\wedge q-r)<-1,\\ C_6 n^{a(p\wedge q-r)}, & ar<-1,\ a(p\wedge q-r)>-1,\\ C_7 n^{ar}, & ar>-1,\ a(p\wedge q-r)<-1,\\ C_8 n^{ap\wedge q+1}, & ar>-1,\ a(p\wedge q-r)>-1,\\ C_9 n^{-1/2}\log(n), & \text{one of the terms is } =-1 \text{ and the other } <-1,\\ C_{10} n^{aL/2}\log(n), & \text{one of the terms is } =-1 \text{ and the other } >-1, \end{cases}\end{aligned}$$

for $r\in\{1,\dots,p\wedge q-\mathbb{1}_{\{p=q\}}\}$ and $2\leqslant m\leqslant p\wedge q$. In a more compact form, as $a/2>(am+1)/2$ when $a\in(-1,-1/(m-1))$, one gets

$$\gamma_{n,p,q,r}\leqslant \Gamma_{a,m}(n) := \begin{cases} C_5 n^{-1/2}, & a<-1\\ C_9 n^{-1/2}\log(n), & a=-1\\ C_6 n^{a/2}, & a\in(-1,-1/(m-1)]\\ C_8 n^{(am+1)/2}, & a\in(-1/(m-1),-1/m). \end{cases}$$

Hence,

$$\sum_{r=1}^{p\wedge q-\mathbb{1}_{\{p=q\}}}\gamma_{n,p,q,r}\leqslant \sum_{r=1}^{p\wedge q-\mathbb{1}_{\{p=q\}}}\Gamma_{a,m}(n)=(p\wedge q)\Gamma_{a,m}(n). \tag{4.53}$$

and therefore (4.53) yields that

$$\Delta_1(n,d,K)\leqslant \Gamma_{a,m}(n)\sum_{p,q\geqslant m}(p\wedge q)pq\frac{(4e^{1/2+\alpha})^{p+q}\sqrt{K}^{p+q}}{(p!q!)^{\beta}}\log^{(p+q)/2-1}(2d^2+e^{(p+q)/2-2}). \tag{4.54}$$

The sum appearing in (4.54) is the same sum as in (4.30), with additional multiplicative term $(p\wedge q)K^{1/2(p+q)}$. Hence, we use exactly the same steps as in the proof of Corollary 4.2 to obtain the following: For $\beta\in(1/2,1]$, apply steps (4.30)-(4.40), from the proof of Corollary 4.2 to obtain $\Delta_1(n,d,K)\leqslant c\Gamma_{a,m}(n)(\Delta_1^{(1)}(d,k)+\Delta_1^{(2)}(K))$, where

$$\begin{aligned}\Delta_1^{(1)}(d,k) &= c_1\frac{e^{k_1K^{1/(2\beta)}\log^{1/(2\beta)}(d)}}{\log_+(d)}, \quad \text{and}\\ \Delta^{(2)}(K) &= c_2\sum_{p\geqslant 0}\frac{(4\sqrt{K}e^{1+\alpha+3/(4e)+k_2K^{1/(2\beta)}})^p}{(p!)^{\beta-1/2}},\end{aligned} \tag{4.55}$$

with explicit constants $c_1 = 2^{4\beta-1}(\pi e^{1/3}/2)^\beta$, $k_1 = \exp\left(2\beta e^{1/2e}(2^{5/2}e^{\alpha+1/2+2/e})^{1/\beta}\right)$, $c_2 = 2^{2\beta+5/4}\pi^{-1/4}(\pi e^{1/3}/2)^{\beta/2}$ and $k_2 = \exp\left(\beta e^{1/(2e)}(4e^{1/2+1/e+\alpha})^{1/\beta}\right)$.

However, compared to (4.40), it is not enough to show that $\Delta^{(2)}(K)$ in(4.55) is finite, we need a bound explicit in K. To that end, for $x > 0$ and $\delta \in (0, 1/2]$, denote $S(x) = \sum_{p\geqslant 0} a_p(x)$, where $a_p = x^p(p!)^{-\delta}$. Given that $a_{p+1}/a_p = x(p+1)^{-\delta}$, by the ratio test, $S(x)$ converges when $p+1 > x^{1/\delta}$, and a_p increases as $p+1$ grows toward $p+1 \leqslant x^{1/\delta}$, and decreases afterwards. Let $M := \lceil x^{1/\delta}\rceil$, then $\max_{p\geqslant 0} a_p$ is either a_M or a_{M-1}, depending if the ration a_{p+1}/a_p crosses 1 before or after M. Hence it holds that $\sum_{p=0}^{2M} a_p \leqslant (2M+1)a_M$. For $p \geqslant 2M$,

$$\frac{a_{p+1}}{a_p} = \frac{x}{(p+1)^\delta} \leqslant \frac{M^\delta}{(2M)^\delta} = 2^{-\delta},$$

and therefore the tail is geometric

$$\sum_{p\geqslant 2M} a_p \leqslant \frac{a_{2M}}{1-2^{-\delta}} \leqslant \frac{a_M}{1-2^{-\delta}}.$$

Using $2M+1 \leqslant 3(1+x^{1/\delta})$, we obtain that

$$S(x) = \sum_{p=0}^{2M} a_p(x) + \sum_{p=2M}^{\infty} a_p(x) \leqslant \left(2M+1+\frac{1}{1-2^{-\delta}}\right) a_M \leqslant \left(3+\frac{1}{1-2^{-\delta}}\right)(1+x^{1/\delta})a_M.$$

As $M^\delta \geqslant x$ and $\exp(\delta m) \leqslant \exp(\delta + \delta x^{1/\delta})$, using Stirling's lower bound (4.18), we get that

$$a_M = \frac{x^M}{(M!)^\delta} \leqslant \frac{1}{(\sqrt{2\pi})^\delta} e^{\delta M}(\frac{x}{M^\delta})^M \leqslant \frac{e^\delta}{(\sqrt{2\pi})^\delta} e^{\delta x^{1/\delta}}.$$

Finally, using $1+x \leqslant e^x$ for all $x \geqslant 0$,

$$S(x) \leqslant (3+\frac{1}{1-2^{-\delta}})(1+x^{1/\delta})\frac{e^\delta}{(\sqrt{2\pi})^\delta} e^{\delta x^{1/\delta}} \leqslant (3+\frac{1}{1-2^{-\delta}})\frac{e^\delta}{(\sqrt{2\pi})^\delta} e^{(\delta+1)x^{1/\delta}}. \tag{4.56}$$

Define $\psi(K) = 4K^{1/2}e^{1+\alpha+3/(4e)+k_2K^{1/(2\beta)}}$, and for $\delta = \beta - 1/2$, and $\beta \in (1/2, 1]$, using (4.55) and (4.56) with $C_\delta := (3+(1-2^{-\delta})^{-1})e^\delta(\sqrt{2\pi})^{-\delta}$ we obtain

$$\Delta_1^{(2)}(K) \leqslant c_3 \exp\left((\beta+1/2)\psi(K)^{1/(\beta-1/2)}\right),$$

where $c_3 = c_2 C_{\beta-1/2}$, which finally yields that

$$\Delta_1(n,d,K) \leqslant c\Gamma_{a,m}(n)(c_1+c_3)\frac{e^{k_1K^{1/(2\beta)}\log^{1/(2\beta)}(d)}}{\log_+(d)} e^{3\psi(K)^{1/(\beta-1/2)}/2}. \tag{4.57}$$

On the other hand, for $\beta = 1/2$, applying (4.30)-(4.37) and (4.41)-(4.42) from the proof of Corollary 4.2, yields that $\Delta_1(n,d,K) \leqslant c\Gamma_{a,m}(n)(\Delta_1^{(1)}(d,k) + \Delta_1^{(2)}(K))$, where

$$\begin{aligned} \Delta_1^{(1)}(d,k) &= \sqrt{2\pi}e^{1/12}\frac{d^{k_3K}}{\log_+(d)}, \quad \text{and} \\ \Delta_1^{(2)}(K) &= 4e^{1/12}\sum_{p\geqslant 0}\left(4\sqrt{K}e^{1+\alpha+3/(4e)+k_4K}\right)^p, \end{aligned} \tag{4.58}$$

where $k_3 = 2^4\exp(2\alpha + 1 + 9/(2e))$ and $k_4 = 8\exp(1 + 2\alpha + 5/(2e))$. $\Delta_1^{(2)}(K)$ is a geometric series that converges when $x := 4\sqrt{K}e^{1+\alpha+3/(4e)+k_4K} < 1$, as $x > 0$. Using the steps (4.42)-(4.43) as in the proof of Corollary 4.2, one gets that the series $\Delta_1^{(2)}(K)$ converges when $\alpha < \alpha_0 - \log(K)/2$, and that $\Delta_1^{(2)}(K) \leqslant 4e^{1/12}(1-\psi(K))^{-1}$. Combining $\alpha < \alpha_0 - \log(K)/2$ with k_3K, yields that $k_3K < 2^4\exp(2\alpha_0 + 1 + 9/(2e))$. Therefore, together with (4.58), for $k_5 := 2^4\exp(2\alpha_0 + 1 + 9/(2e)) \approx 0.768$, it holds that

$$\Delta_1(n, d, K) \leqslant c_4\Gamma_{a,m}(n)\frac{d^{k_5}}{\log_+(d)}\frac{1}{1-\psi(K)}, \tag{4.59}$$

where $c_4 = ce^{1/12}(\sqrt{2\pi} + 4)$. Using Theorem 2.3, (4.57), and (4.59), together with (4.1), we get for $\beta \in (1/2, 1]$

$$\begin{aligned} d_{\mathcal{R}}(\boldsymbol{S}_n, \boldsymbol{Z}_n) &\leqslant C\zeta(d,K)\psi(K)^{1/(\beta-1/2)}\Gamma_{a,m}(n)\log_+(n)(\log_+(\log_+(n))\mathbb{1}_{\{a=-1\}} + \mathbb{1}_{\{a\neq-1\}}) \\ &\times \frac{\log_+(\sigma_*(\boldsymbol{\Lambda}_n))}{\sigma_*(\boldsymbol{\Lambda}_n)}, \quad \text{where} \quad \zeta(d,K) = \log_+^{1/(2\beta)}(d)e^{c_2K^{1/(2\beta)}\log_+^{1/(2\beta)}(d)}, \end{aligned}$$

and for $\beta = 1/2$, it follows similarly that

$$\begin{aligned} d_{\mathcal{R}}(\boldsymbol{S}_n, \boldsymbol{Z}_n) &\leqslant C\frac{\log_+(d)d^{k_5}}{1-\psi(K)}\Gamma_{a,m}(n)\log_+(n)(\log_+(\log_+(n))\mathbb{1}_{\{a=-1\}} + \mathbb{1}_{\{a\neq-1\}}) \\ &\times \frac{\log_+(\sigma_*(\boldsymbol{\Lambda}_n))}{\sigma_*(\boldsymbol{\Lambda}_n)}. \end{aligned}$$

For the convex and 1-Wasserstein distance, as $p \wedge q \leqslant q$, using (4.38) to get $\sqrt{p} \leqslant e^{p/2e}$, together with (4.53) and the assumption (2.6), yields that

$$\begin{aligned} \Delta_2(n,d,K) &\leqslant c\Gamma_{a,m}(n)\sum_{p,q\geqslant m}(p\wedge q)\frac{\sqrt{pq}(2\sqrt{K}e^{\alpha})^{p+q}}{(p!q!)^{\beta-1/2}} \leqslant c\Gamma_{a,m}(n)\sum_{p,q\geqslant m}\frac{\sqrt{p}(2\sqrt{K}e^{\alpha})^{p+q}}{(p!q!)^{\beta-1/2}} \\ &\leqslant c\Gamma_{a,m}(n)\left(\sum_{p\geqslant m}\frac{(2\sqrt{K}e^{\alpha+1/(2e)})^p}{(p!)^{\beta-1/2}}\right)^2 \end{aligned}$$

Now, if $\beta \in (1/2, 1]$, then (4.56) yields that

$$\sum_{p\geqslant m}\frac{(2\sqrt{K}e^{\alpha+1/(2e)})^p}{(p!)^{\beta-1/2}} \leqslant \sum_{p\geqslant 0}\frac{(2\sqrt{K}e^{\alpha+1/(2e)})^p}{(p!)^{\beta-1/2}} \leqslant c_1e^{c_2K^{1/(2\beta-1)}},$$

for $c_1 = (\sqrt{2\pi})^{1/2-\beta}$ and $c_2 = (\beta - 1/2)2^{1/(\beta-1/2)}\exp\left[1/(2e\beta - e) + \alpha/(\beta - 1/2) + 1\right]$. Hence,

$$\Delta_2(n,d,K) \leqslant c_3\Gamma_{a,m}(n)e^{c_4K^{1/(2\beta-1)}},$$

where $c_3 = cc_1^2$ and $c_4 = 2c_2$. On the other hand, if $\beta = 1/2$, then for $\alpha < -1/(2e) - \ln(2\sqrt{K})$,

$$\sum_{p\geqslant 0}(2\sqrt{K}e^{\alpha+1/(2e)})^p = \frac{1}{1-c_5\sqrt{K}}, \tag{4.60}$$

where $c_5 = 2e^{\alpha+1/(2e)}$, as (4.60) is just a geometric series. Hence, the proof is concluded with

$$\Delta_2(n,d,K) \leqslant c_6\Gamma_{a,m}(n)\frac{1}{(1-c_5\sqrt{K})^2}.$$

□

4.6 Proof of Theorem 3.1.

The proof of the finite-analogue to Theorem 2.1 is simpler as there is no need to insure summability of the bound. As it is sharper, it makes sense to present a separate proof. The tools used in this proof are the same as in the Theorem 4.1.

Proof of Theorem 3.1. First, observe that if $\boldsymbol{F}$ belongs entirely to the first Wiener chaos (i.e., $q_i = 1$ for all $i \in 1, \ldots, d$), then $\boldsymbol{F}$ is a centred Gaussian vector with covariance matrix $\boldsymbol{\Sigma}$. Indeed, we may write $F_i = I_1(f_{i,1})$ for some $f_{i,1} \in \mathfrak{H}$. By the linearity of the mapping $h \mapsto I_1(h)$, any linear combination $\sum_{i=1}^d a_i F_i = I_1(\sum_{i=1}^d a_i f_{i,1})$ is a univariate Gaussian random variable. Consequently, $\boldsymbol{F} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$ with entries $\boldsymbol{\Sigma}_{i,j} = \langle f_{i,1}, f_{j,1} \rangle_{\mathfrak{H}}$. Since $\boldsymbol{Z} \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma})$ by definition, we have $\boldsymbol{F} \overset{d}{=} \boldsymbol{Z}$, and thus $d_M(\boldsymbol{F}, \boldsymbol{Z}) = 0$, where $M \in \{\mathcal{R}, \mathscr{C}, \mathcal{W}\}$.

Therefore, for the remainder of the proof, assume that there exists an $i \in \{1, \ldots, d\}$ such that $q_i > 1$. For $d \geqslant 3$, (4.8) yields that $d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C \Delta_{\boldsymbol{F}} (\sigma_*)^{-1} \log(d) \log_+\{\underline{\sigma} \Delta_{\boldsymbol{F}} (\overline{\sigma} \sigma_*)^{-1}\}$, where using (4.20)

$$\begin{aligned} \Delta_{\boldsymbol{F}} := \mathbb{E}\left[\max_{1 \leqslant i,j \leqslant d} \left| \boldsymbol{\Sigma}_{ij} - \boldsymbol{\tau}_{i,j}^{\boldsymbol{F}}(\boldsymbol{F}) \right| \right] &\leqslant \mathbb{E}\left[\max_{1 \leqslant i,j \leqslant d} \left| \boldsymbol{\Sigma}_{ij} - \langle -DL^{-1} F_i, DF_j \rangle_{\mathfrak{H}} \right| \right] \\ &= \mathbb{E}\left[\max_{1 \leqslant i,j \leqslant d} |\Delta_{i,j}| \right], \end{aligned} \tag{4.61}$$

with $\Delta_{i,j}$ defined as

$$\begin{aligned} \Delta_{i,j} &:= \boldsymbol{\Sigma}_{ij} - \langle -DL^{-1} F_i, DF_j \rangle_{\mathfrak{H}} = \mathbb{E}[F_i F_j] - \langle -DL^{-1} F_i, DF_j \rangle_{\mathfrak{H}} \\ &= \sum_{k=1}^{q_i} \sum_{\ell=1}^{q_j} \left(\mathbb{E}[I_k(f_{i,k}) I_\ell(f_{j,\ell})] - \left\langle -DL^{-1} I_k(f_{i,k}), DI_\ell(f_{j,\ell}) \right\rangle_{\mathfrak{H}} \right) \\ &= \sum_{k=1}^{q_i} \sum_{\ell=1}^{q_j} b(i,j,k,\ell), \end{aligned} \tag{4.62}$$

where $b(i,j,k,\ell) := \mathbb{E}[I_k(f_{i,k}) I_\ell(f_{j,\ell})] - \left\langle -DL^{-1} I_k(f_{i,k}), DI_\ell(f_{j,\ell}) \right\rangle_{\mathfrak{H}}$ for $i,j \in \{1, \ldots d\}$. By (4.5), $-L^{-1} I_k(f_{i,k}) = k^{-1} I_k(f_{i,k})$ for all $i \in \{1, \ldots, d\}$ and $k \in \{1, \ldots q_i\}$, and hence

$$b(i,j,k,\ell) = \mathbb{E}[I_k(f_{i,k}) I_\ell(f_{j,\ell})] - \frac{1}{k} \langle DI_k(f_{i,k}), DI_\ell(f_{j,\ell}) \rangle_{\mathfrak{H}} \quad \text{for } i,j \in \{1, \ldots d\}.$$

If $d \in \{1,2\}$, then we use [3, Lem. 5.4] to get

$$d_{\mathcal{R}}(\boldsymbol{F}, \boldsymbol{Z}) \leqslant C \frac{\Delta_{\boldsymbol{F}}}{\sigma_*} \log(d) \log_+ \left(\frac{(\underline{\sigma} \vee 1) \Delta_{\boldsymbol{F}}}{(\overline{\sigma} \wedge 1)(\sigma_* \vee 1)} \right).$$

Note that when $k = \ell = 1$, we have $b(i,j,1,1) = \langle f_{i,1}, f_{j,1} \rangle_{\mathfrak{H}} - I_0(f_{i,1} \otimes_1 f_{j,1}) = 0$ for all $i,j \in \{1, \ldots, d\}$. Furthermore, whenever $q_i = q_j = 1$, it follows that $b(i,j,k,\ell) = 0$ for $1 \leqslant k \leqslant q_i$ and $1 \leqslant \ell < q_j$, and hence $\Delta_{i,j} = 0$ since the only term in the definition (4.62) of $\Delta_{i,j}$ is $b(i,j,1,1) = 0$.

Consequently, to bound $\mathbb{E}[\max_{1 \leqslant i,j \leqslant d} |\Delta_{i,j}|]$, we only need to consider the case $(q_i, q_j) \neq (1,1)$. Define $q := \max_{1 \leqslant i \leqslant d} q_i$; then $q > 1$ by the above argument.

The proof of [39, Lem. 6.2.1] yields that

$$\langle DI_k(f_{i,k}), DI_\ell(f_{j,\ell})\rangle_{\mathfrak{H}} = k\ell \sum_{r=1}^{k\wedge\ell}(r-1)!\binom{k-1}{r-1}\binom{\ell-1}{r-1} I_{k+\ell-2r}(f_{i,k}\tilde{\otimes}_r f_{j,\ell}),$$

which furthers implies that $b(i,j,k,\ell) \in \mathcal{P}_{k+\ell-2}$. Therefore $\Delta_{i,j} \in \mathcal{P}_{q_i+q_j-2}$ for all $i,j \in \{1,\ldots,d\}$, and as $q_i + q_j \leqslant 2q$, we conclude that $\Delta_{i,j} \in \mathcal{P}_{2q-2}$ for all $i,j \in \{1,\ldots,d\}$. Then, as $q > 1$, [32, Prop. A1] yields that $\Delta_{i,j}$ is a sub-$(2q-2)$-th chaos random variable relative to scale $M_{2q-2}\|\Delta_{i,j}\|_2$, where one can take $M_x = (4e/x)^{x/2}$ as discussed in [3]. Lemma 4.4 then yields that

$$\begin{aligned}\Delta_{\boldsymbol{F}} &= \mathbb{E}\left[\max_{1\leqslant i,j\leqslant d}|\Delta_{i,j}|\right] \leqslant \max_{1\leqslant i,j\leqslant d} \log^{\frac{2q-2}{2}}(2d^2-1+e^{\frac{2q-2}{2}-1})M_{2q-2}\|\Delta_{i,j}\|_2 \\ &= \log^{q-1}(2d^2-1+e^{q-2})M_{2q-2}\max_{1\leqslant i,j\leqslant d}\|\Delta_{i,j}\|_2.\end{aligned} \tag{4.63}$$

Next, using the definition of $\Delta_{i,j}$ in (4.62) and the triangle inequality, and dropping the constraint $(k,\ell) \neq (1,1)$ (which merely adds a zero term), we obtain

$$\Delta_{\boldsymbol{F}} = \log^{q-1}(2d^2-1+e^{q-2})M_{2q-2}\max_{1\leqslant i,j\leqslant d}\sum_{k=1}^{q_i}\sum_{\ell=1}^{q_j}\|b(i,j,k,\ell)\|_2. \tag{4.64}$$

To bound $\|b(i,j,k,\ell)\|_2$, for $i,j \in \{1,\ldots d\}$, $k \in \{1,\ldots,q_i\}$, $\ell \in \{1,\ldots,q_j\}$, Lemma 4.3 together with the fact that $k,\ell < q$ for all k and ℓ, imply that

$$\begin{aligned}\|b(i,j,k,\ell)\|_2^2 &= \|\mathbb{E}[I_k(f_{i,k})I_\ell(f_{j,\ell})] - \left\langle -DL^{-1}I_k(f_{i,k}), DI_\ell(f_{j,\ell})\right\rangle_{\mathfrak{H}}\|_2^2 \\ &\leqslant \frac{k\ell!k!}{\ell}2^{2k+2\ell}\sum_{r=1}^{b(k,\ell)}\|f_{i,k}\otimes_r f_{j,\ell}\|^2_{\mathfrak{H}^{\otimes(k+\ell-2r)}} \\ &\leqslant \frac{K_q^2}{\ell}b(k,\ell)\max_{1\leqslant r\leqslant b(k,\ell)}\|f_{i,k}\otimes_r f_{j,\ell}\|^2_{\mathfrak{H}^{\otimes(k+\ell-2r)}},\end{aligned}$$

where $K_q := \sqrt{q}q!4^q$. As $(a+b)^{1/2} \leqslant a^{1/2} + b^{1/2}$ and as $b(k,\ell) \leqslant \ell$, it follows that

$$\|\Delta_{i,j}\|_2 \leqslant \sum_{k=1}^{q_i}\sum_{\ell=1}^{q_j}\|b(i,j,k,\ell)\|_2 \leqslant \sum_{k=1}^{q_i}\sum_{\ell=1}^{q_j}K_q\max_{1\leqslant r\leqslant b(k,\ell)}\|f_{i,k}\otimes_r f_{j,\ell}\|_{\mathfrak{H}^{\otimes(k+\ell-2r)}}.$$

Let $1 \leqslant R \leqslant b(k,\ell)$ and $K \in \{1,\ldots,q_i\}$ and $L \in \{1,\ldots,q_j\}$ be, such that

$$\|f_{i,K}\otimes_R f_{j,L}\|_{\mathfrak{H}^{\otimes(K+L-2R)}} = \max_{\substack{1\leqslant k\leqslant q_i\\1\leqslant\ell\leqslant q_j\\(k,\ell)\neq(1,1)}}\max_{1\leqslant r\leqslant b(k,\ell)}\|f_{i,k}\otimes_r f_{j,\ell}\|_{\mathfrak{H}^{\otimes(k+\ell-2r)}}.$$

Then we, obtain that

$$\begin{aligned}\max_{1\leqslant i,j\leqslant d}\|\Delta_{i,j}\|_2 &\leqslant K_q\max_{1\leqslant i,j\leqslant d}\sum_{k=1}^{q_i}\sum_{\substack{\ell=1\\(k,\ell)\neq(1,1)}}^{q_j}\max_{1\leqslant r\leqslant b(k,\ell)}\|f_{i,k}\otimes_r f_{j,\ell}\|_{\mathfrak{H}^{\otimes(k+\ell-2r)}} \\ &\leqslant q^2K_q\max_{1\leqslant i,j\leqslant d}\|f_{i,K}\otimes_R f_{j,L}\|_{\mathfrak{H}^{\otimes(K+L-2R)}}.\end{aligned} \tag{4.65}$$

As $ed^2, e^{q-2}-1 \geqslant 1$, using (4.1) it follows that

$$\log(2d^2+(e^{q-2}-1)) \leqslant \log_+(2d^2+(e^{q-1})) \leqslant 4\log_+(2)(q-1)\log_+(d).$$

Hence for all $q \geqslant 2$, there exists a constant $\widetilde{C}_q := 4\log_+(2)(q-1)$depending only on q, such that

$$\log^{q-1}(2d^2+(e^{q-2}-1)) \leqslant \widetilde{C}_q \log_+^{q-1}(d). \tag{4.66}$$

Thus, using (4.65), (4.63), together with $\overline{C}_q = \widetilde{C}_q M_{2q-2} q^2 K_q$, yields

$$\Delta_{\boldsymbol{F}} \leqslant \overline{C}_q \log^{q-1}(d) \max_{1\leqslant i,j\leqslant d} \|f_{i,K} \otimes_R f_{j,L}\|_{\mathfrak{H}^{\otimes(K+J-2R)}} = \overline{C}_q \log^{q-1}(d)\overline{\Delta}_{\boldsymbol{F}}, \tag{4.67}$$

where $\overline{\Delta}_{\boldsymbol{F}} := \max_{1\leqslant i,j\leqslant d}\|f_{i,K} \otimes_R f_{j,L}\|_{\mathfrak{H}^{\otimes(K+J-2R)}}$.

Recall that $q=1$ implies that $\boldsymbol{F}$ is Gaussian and hence $d_{\mathcal{R}}(\boldsymbol{F},\boldsymbol{Z})=0$. Using Remark 4.6 together with (4.67) and (4.61), yields

$$d_{\mathcal{R}}(\boldsymbol{F},\boldsymbol{Z}) \leqslant C\frac{\overline{C}_q \log_+^{q-1}(d)\overline{\Delta}_{\boldsymbol{F}}}{\sigma_*(\boldsymbol{\Lambda})}\log(d)\log_+\left(\frac{\overline{C}_q \log_+^{q-1}(d)\overline{\Delta}_{\boldsymbol{F}}}{\sigma_*(\boldsymbol{\Lambda})}\right). \tag{4.68}$$

Moreover, using (4.1) implies that

$$\begin{aligned}
\log_+\left(\frac{\overline{C}_q \log_+^{q-1}(d)\overline{\Delta}_{\boldsymbol{F}}}{\sigma_*(\boldsymbol{\Lambda})}\right) &\leqslant 8\log_+(\overline{C}_q)\log_+(\log_+^{q-1}(d))\log_+(\overline{\Delta}_{\boldsymbol{F}})\log_+(\sigma_*^{-2}(\boldsymbol{\Lambda}))\\
&\leqslant \overline{K}_q(|\log(\log_+(d))| \vee (q-1)^{-1})\log_+(\overline{\Delta}_{\boldsymbol{F}})\log_+(\sigma_*(\boldsymbol{\Lambda}))\\
&\leqslant \overline{K}_q\log_+(\log_+(d))\log_+(\overline{\Delta}_{\boldsymbol{F}})\log_+(\sigma_*(\boldsymbol{\Lambda})),
\end{aligned}$$

where $\overline{K}_q = 16(q-1)(|\log \overline{C}_q| \vee 1)$. Let $C_q := C\overline{K}_q\overline{C}_q$, where $C>0$ is an universal constant. Combining the obtained bound in the display above with (4.68), we get that

$$d_{\mathcal{R}}(\boldsymbol{F},\boldsymbol{Z}) \leqslant C_q \log^q(d)(|\log\log(d)|\vee 1)\overline{\Delta}_{\boldsymbol{F}}(|\log \overline{\Delta}_{\boldsymbol{F}}| \vee 1)\frac{|\log\sigma_*(\boldsymbol{\Lambda})| \vee 1}{\sigma_*(\boldsymbol{\Lambda})}.$$

Using $\overline{C}_q = \widetilde{C}_q M_{2q-2}q^2K_q \leqslant 4log_+(2)q^{5/2}(8e)^q$, together with (4.1), there exists $c \geqslant 0$, such that

$$C_q \leqslant c\log_+(q)q^{9/2}(8e)^q.$$

Convex and 1-Wasserstein distance. Recall (4.9) and (4.10). As in the proof of the Theorem 4.1, take $u_k = -D(L^{-1}F_k)$ for all $k \in \{1,\dots,d\}$, as then $\delta(u_k) = -\delta(DL^{-1}F_k) = F_k$. Hence, to conclude the proof, it suffices to bound $\sqrt{\mathbb{E}[\|M_{\boldsymbol{F}} - \boldsymbol{\Sigma}\|_{\text{H.S.}}^2]}$, where $M_{\boldsymbol{F}}(i,j) = \langle Df_i, -DL^{-1}F_j\rangle_{\mathfrak{H}}$.

Using the same steps as in the proof of Theorem 4.1 Part 2, it holds that

$$\sqrt{\mathbb{E}[\|M_{\boldsymbol{F}} - \boldsymbol{\Sigma}\|_{\text{H.S.}}^2]} \leqslant \sqrt{\mathbb{E}[\|\boldsymbol{\Sigma} - \operatorname{Cov}(\boldsymbol{F})\|_{\text{H.S.}}^2]} + d\max_{1\leqslant i,j\leqslant d}\|\operatorname{Cov}(\boldsymbol{F})_{ij} - M_{\boldsymbol{F}}(i,j)\|_2 \tag{4.69}$$

Definition of $M_{\boldsymbol{F}}$ and (4.5) yield

$$\begin{aligned}
\operatorname{Cov}(\boldsymbol{F})_{ij} - M_{\boldsymbol{F}}(i,j) &= \mathbb{E}[F_iF_j] - \langle DF_i, -DL^{-1}F_j\rangle_{\mathfrak{H}}\\
&= \mathbb{E}[F_iF_j] - \langle D\sum_{k=1}^{q_i} I_k(f_{i,k}), D\sum_{\ell=1}^{q_j}\frac{1}{\ell}I_\ell(f_{j,\ell})\rangle_{\mathfrak{H}}\\
&= \sum_{k=1}^{q_i}\sum_{\ell=1}^{q_j}\left(\mathbb{E}[I_k(f_{i,k})I_\ell(f_{j,\ell})] - \frac{1}{\ell}\langle DI_k(f_{i,k}), DI_\ell(f_{j,\ell})\rangle_{\mathfrak{H}}\right) = \Delta_{i,j},
\end{aligned}$$

where $\Delta_{i,j}$ is defined in (4.62). Since $\mathrm{Cov}(\boldsymbol{F}) = \mathrm{Cov}(\boldsymbol{Z})$, we have $\mathbb{E}[\|\boldsymbol{\Sigma} - \mathrm{Cov}(\boldsymbol{F})\|_{\mathrm{H.S.}}^2] = 0$, so (4.69) reduces to $\sqrt{\mathbb{E}[\|M_{\boldsymbol{F}} - \boldsymbol{\Sigma}\|_{\mathrm{H.S.}}^2]} \leqslant d \max_{1\leqslant i,j\leqslant d} \|\Delta_{i,j}\|_2$. Recalling the definition of $\overline{\Delta}_{\boldsymbol{F}}$ in (4.67), the bound (4.65) implies $\max_{1\leqslant j,k\leqslant d} \|\Delta_{j,k}\|_2 \leqslant q^2 K_q \overline{\Delta}_{\boldsymbol{F}}$. This completes the proof, as

$$\sqrt{\mathbb{E}[\|M_{\boldsymbol{F}} - \boldsymbol{\Sigma}\|_{\mathrm{H.S.}}^2]} \leqslant \sqrt{\mathbb{E}[\|\boldsymbol{\Sigma} - \mathrm{Cov}(\boldsymbol{F})\|_{\mathrm{H.S.}}^2]} + dq^2 K_q \overline{\Delta}_{\boldsymbol{F}}.$$

□

proof of Corollary 3.3. The proof follows exactly as proof to Theorem 3.1 until (4.64). Recall, our goal is to find

$$\max_{1\leqslant i,j\leqslant d} \|\Delta_{i,j}\|_2 \leqslant \max_{1\leqslant i,j\leqslant d} \sum_{k=1}^{q_i} \sum_{\substack{\ell=1 \\ (k,\ell)\neq(1,1)}}^{q_j} \|b(i,j,k,\ell)\|_2. \tag{4.70}$$

Considering the variables $b(i,j,k,\ell) = \mathbb{E}[I_k(f_{i,k}) I_\ell(f_{j,\ell})] - \left\langle -DL^{-1} I_k(f_{i,k}), DI_\ell(f_{j,\ell}) \right\rangle_{\mathfrak{H}}$ for $i,j \in \{1,\dots d\}$, $k \in \{1,\dots,q_i\}$, $\ell \in \{1,\dots,q_j\}$, we use the proof of [39, Thm 6.2.2] to bounds these. First, assume that $k=\ell$, then it follows directly from [39, Eq.(6.2.6)] that

$$\begin{aligned} \|b(i,j,k,\ell)\|_2^2 &= \|\mathbb{E}[I_k(f_{i,k}) I_\ell(f_{j,\ell})] - \left\langle -DL^{-1} I_k(f_{i,k}), DI_\ell(f_{j,\ell}) \right\rangle_{\mathfrak{H}}\|_2^2 \\ &\leqslant \begin{cases} 0, & k=\ell=1, \\ 1/2\left\{\kappa_4(I_k(f_{i,k})) + \kappa_4(I_\ell(f_{j,\ell}))\right\} \sum_{r=1}^{i} \binom{2r}{r}, & k=\ell>1. \end{cases} \end{aligned} \tag{4.71}$$

Since $i \leqslant q_k \vee q_\ell \leqslant q$, and binomial coefficients are positive, we see for $i=\ell>1$, that

$$\frac{1}{2}\sum_{r=1}^{i}\binom{2r}{r} \leqslant \frac{1}{2}\sum_{r=1}^{q_k\vee q_\ell}\binom{2r}{r} \leqslant \frac{1}{2}\sum_{r=1}^{q}\binom{2r}{r} =: C_q,$$

where $C_q = (1/2)\sum_{r=1}^{q}\binom{2r}{r}$. If instead $k \neq \ell$, then [39, Eq. (6.2.7)] yields that

$$\begin{aligned} \|b(i,j,k,\ell)\|_2^2 \leqslant{}& \mathbb{E}[I_k(f_{i,k})^2]\kappa_4(I_\ell(f_{j,\ell}))^{1/2} + \mathbb{E}[I_\ell(f_{j,\ell})^2]\kappa_4(I_k(f_{i,k}))^{1/2} \\ &+ \frac{1}{2}\sum_{r=1}^{k\wedge\ell-1}(k+\ell-2r)!\left\{\binom{\ell}{r}^2\kappa_4(I_k(f_{i,k})) + \binom{k}{r}^2\kappa_4(I_\ell(f_{j,\ell}))\right\}. \end{aligned}$$

Moreover, since $k,\ell < q$ and binomial coefficients are increasing in their upper index, we have the bound

$$\frac{1}{2}\sum_{r=1}^{k\wedge\ell-1}(k+\ell-2r)!\max\left\{\binom{\ell}{r}^2, \binom{k}{r}^2\right\} \leqslant \frac{1}{2}\sum_{r=1}^{q-1}(2q-2r)!\binom{q}{r}^2 =: K_q,$$

where $K_q = (1/2)\sum_{r=1}^{q-1}(2q-2r)!\binom{q}{r}^2$ and thus obtain

$$\begin{aligned} \|b(i,j,k,\ell)\|_2^2 \leqslant{}& \mathbb{E}[I_k(f_{i,k})^2]\kappa_4(I_\ell(f_{j,\ell}))^{1/2} + \mathbb{E}[I_\ell(f_{j,\ell})^2]\kappa_4(I_k(f_{i,k}))^{1/2} \\ &+ K_q\left\{\kappa_4(I_k(f_{i,k})) + \kappa_4(I_\ell(f_{j,\ell}))\right\}. \end{aligned} \tag{4.72}$$

Combining the two cases, namely (4.71) and (4.72), with (4.70), it follows that

$$\|\Delta_{i,j}\|_2 \leqslant \sum_{k=1}^{q_i}\sum_{\ell=1}^{q_\ell}\|b(i,j,k,\ell)\|_2 \leqslant \sum_{k=1}^{q_i}\sum_{\substack{\ell=1\\(k,\ell)\neq(1,1)}}^{q_j}\Big\{\widetilde{K}_q\left[\kappa_4(I_k(f_{i,k}))+\kappa_4(I_\ell(f_{j,\ell}))\right]$$
$$+\mathbb{1}_{\{k\neq\ell\}}\left[\mathbb{E}[I_k(f_{i,k})^2]\kappa_4(I_\ell(f_{j,\ell}))^{1/2}+\mathbb{E}[I_\ell(f_{j,\ell})^2]\kappa_4(I_k(f_{i,k}))^{1/2}\right]\Big\}^{1/2},$$

where $\widetilde{K}_q=\max\{C_q,K_q\}$. Next we want to estimate $\max_{1\leqslant i,j\leqslant d}\|\Delta_{i,j}\|_2$. It follows that

$$\kappa_4(I_k(f_{i,k}))+\kappa_4(I_\ell(f_{j,\ell}))\leqslant 2\max_{1\leqslant i\leqslant d}\max_{1\leqslant k\leqslant q_i}\kappa_4(I_k(f_{i,k}))$$

and

$$\mathbb{E}[I_k(f_{i,k})^2]\kappa_4(I_\ell(f_{j,\ell}))^{1/2}+\mathbb{E}[I_\ell(f_{j,\ell})^2]\kappa_4(I_k(f_{i,k}))^{1/2}$$
$$\leqslant \max_{1\leqslant i,j\leqslant d}\max_{\substack{1\leqslant k\leqslant q_i\\1\leqslant \ell\leqslant q_k}}\left(\mathbb{E}[I_k(f_{i,k})^2]\kappa_4(I_\ell(f_{j,\ell}))^{1/2}\right).$$

Thus, it follows that

$$\max_{1\leqslant i,j\leqslant d}\|\Delta_{i,j}\|_2\leqslant \max_{1\leqslant i,j\leqslant d}\sum_{k=1}^{q_i}\sum_{\substack{\ell=1\\(k,\ell)\neq(1,1)}}^{q_j}\Big\{2\widetilde{K}_q\max_{1\leqslant i\leqslant d}\max_{1\leqslant k\leqslant q_i}\kappa_4(I_k(f_{i,k}))$$
$$+2\max_{1\leqslant i,j\leqslant d}\max_{\substack{1\leqslant k\leqslant q_i\\1\leqslant \ell\leqslant q_j}}\left[\mathbb{E}[I_k(f_{i,k})^2]\kappa_4(I_\ell(f_{j,\ell}))^{1/2}\right]\mathbb{1}_{\{k\neq\ell\}}\Big\}^{1/2},$$

and as square root is sub-additive, and $q_i,q_j\leqslant q$, it further follows that

$$\max_{1\leqslant i,j\leqslant d}\|\Delta_{i,j}\|_2\leqslant \max_{1\leqslant i,j\leqslant d}q_iq_j\Big\{(2\widetilde{K}_q)^{1/2}\max_{1\leqslant i\leqslant d}\max_{1\leqslant k\leqslant q_i}(\kappa_4(I_k(f_{i,k})))^{1/2}$$
$$+\sqrt{2}\max_{1\leqslant i,j\leqslant d}\max_{\substack{1\leqslant k\leqslant q_i\\1\leqslant \ell\leqslant q_\ell}}\left[\mathbb{E}[I_k(f_{i,k})^2]\kappa_4(I_\ell(f_{j,\ell}))^{1/2}\right]^{1/2}\mathbb{1}_{\{k\neq\ell\}}\Big\}$$
$$\leqslant q^2\max\{\sqrt{2},(2\widetilde{K}_q)^{1/2}\}\Big\{\max_{1\leqslant i\leqslant d}\max_{1\leqslant k\leqslant q_i}(\kappa_4(I_k(f_{i,k})))^{1/2}$$
$$+\max_{1\leqslant i,j\leqslant d}\max_{\substack{1\leqslant k\leqslant q_i\\1\leqslant \ell\leqslant q_j}}\left[\mathbb{E}[I_k(f_{i,k})^2]\kappa_4(I_\ell(f_{j\ell}))^{1/2}\right]^{1/2}\mathbb{1}_{\{k\neq\ell\}}\Big\}.$$

Let $\overline{K}_q:=q^2\max\{\sqrt{2},(2\widetilde{K}_q)^{1/2}\}$, and combine the obtained bound on $\max_{1\leqslant i,j\leqslant d}\|\Delta_{i,j}\|_2$ with (4.63) to obtain

$$\Delta_{\boldsymbol{F}}\leqslant \widetilde{C}_q\log^{q-1}(d)\Big\{\max_{1\leqslant i\leqslant d}\max_{1\leqslant k\leqslant q_i}(\kappa_4(I_k(f_{i,k})))^{1/2}$$
$$+\max_{1\leqslant i,j\leqslant d}\max_{\substack{1\leqslant k\leqslant q_i\\1\leqslant \ell\leqslant q_j}}\left[\mathbb{E}[I_k(f_{i,k})^2]\kappa_4(I_\ell(f_{j,\ell}))^{1/2}\right]^{1/2}\mathbb{1}_{\{k\neq\ell\}}\Big\},$$

where $\widetilde{C}_q$ is such that $M_{2q-2}\overline{K}_q \log^{q-1}(2d^2-1+e^{q-2}) \leqslant \widetilde{C}_q \log^{q-1}(d)$. For the existence of such a constant $\widetilde{C}_q$ see (4.66). Define

$$\overline{\Delta}_{\boldsymbol{F}} := \max_{1\leqslant i\leqslant d}\max_{1\leqslant k\leqslant q_i}(\kappa_4(I_k(f_{i,k})))^{1/2} + \max_{1\leqslant i,j\leqslant d}\max_{\substack{1\leqslant k\leqslant q_i\\ 1\leqslant \ell\leqslant q_j}}\left[\mathbb{E}[I_k(f_{i,k})^2]\kappa_4(I_\ell(f_{j,\ell}))^{1/2}\right]^{1/2}\mathbb{1}_{\{k\neq\ell\}}.$$

Hence $\Delta_{\boldsymbol{F}} \leqslant \widetilde{C}_q \log^{q-1}(d)\overline{\Delta}_{\boldsymbol{F}}$, and the remainder of the proof follows as in Theorem 3.1 (from (4.68) onwards). This completes the proof. □

4.7 Proof of Theorem 3.4

In order to prove Theorem 3.4, we use the ensuing lemma, which is proven in Appendix A, below.

Lemma 4.7. *Let $I := \{(i,j,r) \in \mathbb{N}^3 : 1 \leqslant i,j \leqslant d, 1 \leqslant r \leqslant p\}$. Define $\lambda : I \to \{1,2,3,\ldots,d^2p\}$ by $\lambda(i,j,r) = (r-1)d^2 + (j-1)d + i$. Then λ is a bijection.*

Proof of Theorem 3.4. Begin with re-writing $\boldsymbol{F}_n$ as follows

$$\boldsymbol{F}_n = n^{-1/2}(\boldsymbol{Z}\otimes\boldsymbol{I}_d)\boldsymbol{\varpi} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\begin{pmatrix} y_{k-1}^{(1)}\boldsymbol{u}_k \\ \vdots \\ y_{k-1}^{(d)}\boldsymbol{u}_k \\ \vdots \\ y_{k-p}^{(1)}\boldsymbol{u}_k \\ \vdots \\ y_{k-p}^{(d)}\boldsymbol{u}_k \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Phi\begin{pmatrix}\boldsymbol{u}_k \\ \boldsymbol{y}_{k-1} \\ \vdots \\ \boldsymbol{y}_{k-p}\end{pmatrix} =: \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Phi\left(\boldsymbol{G}_k\right). \tag{4.73}$$

Then $\boldsymbol{G}_k = (G_k^{(a)})_{1\leqslant a\leqslant d(p+1)}$ is a Gaussian sequence due to (3.2), i.e.

$$\boldsymbol{G}_k = \begin{pmatrix}\boldsymbol{u}_k \\ \boldsymbol{Z}_{k-1}\end{pmatrix} \sim \mathcal{N}_{d(p+1)}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_G\right), \quad \text{where } \boldsymbol{\Sigma}_G := \begin{pmatrix}\boldsymbol{\Sigma}_u & \boldsymbol{0}_{d\times dp} \\ \boldsymbol{0}_{dp\times d} & \boldsymbol{\Sigma}_Z\end{pmatrix}, \tag{4.74}$$

where, recall, $\boldsymbol{Z}_k = (\boldsymbol{y}_k, \boldsymbol{y}_{k-1}, \ldots, \boldsymbol{y}_{k-p+1})^\top \in \mathbb{R}^{dp}$ is the stacked vector of p-lags and $\boldsymbol{\Sigma}_Z = \mathrm{Cov}(\boldsymbol{Z}_k, \boldsymbol{Z}_k)$ for $k \in \mathbb{Z}$ is its covariance matrix. Next, denote $\rho^{(a,b)}(k-\ell) := \mathbb{E}[G_k^{(a)}G_\ell^{(b)}]$ and note that $\rho^{(a,b)}(k-\ell) = \rho^{(b,a)}(\ell-k)$ for all $a,b \in \{1,\ldots,d(p+1)\}$ and $k,\ell \in \mathbb{N}$. Recall the definition of $\boldsymbol{\Gamma}_G$,

$$\boldsymbol{\Gamma}_G(\ell) := \begin{cases}\mathrm{Cov}(\boldsymbol{G}_k, \boldsymbol{G}_{k-\ell}) = (\rho^{(i,j)}(\ell))_{i,j\in\{1,\ldots,d(p+1)\}}, & \ell > 0 \\ \boldsymbol{\Gamma}_G(-\ell)^\top, & \ell < 0,\end{cases} \quad \text{and} \quad \boldsymbol{\Gamma}_G(0) = \boldsymbol{\Sigma}_G.$$

Define $I := \{(i,j,r) \in \mathbb{N}^3 : 1 \leqslant i,j \leqslant d, 1 \leqslant r \leqslant p\}$, with exactly d^2p distinct elements. Next, by Lemma 4.7, the function $\lambda : I \to \{1,2,3,\ldots,d^2p\}$ given by $\lambda(i,j,r) = (r-1)d^2 + (j-1)d + i$ is a bijection. Therefore, to each triple $(i,j,r) \in I$ we can assign a unique element

$\lambda(i,j,r) = v \in \{1,\ldots,d^2p\}$, and the other way around. Hence, for $\Phi = (\phi_1,\ldots,\phi_{d^2p})$, its coordinates $\phi_v : \mathbb{R}^{2d} \to \mathbb{R}$ for $v \in \{1,\ldots,d^2p\}$ are given as

$$\phi_v(\boldsymbol{z}) = z^{(i)}z^{(rd+j)} = H_1(z^{(i)})H_1(z^{(rd+j)}), \quad i,j \in \{1,\ldots,d\},\ r \in \{1,\ldots,p\} \tag{4.75}$$

for $\boldsymbol{z} \in \mathbb{R}^{d(p+1)}$, and where there are unique $i,j \in \{1,\ldots,d\}$ and $r \in \{1,\ldots,p\}$ such that $(i,j,r) = \lambda^{-1}(v)$.Since ϕ_v has Hermite rank 2, we want to use Theorem 3.1. Hence, we show how to represent $\boldsymbol{S}_n$ via multiple integrals, in order to apply Theorem 3.1. Recall from (4.73), that $\boldsymbol{S}_n = n^{-1/2}\sum_{k=1}^n \Phi(\boldsymbol{G}_k)$. As seen in the beginning of the proof of Theorem 2.3, let $\mathfrak{H}$ be a separable Hilbert space and $\boldsymbol{X}$ an isonormal process, such that $\boldsymbol{G}_k^{(i)} \overset{\mathcal{D}}{=} X(e_{k,i})$ with the covariance structure $\langle e_{k,i}, e_{k',i'}\rangle_{\mathfrak{H}} = \rho^{(i,i')}(k-k')$ for $i,i' \in \{1,\ldots d(p+1)\}$ and $k,k' \in \mathbb{N}$, for for centred Gaussian sequence $(\boldsymbol{G}_k)_{k\in\mathbb{N}}$ from (4.74).

To study the structure of the components of $\boldsymbol{S}_n = (S_{n,1},\ldots,S_{n,d^2p})$, fix $v \in \{1,\ldots,d^2p\}$ and denote the corresponding $(i,j,r) \in I$, according to Lemma 4.7. By (4.75), it holds that

$$S_{n,v} = \frac{1}{\sqrt{n}}\sum_{k=1}^n G_k^{(i)}G_k^{(rd+j)} = \frac{1}{\sqrt{n}}\sum_{k=1}^n H_1(X(e_{k,i}))H_1(X(e_{k,rd+j})). \tag{4.76}$$

Next, we show that $S_{n,v} \overset{\mathcal{D}}{=} I_2(f(n,v))$, for some appropriate kernel $f(n,v) \in \mathfrak{H}^{\odot 2}$. Due to [39, Thm 2.7.7] (which also holds for H_1 in our setting), (4.76) and [39, Thm 2.7.10], it holds that

$$S_{n,v} \overset{\mathcal{D}}{=} \frac{1}{\sqrt{n}}\sum_{k=1}^n I_1(e_{k,i})I_1(e_{k,rd+j}) = \frac{1}{\sqrt{n}}\sum_{k=1}^n\sum_{s=0}^1 s!\binom{1}{s}^2 I_{2-2s}(e_{k,i}\widetilde{\otimes}_s e_{k,rd+j}).$$

Note, that $\binom{1}{s} = 1$ for $s = 0,1$ and $s! = 1$ for $s = 0,1$. Hence, by (4.74), it follows that

$$I_0(e_{k,i}\widetilde{\otimes}_1 e_{k,rd+j}) = e_{k,i}\widetilde{\otimes}_1 e_{k,rd+j} = \langle e_{k,i}, e_{k,rd+j}\rangle_{\mathfrak{H}} = \rho^{(i,rd+j)}(0) = (\boldsymbol{\Sigma}_G)_{i,rd+j} = 0,$$

since $i \in \{1,\ldots,d\}$ and $rd+j \in \{d+1,\ldots d(p+1)\}$. Thus,

$$S_{n,v} \overset{\mathcal{D}}{=} I_2(f_{n,v}), \quad \text{with} \quad f_{n,v} = \frac{1}{2\sqrt{n}}\sum_{k=1}^n \left(e_{k,i}\otimes e_{k,rd+j} + e_{k,rd+j}\otimes e_{k,i}\right). \tag{4.77}$$

For the remainder of the proof, take v and v' from $\{1,\ldots,d^2p\}$ and let $i,j,i',j' \in \{1,\ldots,d\}$ and $r,r' \in \{1,\ldots,p\}$ be the corresponding elements form I, according to Lemma 4.7. To apply Theorem 3.1 with $q = 2$, we calculate $\|f_{n,v}\otimes_1 f_{n,v'}\|^2_{\mathfrak{H}^{\otimes 2}}$. Using linearity of the contraction operator, together with (4.7) and that $\langle e_{k,i}, e_{k',i'}\rangle_{\mathfrak{H}} = \rho^{(i,i')}(k-k')$ for $i,i' \in \{1,\ldots d(p+1)\}$ and $k,k' \in \mathbb{N}$, we get

$$\begin{aligned}
f_{n,v}\otimes_1 f_{n,v'} &= \frac{1}{4n}\sum_{k,k'=1}^n (e_{k,i}\otimes e_{k,rd+j} + e_{k,rd+j}\otimes e_{k,i})\otimes_1 (e_{k',i'}\otimes e_{k',r'd+j'} + e_{k',r'd+j'}\otimes e_{k',i'}) \\
&= \frac{1}{4n}\sum_{k,k'=1}^n \Big(\rho^{(rd+j,r'd+j')}(k-k')e_{k,i}\otimes e_{k',i'} + \rho^{(rd+j,i')}(k-k')e_{k,i}\otimes e_{k',r'd+j'} \\
&\quad + \rho^{(i,r'd+j')}(k-k')e_{k,rd+j}\otimes e_{k',i'} + \rho^{(i,i')}(k-k')e_{k,rd+j}\otimes e_{k',r'd+j'}\Big).
\end{aligned}$$

Expanding the definition of $f_{n,v} \otimes_1 f_{n,v'}$ yields a sum of 16 terms of the form

$$\rho^{(\cdot,\cdot)}(k-k')\rho^{(\cdot,\cdot)}(a-a')\rho^{(\cdot,\cdot)}(k-a)\rho^{(\cdot,\cdot)}(k'-a'), \qquad k,k',a,a' \in \{1,\dots,n\}. \tag{4.78}$$

This can be written compactly as,

$$\begin{aligned}\|f_{n,v} \otimes_1 f_{n,v'}\|^2_{\mathfrak{H}^{\otimes 2}} \\ = \frac{1}{16n^2} \sum_{k,k',a,a'=1}^{n} \sum_{\substack{\alpha_m \in R \\ m \in \{1,2,3,4\}}} \rho^{\alpha_1}(k-k')\rho^{\alpha_2}(a-a')\rho^{\alpha_3}(k-a)\rho^{\alpha_4}(k'-a'),\end{aligned} \tag{4.79}$$

where $R = \{(u,v) : u,v \in \{i,i',rd+j,r'd+j'\}\}$. Recall that, $i,j,i',j' \in \{1,\dots,d\}$ and $r,r' \in \{1,\dots,p\}$. Now, Lemma 3.8 gives us an explicit from of covariances of the type $\rho^{(u,v)}(x,y)$ in (4.79). Indeed, for $\mathbf{\Gamma}_Z(\ell) = \sum_{i=0}^{\infty} \boldsymbol{A}^i \mathbf{\Sigma}_U (\boldsymbol{A}^\top)^i$,

$$\mathbb{R}^{d(p+1)\times d(p+1)} \ni \mathbf{\Gamma}_G(\ell) = \begin{cases} \begin{pmatrix} \mathbf{0}_{d\times d} & \mathbf{0}_{d\times dp} \\ \boldsymbol{A}^{\ell-1}\mathbf{\Sigma}_U \boldsymbol{J}^\top & \mathbf{\Gamma}_Z(\ell) \end{pmatrix}, & \ell > 0, \\ \mathbf{\Sigma}_G = \begin{pmatrix} \mathbf{\Sigma}_u & \mathbf{0}_{d\times dp} \\ \mathbf{0}_{dp\times d} & \mathbf{\Sigma}_y. \end{pmatrix} & \ell = 0. \end{cases}$$

This explicit structure of $\mathbf{\Gamma}_G$ determines the terms of the form (4.78), as $\rho^{(u,v)}(x-y)$ is an entry in $\mathbf{\Gamma}_G(\ell)$, at with coordinates (u,v) where $\ell = x-y$, and $u,v \in \{1,\dots,d(p+1)\}$. By Lemma 3.8 and the symmetry $\rho^{(u,v)}(\ell) = \rho^{(v,u)}(-\ell)$, for $\ell \in \mathbb{Z}$, we have the following consequences of the block form of $\mathbf{\Gamma}_G(\ell)$:

$$\text{(i)} \quad \rho^{(i,i')}(\ell) = 0 \text{ for } \ell \neq 0, \qquad \rho^{(i,i')}(0) = (\mathbf{\Sigma}_u)_{i,i'}, \tag{4.80}$$

$$\text{(ii)} \quad \rho^{(i,r'd+j')}(\ell) = 0 \text{ for } \ell \geqslant 0, \tag{4.81}$$

$$\text{(iii)} \quad \rho^{(rd+j,i')}(\ell) = 0 \text{ for } \ell \leqslant 0. \tag{4.82}$$

Moreover, for $u,u' \in \{1,\dots,d\}$ and all $x,y \in \mathbb{Z}$,

$$|\rho^{(u,u')}(x-y)| \leqslant \|\mathbf{\Sigma}_u\|_{\max} \mathbb{1}_{\{x=y\}}. \tag{4.83}$$

Indeed, (4.80) holds because the upper-left $d \times d$ block of $\mathbf{\Gamma}_G(\ell)$ is zero for $\ell \neq 0$ and equals $\mathbf{\Sigma}_u$ at $\ell = 0$. Moreover, (4.81) follows since the upper-right block of $\mathbf{\Gamma}_G(\ell)$ is zero for all $\ell > 0$, and at $\ell = 0$ it is also zero by the definition of $\mathbf{\Sigma}_G$. Finally, (4.82) follows from (4.81) by symmetry: $\rho^{(rd+j,i')}(\ell) = \rho^{(i',rd+j)}(-\ell)$.

Hence, we determine $\rho^{(u,v)}(x-y)$ based on which part of $\mathbf{\Gamma}_G(x-y)$ it is. In what follows, we go through all 16 terms of the form (4.78) to bound (4.79).

Step 1. First, using (4.81), and the symmetry $\rho^{(u,v)}(x-y) = \rho^{(v,u)}(y-x)$, yields

$$\rho^{(rd+j,i')}(k-k')\rho^{(i,r'd+j')}(a-a')\rho^{(r'd+j',i')}(k'-a')\rho^{(i,rd+j)}(k-a) = 0, \tag{4.84}$$

$$\rho^{(i,r'd+j')}(k-k')\rho^{(rd+j,i')}(a-a')\rho^{(rd+j,i)}(k-a)\rho^{(i',r'd+j')}(k'-a') = 0. \tag{4.85}$$

Indeed, in (4.84), $\rho^{(rd+j,i')}(k-k') \neq 0$, when $k > k'$; $\rho^{(i,r'd+j')}(a-a') \neq 0$, when $a < a'$, and $\rho^{(r'd+j',i')}(k'-a') \neq 0$, when $k' > a'$. Hence, we have obtained that $k > k' > a' > a$, specially that $k > a$. However, when $k > a$ the final term, $\rho^{(i,rd+j)}(k-a) = 0$. Hence the whole expression (4.84) is zero. Same logic applies to (4.85).

Step 2. After eliminating the identically zero terms in **Step 1**, each remaining (non-trivial) summand in the expansion of $\|f_{n,v} \otimes_1 f_{n,v'}\|^2_{\mathfrak{H}^{\otimes 2}}$ is of the form

$$\frac{1}{16n^2} \sum_{k,k',a,a'=1}^{n} \rho^{\beta_1}(k-k')\rho^{\beta_2}(a-a')\rho^{\beta_3}(k-a)\rho^{\beta_4}(k'-a'), \qquad \beta_m \in R, \tag{4.86}$$

where $R = \{(u,v) : u,v \in \{i,i',rd+j,r'd+j'\}\}$. Moreover, in each non-trivial term (4.86) there is at least one factor $\rho^{(u,v)}(\cdot)$ with $u,v \in \{i,i'\}$ (i.e., both indices in the $\mathbf{\Sigma}_u$-block of $\mathbf{\Gamma}_G$). Applying (4.83), this factor can be bounded by $\|\mathbf{\Sigma}_u\|_{\max}\mathbb{1}_{\{\cdot=0\}}$, which collapses one of the four products.

Concretely, if the $\mathbf{\Sigma}_u$-block factor is $\rho^{(u,v)}(a-a')$ (with $u,v \in \{i,i'\}$), then by (4.83) we have

$$\sum_{a,a'=1}^{n} |\rho^{(u,v)}(a-a')| X_{a,a'} \leqslant \|\mathbf{\Sigma}_u\|_{\max} \sum_{a=1}^{n} X_{a,a'},$$

and hence the corresponding term (4.86) is bounded by

$$\frac{\|\mathbf{\Sigma}_u\|_{\max}}{16n^2} \sum_{k,k',a=1}^{n} \left|\rho^{\beta_1}(k-k')\rho^{\beta_3}(k-a)\rho^{\beta_4}(k'-a)\right|.$$

The same argument applies if the $\mathbf{\Sigma}_u$-block factor is $\rho^{(u,v)}(k-a)$, $\rho^{(u,v)}(k'-a')$, or $\rho^{(u,v)}(k-k')$, collapsing the corresponding index.

Finally, by the symmetry $\rho^{(u,v)}(\ell) = \rho^{(v,u)}(-\ell)$ for integer lags ℓ, we may rewrite each remaining factor so that its argument lies in $\{k-k', k-a, k'-a\}$, possibly after swapping the ordered pair, which remains in R. Therefore, there exists a subset $J' := R\backslash(\{i,i'\} \times \{i,i'\}) \subset R$, such that every non-trivial term is bounded by an expression of the form

$$\frac{\|\mathbf{\Sigma}_u\|_{\max}}{16n^2} \sum_{k,k',a=1}^{n} \sum_{\substack{\alpha_m \in J' \\ m=1,2,3}} \rho^{\alpha_1}(k-k')\rho^{\alpha_2}(k-a)\rho^{\alpha_3}(k'-a).$$

Summing over the remaining terms in (4.79), yields

$$\|f_{n,v} \otimes_1 f_{n,v'}\|^2_{\mathfrak{H}^{\otimes 2}} \leqslant \frac{1}{16n^2}\|\mathbf{\Sigma}_u\|_{\max} \sum_{k,k',a=1}^{n} \sum_{\substack{\alpha_m \in J' \\ m\in\{1,2,3\}}} \rho^{\alpha_1}(k-k')\rho^{\alpha_2}(k-a)\rho^{\alpha_3}(k'-a),$$

where $J' \subset R$. Using substitutions $t = k-a$ and $s = k'-a$, one gets

$$\begin{aligned}
\sum_{k,k',a=1}^{n} \sum_{\substack{\alpha_m \in J' \\ m\in\{1,2,3\}}} \rho^{\alpha_1}(k-k')\rho^{\alpha_2}(k-a)\rho^{\alpha_3}(k'-a) &\leqslant \sum_{|t|<n} \rho^{\alpha_2}(t) \sum_{|s|<n} \sum_{a'=1}^{n} \rho^{\alpha_3}(s)\rho^{\alpha_1}(t-s) \\
&\leqslant n \sum_{|t|<n} \rho^{\alpha_2}(t)(\rho_n^{\alpha_3} * \rho_n^{\alpha_1})(t) = n\|\rho_n^{\alpha_2}(\rho_n^{\alpha_3} * \rho_n^{\alpha_1})\|_{\ell^1(\mathbb{Z})} \leqslant n\|\rho_n^{\alpha_2}\|_{\ell^1(\mathbb{Z})}\|\rho_n^{\alpha_3} * \rho_n^{\alpha_1}\|_{\ell^1(\mathbb{Z})} \\
&\leqslant n\|\rho_n^{\alpha_1}\|_{\ell^1(\mathbb{Z})}\|\rho_n^{\alpha_2}\|_{\ell^1(\mathbb{Z})}\|\rho_n^{\alpha_3}\|_{\ell^1(\mathbb{Z})} \leqslant n\Big(\sum_{|\ell|<n} \|\mathbf{\Gamma}_G(\ell)\|_{\max}\Big)^3,
\end{aligned}$$

where $\|M\|_{\max} = \max_{1\leqslant i,j\leqslant d}|M_{i,j}|$, for a matrix $M \in \mathbb{R}^{d\times d}$ and recalling that, for a sequence $s = (s_k)_{k\in\mathbb{Z}}$ and $p \geqslant 1$, $\|s\|^p_{\ell^p(\mathbb{Z})} = \sum_{k\in\mathbb{Z}}|s_k|^p$, and that $\rho_n^{(u,u)}(k) = |\rho^{(u,v)}(k)|\mathbb{1}_{\{|k|<n\}}$. It was also used that we can bound any of the covariances $\rho^{\alpha_i}(\ell)$ with the maximal entry of the matrix $\mathbf{\Gamma}_G(\ell)$. Hence we obtain

$$\|f_{n,v} \otimes_1 f_{n,v'}\|_{\mathfrak{H}^{\otimes 2}} \leqslant \frac{\sqrt{7}}{\sqrt{8n}}\|\Sigma_u\|^{1/2}_{\max}\Big(\sum_{|\ell|<n}\|\mathbf{\Gamma}_G(\ell)\|_{\max}\Big)^{3/2}. \tag{4.87}$$

Next, we observe that $\mathrm{Cov}(\boldsymbol{S}_n) = \mathrm{Cov}(\boldsymbol{N})$. Indeed, by (4.77) it follows that

$$\begin{aligned}
\mathbb{E}[\boldsymbol{S}_{n,v}\boldsymbol{S}_{n,v'}] &= \mathbb{E}[I_2(f(n,v))I_2(f(n,v'))] = 2\langle f(n,v), f(n,v')\rangle_{\mathfrak{H}^{\otimes 2}} \\
&= \frac{1}{2n}\sum_{k,k'=1}^{n}\Big\langle e_{k,i}\otimes e_{k,rd+j} + e_{k,rd+j}\otimes e_{k,i}, e_{k',i'}\otimes e_{k',r'd+j'} + e_{k',r'd+j'}\otimes e_{k,i'}\Big\rangle_{\mathfrak{H}^{\otimes 2}} \\
&= \frac{1}{n}\sum_{k,k'=1}^{n}\langle e_{k,i}, e_{k',i'}\rangle_{\mathfrak{H}}\langle e_{k,rd+j}, e_{k',r'd+j'}\rangle_{\mathfrak{H}} + \langle e_{k,i}, e_{k',r'd+j'}\rangle_{\mathfrak{H}}\langle e_{k,rd+j,e_{k',i'}}\rangle_{\mathfrak{H}} \\
&= \frac{1}{n}\sum_{k,k'=1}^{n}\rho^{(i,i')}(k-k')\rho^{(rd+j,r'd+j')}(k-k') \\
&\quad + \frac{1}{n}\sum_{k,k'=1}^{n}\rho^{(i,r'd+j')}(k-k')\rho^{(rd+j,i')}(k-k').
\end{aligned}$$

To evaluate the two sums, set $\ell := k-k'$. Using (4.80), the first sum reduces to the diagonal $k = k'$:

$$\begin{aligned}
\frac{1}{n}\sum_{k,k'=1}^{n}\rho^{(i,i')}(k-k')\rho^{(rd+j,r'd+j')}(k-k') &= \frac{1}{n}\sum_{k=1}^{n}\rho^{(i,i')}(0)\rho^{(rd+j,r'd+j')}(0) \\
&= (\mathbf{\Sigma}_u)_{i,i'}(\mathbf{\Sigma}_Y)_{rd+j,r'd+j'}.
\end{aligned}$$

For the second sum, for each pair (k,k') with $\ell = k-k'$ we have: if $\ell \geqslant 0$ then $\rho^{(i,r'd+j')}(\ell) = 0$ by (4.81), while if $\ell \leqslant 0$ then $\rho^{(rd+j,i')}(\ell) = 0$ by (4.82). Hence the product is identically zero for all k, k' and

$$\frac{1}{n}\sum_{k,k'=1}^{n}\rho^{(i,r'd+j')}(k-k')\rho^{(rd+j,i')}(k-k') = 0.$$

Hence, we obtain that

$$\mathbb{E}[\boldsymbol{S}_{n,v}\boldsymbol{S}_{n,v'}] = (\mathbf{\Sigma}_u)_{i,i'}(\mathbf{\Sigma}_Y)_{rd+j,r'd+j'} = (\mathbf{\Sigma}_Z\otimes\mathbf{\Sigma}_u)_{v,v'}.$$

Indeed, note that for Kronecker product it holds that

$$(\mathbf{\Sigma}_Z)_{\alpha,\beta}(\mathbf{\Sigma}_u)_{i,i'} = (\mathbf{\Sigma}_Z\otimes\mathbf{\Sigma}_u)_{(\alpha-1)d+i,(\beta-1)d+i'},$$

and in our case we have $\alpha = rd+j$ and $\beta = r'd+j'$, which substituting back yields that

$$(\mathbf{\Sigma}_Z)_{rd+j,r'd+j'}(\mathbf{\Sigma}_u)_{i,i'} = (\mathbf{\Sigma}_Z\otimes\mathbf{\Sigma}_u)_{((rd+j)-1)d+i,((r'd+j')-1)d+i'} = (\mathbf{\Sigma}_Z\otimes\mathbf{\Sigma}_u)_{v,v'},$$

where we recall bijection λ from Lemma 4.7. All in all, we have obtained that $\operatorname{Cov}(\boldsymbol{S}_n) = \operatorname{Cov}(\boldsymbol{N})$, and we can use Theorem 3.1 with $q = 2$ together with (4.87) to get

$$d_{\mathcal{R}}(\boldsymbol{S}_n, \boldsymbol{N}) \leqslant C \log^2(d^2p) \log_+ \log(d^2p) \Delta_{\boldsymbol{S}_n} \log_+(\Delta_{\boldsymbol{S}_n}) \frac{\log_+(\sigma_*)}{\sigma_*}, \tag{4.88}$$

$$d_{\mathscr{C}}(\boldsymbol{S}_n, \boldsymbol{N}) \leqslant C(d^2p)^{65/24} \Delta_{\boldsymbol{S}_n} \frac{1}{\sigma_*(\boldsymbol{\Lambda})^{3/2}}, \quad \text{and}$$

$$d_{\mathcal{W}}(\boldsymbol{S}_n, \boldsymbol{N}) \leqslant C(d^2p)^{3/2} \Delta_{\boldsymbol{S}_n} \frac{\sigma_*(\boldsymbol{\Sigma})^{1/2}}{\sigma_*(\boldsymbol{\Sigma})}.$$

where C, and $\Delta_{\boldsymbol{S}_n} := \sqrt{7}(\sqrt{8n})^{-1} \|\Sigma_u\|_{\max}^{1/2} \Big(\sum_{|\ell|<n} \|\boldsymbol{\Gamma}_G(\ell)\|_{\max} \Big)^{3/2}$. Using (4.1), together with (4.88), conclude the proof. □

Proof of Lemma 3.8. **Part 1.** Let $p = 1$. We start by considering the case $\ell = 0$, in which case $\boldsymbol{\Sigma}_G = \operatorname{Cov}(\boldsymbol{G}_k, \boldsymbol{G}_k) = \operatorname{Var}(\boldsymbol{G}_k)$, and hence the desired form of $\boldsymbol{\Sigma}_G$ follows from (4.74). That $\boldsymbol{\Sigma}_y = \sum_{i=0}^{\infty} \boldsymbol{A}_1^i \boldsymbol{\Sigma}_u (\boldsymbol{A}_1^i)^\top$ holds, follows directly from [34, Eq. (2.1.6)]. Next, assume that $\ell > 0$. And observe the following 4 exhaustive cases.

Case 1.1. Assume that $i, j \in \{1, \ldots, d\}$. Since $(\boldsymbol{u}_k)_{k\in\mathbb{N}}$ is assumed to be white noise with a non-singular covariance matrix, it follows that $\boldsymbol{u}_k$ is uncorrelated across time, i.e.

$$(\boldsymbol{\Gamma}_G(\ell))_{i,j} = \operatorname{Cov}(G_k^{(i)}, G_{k-\ell}^{(j)}) = \operatorname{Cov}(u_k^{(i)}, u_{k-\ell}^{(j)}) = 0.$$

Case 1.2. Assume that $i, j \in \{d+1, \ldots, 2d\}$. Then, by [34, Eq. (2.1.6)],

$$(\boldsymbol{\Gamma}_G(\ell))_{i,j} = \operatorname{Cov}(y_{k-1}^{(i-d)}, y_{k-1-\ell}^{(j-d)}) = \big(\mathbb{E}[\boldsymbol{y}_{k-1}\boldsymbol{y}_{k-1-\ell}^\top]\big)_{i-d,j-d} = \Big(\boldsymbol{A}^\ell \boldsymbol{\Sigma}_y\Big)_{i-d,j-d}.$$

Case 1.3. Assume that $i \in \{1, \ldots, d\}$, $j \in \{d+1, \ldots, 2d\}$. Then, since $\ell > 0$, it follows that

$$(\boldsymbol{\Gamma}_G(\ell))_{i,j} = \operatorname{Cov}(u_k^{(i)}, y_{k-1-\ell}^{(j-d)}) = 0.$$

Case 1.4. Let $i, j \in \{1, \ldots, d\}$, then $(\boldsymbol{\Gamma}_G(\ell))_{i+d,j} = \operatorname{Cov}(y_{k-1}^{(i)}, u_{k-\ell}^{(j)})$. To find the form of the covariance, it suffices to show that

$$\operatorname{Cov}(\boldsymbol{y}_{k-1}, \boldsymbol{u}_{k-\ell}) = \boldsymbol{A}_1^{\ell-1} \boldsymbol{\Sigma}_u. \tag{4.89}$$

In order to prove (4.89), assume for induction $\ell = 1$. Then

$$\operatorname{Cov}(\boldsymbol{y}_{k-1}, \boldsymbol{u}_{k-1}) = \operatorname{Cov}(\boldsymbol{A}_1 \boldsymbol{y}_{k-2} + \boldsymbol{u}_{k-1}, \boldsymbol{u}_{k-1}) = \operatorname{Cov}(\boldsymbol{u}_{k-1}, \boldsymbol{u}_{k-1}) = \boldsymbol{\Sigma}_u.$$

Next, we assume that (4.89) holds for some $\ell > 1$, and prove that it holds for $\ell + 1$:

$$\operatorname{Cov}(\boldsymbol{y}_{k-1}, \boldsymbol{u}_{k-\ell-1}) = \operatorname{Cov}(\boldsymbol{A}_1 \boldsymbol{y}_{k-2} + \boldsymbol{u}_{k-1}, \boldsymbol{u}_{k-\ell-1}) = \boldsymbol{A}_1 \operatorname{Cov}(\boldsymbol{y}_{k-2}, \boldsymbol{u}_{k-\ell-1}).$$

Hence, by induction assumption, we can conclude (4.89), which concludes the proof:

$$\operatorname{Cov}(\boldsymbol{y}_{k-1}, \boldsymbol{u}_{k-\ell-1}) = \boldsymbol{A}_1 \boldsymbol{A}_1^{\ell-1} \boldsymbol{\Sigma}_u = \boldsymbol{A}_1^\ell \boldsymbol{\Sigma}_u.$$

Part 2. Now assume $p > 1$. First note that $\boldsymbol{\Gamma}_Y(\ell) = \sum_{i=0}^{\infty} \boldsymbol{A}^{\ell+i} \boldsymbol{\Sigma}_U (\boldsymbol{A}^i)^\top$ follows directly from [34, Eq. (2.1.10)]. Let $\ell > 0$ and split into the following exhaustive cases. The proof is in four steps, one for each sub-matrix of $\boldsymbol{\Gamma}_G(\ell)$. Fix $i, j \in \{1, \ldots, d\}$ and $r \in \{1, \ldots, p\}$.

Case 2.1. Consider entries in the top left part: $(\boldsymbol{\Gamma}_G(\ell))_{i,j} = 0$, with the same arguments as in **Case 1.1**.
Case 2.2. For the entries in the top right part: by the definition, it follows that $(\boldsymbol{\Gamma}_G(\ell))_{i,j+rd} = \mathrm{Cov}(u_k^{(i)}, y_{k-\ell-r}^{(j)}) = 0$, as $\ell + r > 0$.
Case 2.3. For the entries in the bottom left part: Then $(\boldsymbol{\Gamma}_G(\ell))_{rd+j,i} = \mathrm{Cov}(y_{k-r}^{(i)}, u_{k-\ell}^{(j)})$. To find the form of the covariance, we start by showing that

$$\mathrm{Cov}(\boldsymbol{Y}_{k-1}, \boldsymbol{U}_{k-\ell}) = \boldsymbol{A}^{\ell-1}\boldsymbol{\Sigma}_U. \tag{4.90}$$

Equation (4.90) can be shown using induction and the fact that $\boldsymbol{Y}_k$ is a VAR(1) process as in (3.8). Indeed, for $\ell = 1$, we see directly

$$\mathrm{Cov}(\boldsymbol{Y}_{k-1}, \boldsymbol{U}_{k-1}) = \mathrm{Cov}(\boldsymbol{A}\boldsymbol{Y}_{k-2} + \boldsymbol{U}_{k-1}, \boldsymbol{U}_{k-1}) = \mathrm{Cov}(\boldsymbol{U}_{k-1}, \boldsymbol{U}_{k-1}) = \boldsymbol{\Sigma}_U.$$

Next, assume that (4.90) holds for $\ell > 1$, and that it holds for $\ell + 1$, that

$$\mathrm{Cov}(\boldsymbol{Y}_{k-1}, \boldsymbol{U}_{k-\ell-1}) = \mathrm{Cov}(\boldsymbol{A}\boldsymbol{Y}_{k-2} + \boldsymbol{U}_{k-1}, \boldsymbol{U}_{k-\ell-1}) = \boldsymbol{A}\mathrm{Cov}(\boldsymbol{Y}_{k-2}, \boldsymbol{U}_{k-\ell-1}).$$

Hence, by induction assumption, it follows that

$$\mathrm{Cov}(\boldsymbol{Y}_{k-1}, \boldsymbol{U}_{k-\ell-1}) = \boldsymbol{A}\boldsymbol{A}^{\ell-1}\boldsymbol{\Sigma}_U = \boldsymbol{A}^{\ell}\boldsymbol{\Sigma}_U,$$

concluding (4.90).

Let $\boldsymbol{e}_k := (\boldsymbol{0} : \cdots : \boldsymbol{I}_d : \cdots : \boldsymbol{0}) \in \mathbb{R}^{d\times dp}$ denote the matrix with the zero matrix in all blocks except in the k-th block, in which is the identity matrix $\boldsymbol{I}_d$. Finally, by (4.90), it follows that for all $r \in \{0, \ldots, p-1\}$

$$\mathrm{Cov}(\boldsymbol{y}_{k-r}, \boldsymbol{u}_{k-\ell}) = \mathbb{E}[\boldsymbol{y}_{k-r}\boldsymbol{u}_{k-\ell}^\top] = \mathbb{E}[\boldsymbol{e}_r\boldsymbol{Y}_{k-1}\boldsymbol{U}_{k-\ell}^\top\boldsymbol{J}^\top] = \boldsymbol{e}_r\boldsymbol{A}^{\ell-1}\boldsymbol{\Sigma}_U\boldsymbol{J}^\top.$$

Case 2.4. Finally, for the entries in the bottom right part, let $a, b \in \{1+d, \ldots, (p+1)d\}$, then

$$(\boldsymbol{\Gamma}_G(\ell))_{a,b} = \mathrm{Cov}(Y_{k-1}^{(a-d)}, Y_{k-1-\ell}^{(b-d)}) = (\boldsymbol{\Gamma}_Y(\ell))_{a-d,b-d},$$

where both $a-d, b-d \in \{1, \ldots, p+1\}$. $\square$

Appendix

4.8 Appendix A: Proofs of auxiliary lemmas

This section is a collection of proofs for lemmas that are mentioned as small extensions of already existing results and/or some technical result, that therefore do not contain new ideas.

Proof of Lemma 4.3. From [39, Lem. 6.2.1] the following holds for some $\alpha \in \mathbb{R}$. If $p = q$,

$$\mathbb{E}\left[\left(\alpha - \frac{1}{p}\langle DF, DG\rangle_{\mathfrak{H}}\right)^2\right] \leqslant (\alpha - \mathbb{E}[FG])^2 + p^2\sum_{r=1}^{p-1}(r-1)!^2\binom{p-1}{r-1}^4(2p-2r)!\|f \otimes_r g\|^2_{\mathfrak{H}^{\otimes(2p-2r)}}.$$

On the other hand if $p \neq q$,

$$\mathbb{E}\left[\left(\alpha - \frac{1}{p}\langle DF, DG\rangle_{\mathfrak{H}}\right)^2\right] \leqslant \alpha^2 + q^2 \sum_{r=1}^{p\wedge q}(r-1)!^2 \binom{p-1}{r-1}^2 \binom{q-1}{r-1}^2 \times (p+q-2r)! \|f \otimes_r g\|^2_{\mathfrak{H}^{\otimes(p+q-2r)}}.$$

To obtain this version of the statement, one stops at [39, Eq.(6.2.3)] and applies that for $r \leqslant p$, $r \leqslant q$, and $p, q \geqslant 2$ it holds that

$$\|f\widetilde{\otimes}_r g\|^2_{\mathfrak{H}^{\otimes p+q-2r}} \leqslant \|f \otimes_r g\|^2_{\mathfrak{H}^{\otimes p+q-2r}}.$$

Now let $\alpha = \mathbb{E}[FG]$ in the above. For $p \neq q$ it holds that $\alpha = \mathbb{E}[FG] = 0$, as multiple integrals of different order are orthogonal [39, Eq.(2.7.4)]. Then we can write the both cases into one, with the help of $b(p,q) := p \wedge q - \mathbb{1}_{\{p=q\}}$, as follows

$$\mathbb{E}\left[(\mathbb{E}[FG] - \frac{1}{p}\langle DF, DG\rangle_{\mathfrak{H}})^2\right] \leqslant q^2 \sum_{r=1}^{b(p,q)} (r-1)!^2 \binom{p-1}{r-1}^2 \binom{q-1}{r-1}^2 (p+q-2r)! \|f \otimes_r g\|^2_{\mathfrak{H}^{\otimes(p+q-2r)}}. \tag{4.91}$$

Next, observe that

$$q^2(r-1)!^2 \binom{p-1}{r-1}^2 \binom{q-1}{r-1}^2 (p+q-2r)! = \frac{qp!q!}{p} \binom{p-1}{r-1}\binom{q-1}{r-1}\binom{p+q-2r}{p-r}.$$

and that for some positive a_i, b_i, c_i, d_i, where $i \in \{1, \dots n\}$ it holds that

$$\sum_{i=1}^n a_i b_i c_i d_i \leqslant \sum_{i=1}^n a_i \sum_{i=1}^n b_i \sum_{i=1}^n c_i \sum_{i=1}^n d_i.$$

This follows simply by distribution property of sum and multiplication and the fact that a_i, b_i, c_i are positive. Further, we have the following well-know identities for binomial coefficients

$$\sum_{r=1}^{p} \binom{p}{r} = 2^p, \quad \binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

and hence as binomial coefficients are positive it also holds that

$$\binom{n-1}{r-1} \leqslant \binom{n}{r}.$$

Now combining these four arguments in (4.91), one gets the wanted bound as follows

$$\begin{aligned}
&\mathbb{E}\left[\left(\mathbb{E}[FG]\frac{1}{p}\langle DF, DG\rangle_{\mathfrak{H}}\right)^2\right] \\
&\leqslant \frac{qp!q!}{p} \sum_{r=1}^{p}\binom{p}{r} \sum_{r=1}^{q}\binom{q}{r} \sum_{r=1}^{b(p,q)}\binom{p+q-2r}{p-r} \sum_{r=1}^{b(p,q)} \|f \otimes_r g\|^2_{\mathfrak{H}^{\otimes(p+q-2r)}} \\
&\leqslant \frac{qp!q!}{p} 2^p 2^q 2^{p+q} \sum_{r=1}^{b(p,q)} \|f \otimes_r g\|^2_{\mathfrak{H}^{\otimes(p+q-2r)}} = \frac{qp!q!}{p} 2^{2p+2q} \sum_{r=1}^{b(p,q)} \|f \otimes_r g\|^2_{\mathfrak{H}^{\otimes(p+q-2r)}}.
\end{aligned}$$

It was used that $p \wedge q$ is less or equal to p and q, and again, by adding more coefficients, as they are positive, we get something bigger. Substitution $t = p - r$ was used together with the above binomial-tricks to obtain

$$\sum_{r=1}^{b(p,q)} \binom{p+q-2r}{p-r} \leqslant \sum_{r=1}^{b(p,q)} \frac{(p+q-2r)!}{(p-r)!(q-r)!} \leqslant \sum_{r=1}^{b(p,q)} \frac{(p+q-2)^{p-r}}{(p-r)!} \leqslant \sum_{t=0}^{p+q} \frac{(p+q-2)^t}{t!} \leqslant 2^{p+q},$$

for $p, q \geqslant 1$. □

Proof of Lemma 4.4. First, assume that $q \geqslant 2$. Then the claim follows from the [32, Prop. A.2] for $p = 1$. Hence it remains to show the case when $q = 1$. To that end define $g(x) := \log^{1/2}(1+x)$, $x \geqslant 0$, and note that it is concave for all $x \geqslant 0$. Then we can write, as $e^{|X|} - 1 \geqslant 0$

$$\begin{aligned} \mathbb{E}[|X|^{1/2}] = \mathbb{E}[\log^{1/2}(e^{|X|} - 1 + 1)] = \mathbb{E}[g(e^{|X|} - 1)] &\leqslant g(\mathbb{E}[e^{|X|} - 1] + 0) \\ &= \log^{1/2}(\mathbb{E}[e^{|X|}]) \end{aligned} \tag{4.92}$$

where we used Jensen's inequality for partially concave functions with $c = 0$ (see [10, Lem 14.6]). Now we can use the same logic from the proof of [32, Prop. A.2]. Recall that Y_k being a sub-q-th chaos variable relative to scale M_k it means that

$$\mathbb{E}\left[\exp\left(\left(\frac{|Y_k|}{M_k}\right)^{2/q}\right)\right] \leqslant 2.$$

Set $|X| = \max_{1 \leqslant k \leqslant m}(|Y_k|/M_k)^{2/q}$. Recall $q = 1$. Then it follows from (4.92) that

$$\mathbb{E}\left[\left(\max_{1\leqslant k\leqslant m}\left(\frac{|Y_k|}{M_k}\right)^{2/q}\right)^{1/2}\right] \leqslant \log^{1/2}\left(\mathbb{E}\left[\exp\left(\max_{1\leqslant k\leqslant m}\left(\frac{|Y_k|}{M_k}\right)^{2/q}\right)\right]\right) \leqslant \log^{1/2}(2m),$$

where we used that

$$\begin{aligned} \mathbb{E}\left[\exp\left(\max_{1\leqslant k\leqslant m}\left(\frac{|Y_k|}{M_k}\right)^{2/q}\right)\right] = \mathbb{E}\left[\max_{1\leqslant k\leqslant m}\exp\left(\left(\frac{|Y_k|}{M_k}\right)^{2/q}\right)\right] &\leqslant \sum_{k=1}^{m}\mathbb{E}\left[\exp\left(\left(\frac{|Y_k|}{M_k}\right)^{2/q}\right)\right] \\ &\leqslant 2m. \end{aligned}$$

To put it all together, write

$$\begin{aligned} \mathbb{E}\left[\max_{1\leqslant k\leqslant m}|Y_k|\right] &= \mathbb{E}\left[\left(\max_{1\leqslant k\leqslant m}|Y_k|^2\right)^{1/2}\right]\frac{\max_{1\leqslant k\leqslant m}M_k}{\left(\max_{1\leqslant k\leqslant m}M_k^2\right)^{1/2}} \\ &= \mathbb{E}\left[\left(\frac{\max_{1\leqslant k\leqslant m}|Y_k|^2}{\max_{1\leqslant k\leqslant m}M_k^2}\right)^{1/2}\right]\max_{1\leqslant k\leqslant m}M_k \\ &\leqslant \mathbb{E}\left[\left(\max_{1\leqslant k\leqslant m}\left(\frac{|Y_k|}{M_k}\right)^2\right)^{1/2}\right]\max_{1\leqslant k\leqslant m}M_k \leqslant \log^{1/2}(2m)\max_{1\leqslant k\leqslant m}M_k \\ &\leqslant \log^{1/2}(2m+1)\max_{1\leqslant k\leqslant m}M_k, \end{aligned}$$

and note that $1 = e^{q/2-1}$ for $q = 1$. Hence, we are done. □

Proof of Lemma 4.5. Using the fact that ℓ^p norms for sequences are decreasing, i.e. that $\|x\|_b \leqslant \|x\|_a$ for $a \leqslant 1 \leqslant b$, where $\|x\|_p = (\sum_{k\geqslant 0}|x_k|^p)^{1/p}$ for $a = 1$ and $b = 2\varpi$, one gets

$$\sum_{q\geqslant 0}\frac{u(d)^q}{(q!)^\varpi} = \left(\sum_{q\geqslant 0}\sqrt{\frac{u(d)^{\frac{q}{\varpi}}}{q!}}^{2\varpi}\right)^{\frac{2\varpi}{2\varpi}} \leqslant \left(\sum_{q\geqslant 0}\sqrt{\frac{u(d)^{\frac{q}{\varpi}}}{q!}}\right)^{2\varpi} \leqslant 2^{2\varpi}\left(\frac{\pi e^{\frac{1}{3}}}{2}\right)^{\frac{\varpi}{2}} e^{\varpi e^{1/(2e)}u(d)^{1/\varpi}},$$

where in the last equation, we used two times Stirling's inequality (4.18) as follows. Define $\widetilde{u}(d) := u(d)^{1/\varpi}$. Then using (4.38) with $\alpha = 1/4$, to get that $q^{1/4} \leqslant (e^{1/(2e)})^{q/2}$, it follows

$$\begin{aligned}\sum_{q\geqslant 0}\frac{\widetilde{u}(d)^{q/2}}{\sqrt{q!}} &\leqslant \sum_{q\geqslant 0}\frac{(\widetilde{u}(d)/2)^{q/2}}{(2\pi q)^{1/4}(q/(2e))^{q/2}e^{1/12}}\frac{e^{1/12}(2\pi q)^{1/4}}{(2\pi q)^{1/4}} \\ &\leqslant \left(\frac{\pi e^{1/3}}{2}\right)^{1/4}\sum_{q\geqslant 2}\frac{q^{1/4}(\widetilde{u}(d)/2)^{q/2}}{(2\pi(q/2))^{1/4}(q/(2e))^{q/2}e^{1/12}} \\ &\leqslant \left(\frac{\pi e^{1/3}}{2}\right)^{1/4}\sum_{q\geqslant 0}\frac{(\widetilde{u}(d)/2e^{1/(2e)})^{q/2}}{(q/2))!} \\ &\leqslant \left(\frac{\pi e^{1/3}}{2}\right)^{1/4} e^{e^{1/(2e)}\widetilde{u}(d)/2}\left(\operatorname{erf}\left(\frac{\widetilde{u}(d)}{2}e^{1/(2e)}\right)+1\right) \\ &\leqslant 2\left(\frac{\pi e^{1/3}}{2}\right)^{1/4}\exp\left(\frac{\widetilde{u}(d)}{2}e^{1/(2e)}\right),\end{aligned} \tag{4.93}$$

where in the display above, we used the identity $\sum_{q\geqslant 0} x^{q/2}/(q/2)! = e^x(\operatorname{erf}(\sqrt{x})+1) \leqslant 2e^x$, where $\operatorname{erf}(x) := 2\pi^{-1/2}\int_0^x e^{-t^2}\mathrm{d}t \leqslant 2$, a special case of the Mittag-Leffler function. □

Proof of Lemma 4.7. Indeed to see this, first note that the amount of elements in I is the same as in $\{1, \ldots, d^2p\}$, i.e. $|I| = |\{1, \ldots, d^2p\}| = d^2p$, hence it is enough to show that λ is injection (or surjection). Assume that

$$(r-1)d^2 + (j-1)d + i = (r'-1)d^2 + (j'-1)d + i', \tag{4.94}$$

where $i, i', j, j' \in \{1, \ldots, d\}$, $r, r' \in \{1, \ldots, p\}$, and reduce both sides modulo d. As $(r-1)d^2$, $(r'-1)d^2$, $(j-1)d$ and $(j'-1)d$ are divisible by d, we get $i \equiv i' (\mod d)$. However, since $i, i' \in \{1, \ldots, d\}$, this implies that $i = i'$. Hence, subtracting i, dividing by d, (4.94) becomes

$$(r-1)d + j = (r'-1)d + j'.$$

Again, reduce this equality modulo d, then as $j, j' \in \{1, \ldots, d\}$, it yields that $j = j'$. Substituting back, we get $r = r'$. Hence the two triplets are equal, proving injectivity. By finite cardinality, λ is therefore a bijection. □

References

[1] M. A. Arcones. Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.*, 22(4):2242–2274, 1994. ISSN 0091-1798,2168-894X. URL http://links.jstor.org/sici?sici=0091-1798(199410)22:4<2242:LTFNFO>2.0.CO;2-L&origin=MSN.

[2] S. Bai and M. S. Taqqu. Multivariate limit theorems in the context of long-range dependence. *J. Time Series Anal.*, 34(6):717–743, 2013. ISSN 0143-9782,1467-9892. doi: 10.1111/jtsa.12046. URL https://doi.org/10.1111/jtsa.12046.

[3] A. Basse-O'Connor and D. Kramer-Bang. Quantative bounds for high-dimensional non-linear functionals of gaussian processes, 2025. URL https://arxiv.org/abs/2502.17718.

[4] S. Basu and G. Michailidis. Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.*, 43(4):1535–1567, 2015. ISSN 0090-5364. doi: 10.1214/15-AOS1315. URL https://doi.org/10.1214/15-AOS1315.

[5] V. Bentkus. On the dependence of the Berry-Esseen bound on dimension. *J. Statist. Plann. Inference*, 113(2):385–402, 2003. ISSN 0378-3758. doi: 10.1016/S0378-3758(02)00094-0. URL https://doi.org/10.1016/S0378-3758(02)00094-0.

[6] V. Bentkus. A Lyapunov type bound in $\mathbf{R}^d$. *Teor. Veroyatn. Primen.*, 49(2):400–410, 2004. ISSN 0040-361X. doi: 10.1137/S0040585X97981123. URL https://doi.org/10.1137/S0040585X97981123.

[7] A. C. Berry. The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.*, 49:122–136, 1941. ISSN 0002-9947. doi: 10.2307/1990053. URL https://doi.org/10.2307/1990053.

[8] A. Bordino, S. Favaro, and S. Fortini. Non-asymptotic approximations of Gaussian neural networks via second-order Poincaré inequalities. In *Symposium on Advances in Approximate Bayesian Inference*, volume 253 of *Proc. Mach. Learn. Res. (PMLR)*, page 34. Proceedings of Machine Learning Research PMLR, [place of publication not identified], 2024.

[9] P. Breuer and P. Major. Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.*, 13(3):425–441, 1983. ISSN 0047-259X. doi: 10.1016/0047-259X(83)90019-2. URL https://doi.org/10.1016/0047-259X(83)90019-2.

[10] P. Bühlmann and S. Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

[11] S. Campese, I. Nourdin, and D. Nualart. Continuous Breuer-Major theorem: tightness and nonstationarity. *Ann. Probab.*, 48(1):147–177, 2020. ISSN 0091-1798. doi: 10.1214/19-AOP1357. URL https://doi.org/10.1214/19-AOP1357.

[12] D. Chambers and E. Slud. Central limit theorems for nonlinear functionals of stationary Gaussian processes. *Probab. Theory Related Fields*, 80(3):323–346, 1989. ISSN 0178-8051. doi: 10.1007/BF01794427. URL https://doi.org/10.1007/BF01794427.

[13] J. Chang, X. Chen, and M. Wu. Central limit theorems for high dimensional dependent data. *Bernoulli*, 30(1):712–742, 2024. ISSN 1350-7265. doi: 10.3150/23-bej1614. URL https://doi.org/10.3150/23-bej1614.

[14] W. Chang, K. Takatsu, K. Urban, and A. K. Kuchibhotla. The berry-esseen bound for high-dimensional self-normalized sums, 2025. URL https://arxiv.org/abs/2501.08979.

[15] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao. *Normal Approximation by Stein's Method*. Springer, 2011. ISBN 978-0-387-85725-8.

[16] V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximation of suprema of empirical processes. *Ann. Statist.*, 42(4):1564–1597, 2014. ISSN 0090-5364. doi: 10.1214/14-AOS1230. URL https://doi.org/10.1214/14-AOS1230.

[17] V. Chernozhukov, D. Chetverikov, and K. Kato. Central limit theorems and bootstrap in high dimensions. *Ann. Probab.*, 45(4):2309–2352, 2017. ISSN 0091-1798. doi: 10.1214/16-AOP1113. URL https://doi.org/10.1214/16-AOP1113.

[18] V. Chernozhukov, D. Chetverikov, K. Kato, and Y. Koike. High-dimensional data bootstrap. *Annu. Rev. Stat. Appl.*, 10:427–449, 2023. ISSN 2326-8298. doi: 10.1146/annurev-statistics-040120-022239. URL https://doi.org/10.1146/annurev-statistics-040120-022239.

[19] V. Chernozhukov, D. Chetverikov, and Y. Koike. Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *Ann. Appl. Probab.*, 33(3):2374–2425, 2023. ISSN 1050-5164. doi: 10.1214/22-aap1870. URL https://doi.org/10.1214/22-aap1870.

[20] G. Ciołek, D. Marushkevych, and M. Podolskij. On Dantzig and Lasso estimators of the drift in a high dimensional Ornstein-Uhlenbeck model. *Electron. J. Stat.*, 14(2):4395–4420, 2020. ISSN 1935-7524. doi: 10.1214/20-EJS1775. URL https://doi.org/10.1214/20-EJS1775.

[21] H. Dehling, A. Rooch, and M. S. Taqqu. Non-parametric change-point tests for long-range dependent data. *Scand. J. Stat.*, 40(1):153–173, 2013. ISSN 0303-6898,1467-9469. doi: 10.1111/j.1467-9469.2012.00799.x. URL https://doi.org/10.1111/j.1467-9469.2012.00799.x.

[22] H. Dehling, A. Rooch, and M. Wendler. Two-sample U-statistic processes for long-range dependent data. *Statistics*, 51(1):84–104, 2017. ISSN 0233-1888,1029-4910. doi: 10.1080/02331888.2016.1270542. URL https://doi.org/10.1080/02331888.2016.1270542.

[23] S. Douissi, K. Es-Sebaiy, F. Alshahrani, and F. G. Viens. AR(1) processes driven by second-chaos white noise: Berry-Esséen bounds for quadratic variation and parameter estimation. *Stochastic Process. Appl.*, 150:886–918, 2022. ISSN 0304-4149. doi: 10.1016/j.spa.2020.02.007. URL https://doi.org/10.1016/j.spa.2020.02.007.

[24] C.-G. Esseen. On the Liapounoff limit of error in the theory of probability. *Ark. Mat. Astr. Fys.*, 28A(9):19, 1942.

[25] X. Fang and Y. Koike. High-dimensional central limit theorems by Stein's method. *Ann. Appl. Probab.*, 31(4):1660–1686, 2021. ISSN 1050-5164. doi: 10.1214/20-aap1629. URL https://doi.org/10.1214/20-aap1629.

[26] S. Favaro, B. Hanin, D. Marinucci, I. Nourdin, and G. Peccati. Quantitative CLTs in deep neural networks. *Probab. Theory Related Fields*, 191(3-4):933–977, 2025. ISSN 0178-8051,1432-2064. doi: 10.1007/s00440-025-01360-1. URL https://doi.org/10.1007/s00440-025-01360-1.

[27] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, 1994. ISBN 9780691218632. doi: doi:10.1515/9780691218632. URL https://doi.org/10.1515/9780691218632.

[28] M. Jirak. Weak dependence and optimal quantitative self-normalized central limit theorems, 2025. URL https://arxiv.org/abs/2504.14403.

[29] Y. T. Kim and H. S. Park. Kolmogorov distance for the central limit theorems of the Wiener chaos expansion and applications. *J. Korean Statist. Soc.*, 44(4): 565–576, 2015. ISSN 1226-3192,2005-2863. doi: 10.1016/j.jkss.2015.03.003. URL https://doi.org/10.1016/j.jkss.2015.03.003.

[30] Y. T. Kim and H. S. Park. Optimal Berry-Esseen bound for an estimator of parameter in the Ornstein-Uhlenbeck process. *J. Korean Statist. Soc.*, 46(3):413–425, 2017. ISSN 1226-3192. doi: 10.1016/j.jkss.2017.01.002. URL https://doi.org/10.1016/j.jkss.2017.01.002.

[31] Y. Koike. Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data. *Ann. Statist.*, 47(3):1663–1687, 2019. ISSN 0090-5364. doi: 10.1214/18-AOS1731. URL https://doi.org/10.1214/18-AOS1731.

[32] Y. Koike. Supplement to "gaussian approximation of maxima of wiener functionals and its application to high-frequency data.", 2019.

[33] C. Lévy-Leduc, H. Boistard, E. Moulines, M. S. Taqqu, and V. A. Reisen. Asymptotic properties of U-processes under long-range dependence. *Ann. Statist.*, 39(3):1399–1426, 2011. ISSN 0090-5364,2168-8966. doi: 10.1214/10-AOS867. URL https://doi.org/10.1214/10-AOS867.

[34] H. Lütkepohl. *New introduction to multiple time series analysis*. Springer-Verlag, Berlin, 2005. ISBN 3-540-40172-5. doi: 10.1007/978-3-540-27752-1. URL https://doi.org/10.1007/978-3-540-27752-1.

[35] L. Maini and I. Nourdin. Spectral central limit theorem for additive functionals of isotropic and stationary Gaussian fields. *Ann. Probab.*, 52(2):737–763, 2024. ISSN 0091-1798. doi: 10.1214/23-aop1669. URL https://doi.org/10.1214/23-aop1669.

[36] D. Marinucci. The empirical process for bivariate sequences with long memory. *Stat. Inference Stoch. Process.*, 8(2):205–223, 2005. ISSN 1387-0874,1572-9311. doi: 10.1007/s11203-004-2790-9. URL https://doi.org/10.1007/s11203-004-2790-9.

[37] I. Nourdin and D. Nualart. The functional Breuer-Major theorem. *Probab. Theory Related Fields*, 176(1-2):203–218, 2020. ISSN 0178-8051. doi: 10.1007/s00440-019-00917-1. URL https://doi.org/10.1007/s00440-019-00917-1.

[38] I. Nourdin and G. Peccati. Stein's method on Wiener chaos. *Probab. Theory Related Fields*, 145(1-2):75–118, 2009. ISSN 0178-8051. doi: 10.1007/s00440-008-0162-x. URL https://doi.org/10.1007/s00440-008-0162-x.

[39] I. Nourdin and G. Peccati. *Normal approximations with Malliavin calculus*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2012. ISBN 978-1-107-01777-1. doi: 10.1017/CBO9781139084659. URL https://doi.org/10.1017/CBO9781139084659. From Stein's method to universality.

[40] I. Nourdin, G. Peccati, and A. Réveillac. Multivariate normal approximation using Stein's method and Malliavin calculus. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46 (1):45–58, 2010. ISSN 0246-0203. doi: 10.1214/08-AIHP308. URL https://doi.org/10.1214/08-AIHP308.

[41] I. Nourdin, G. Peccati, and M. Podolskij. Quantitative Breuer-Major theorems. *Stochastic Process. Appl.*, 121(4):793–812, 2011. ISSN 0304-4149. doi: 10.1016/j.spa.2010.12.006. URL https://doi.org/10.1016/j.spa.2010.12.006.

[42] I. Nourdin, G. Peccati, and Y. Swan. Entropy and the fourth moment phenomenon. *J. Funct. Anal.*, 266(5):3170–3207, 2014. ISSN 0022-1236. doi: 10.1016/j.jfa.2013.09.017. URL https://doi.org/10.1016/j.jfa.2013.09.017.

[43] I. Nourdin, G. Peccati, and X. Yang. Berry-Esseen bounds in the Breuer-Major CLT and Gebelein's inequality. *Electron. Commun. Probab.*, 24:Paper No. 34, 12, 2019. doi: 10.1214/19-ECP241. URL https://doi.org/10.1214/19-ECP241.

[44] I. Nourdin, G. Peccati, and X. Yang. Multivariate normal approximation on the Wiener space: new bounds in the convex distance. *J. Theoret. Probab.*, 35(3):2020–2037, 2022. ISSN 0894-9840. doi: 10.1007/s10959-021-01112-6. URL https://doi.org/10.1007/s10959-021-01112-6.

[45] D. Nualart. *Malliavin calculus and its applications*, volume 110 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of

the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2009. ISBN 978-0-8218-4779-4. doi: 10.1090/cbms/110. URL https://doi.org/10.1090/cbms/110.

[46] D. Nualart and E. Nualart. *Introduction to Malliavin calculus*. Cambridge University Press, 2018.

[47] D. Nualart and G. Peccati. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.*, 33(1):177–193, 2005. ISSN 0091-1798. doi: 10.1214/009117904000000621. URL https://doi.org/10.1214/009117904000000621.

[48] D. Nualart and A. Tilva. Continuous Breuer-Major theorem for vector valued fields. *Stoch. Anal. Appl.*, 38(4):668–685, 2020. ISSN 0736-2994. doi: 10.1080/07362994.2019.1711118. URL https://doi.org/10.1080/07362994.2019.1711118.

[49] D. Nualart and H. Zhou. Total variation estimates in the Breuer-Major theorem. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(2):740–777, 2021. ISSN 0246-0203. doi: 10.1214/20-aihp1094. URL https://doi.org/10.1214/20-aihp1094.

[50] H. Robbins. A remark on Stirling's formula. *Amer. Math. Monthly*, 62:26–29, 1955. ISSN 0002-9890. doi: 10.2307/2308012. URL https://doi.org/10.2307/2308012.

[51] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602. Univ. California Press, Berkeley, CA, 1972.

[52] W. B. Wu. Empirical processes of long-memory sequences. *Bernoulli*, 9(5):809–831, 2003. ISSN 1350-7265,1573-9759. doi: 10.3150/bj/1066418879. URL https://doi.org/10.3150/bj/1066418879.

Paper B

Fitting Structural Models to Electricity Markets

Andreas Basse-O'Connor, Tor Bonde, Lota Copić and Jan Pedersen

Abstract

We develop a structural model for electricity spot prices that incorporates renewable generation into the multi-fuel bid stack framework. We show that renewables can be modeled by adjusting the demand curve to account for renewable generation - effectively reducing the market to one with only conventional fuels. Under standard assumptions of inelastic demand and market equilibrium, we derive a general pricing model highlighting two special cases: a single conventional fuel and a proportional multi-fuel setting. Using German market data, we evaluate several structural models within this framework and benchmark their performance against machine-learning models and a simple reduced-form model. Our findings show that a theory-driven one-fuel structural model, using gas prices, EUA prices, and residual demand as driving variables delivers the most robust out-of-sample performance. These results demonstrate the practical value of theoretical structural models, highlighting their simplicity and interpretability.

1 Introduction

1.1 Background

The dependency between electricity spot prices and their driving factors is crucial for decision-makers in the energy market, as it improves their ability to navigate uncertainties and manage risks. A key approach and an essential first step in understanding these uncertainties is to model the prices. This brings us to the rich family of electricity spot price models. They can be broadly categorized into three types: fundamental models, reduced-form models, and structural models. See Deschatre et al. [8] and Weron [22] for surveys of these models.

Fundamental models incorporate detailed knowledge of all generation units, production costs, and market constraints. These models typically yield electricity prices by solving a complex optimization problem, which can lead to significant challenges in computational complexity and difficulty adapting to changing market conditions. In addition, they are limited in their ability to model uncertainty.

In contrast, reduced form models, often presented as stochastic differential equations, are mathematically elegant but lack explicit causal relationships between prices and driving factors. For example, one of the key factors are the fuel prices. Reduced-form models either ignore fuel prices or include them as correlated processes, failing to capture the dependence structure between fuels and electricity prices, and it is exactly this dependence that we are interested into. Furthermore, reduced form models often represent price spikes, a feature inherent to energy prices, through jump processes or regime switches, providing limited insight into the underlying causes of sudden price swings (as argued in Section 1 in Carmona et al. [4]).

In between fundamental and reduced-form models, one finds structural models, which blend elements of both approaches. In structural models, the electricity spot price S_t at time t is given by a function of a vector of underlying state variables X_t at the same time, such as fuel costs, demand, and the capacity of power plants. These variables are treated as exogenous in the spot price equation: the model specifies how X_t affects S_t, but not any feedback from prices to X_t.

There are two aspects to these models: one is obtaining the desired functional form, typically using equilibrium analysis, and the other is modeling the dynamics of the state variables X_t. Combining these two aspects yields a dynamic model for the electricity spot price. This approach is convenient for understanding the relationship between spot prices and their fundamental driving variables X_t, such as identifying correlations, and, eventually, for pricing. For an extensive survey of structural models, see Carmona and Coulon [3] and Deschatre et al. [8].

Numerous studies have advanced the development of structural models for electricity pricing, with a central theme of capturing the supply-demand relationship. The seminal work of Barlow [1] introduced a model in which electricity spot prices are derived using a vertical demand curve and a supply curve represented as a power function of a simple diffusion process. This approach effectively captures stylized features, such as price spikes, without relying on jump processes. The vertical demand curve reflects the widely accepted assumption of inelastic demand with respect to electricity prices, as discussed in Burger et al. [2], Carmona et al. [4], Cartea and Villaplana [5], Howison and Coulon [12].

Building on Barlow's foundation, Kanamura and Ōhashi [14] incorporated explicit seasonality into the demand function, while Cartea and Villaplana [5] introduced the capacity of the power plant as an additional driving variable. Burger et al. [2] addressed short-term and long-term volatility by including components such as an adjusted load function for prices and a residual process, which they described as the "psychology of the market". Their use of cubic splines to fit the supply curve closely aligns with the approach used in this paper.

These earlier versions of structural models often consider demand and capacity as short-term drivers. Extending this, Howison and Coulon [12], Carmona et al. [4], and Füss et al. [10] introduced multi-fuel models, recognizing fuel prices as key drivers of medium- to long-term dynamics. Additionally, Pirrong and Jermakyan [17] employed a simple two-factor model using natural gas prices and demand, which de Maere d'Aertrycke and Smeers [7] later expanded to include multiple fuel types. The fuel-inclusive models revolve around modeling bid stack functions to capture the supply-demand relationship.

Our model builds on this framework, expanding it to incorporate renewable energy through the concept of "residual demand". Related to this, Wagner [21] extended Barlow's seminal work by modeling the dynamics of residual demand to account for renewable energy sources. Similarly, Coskun and Korn [6] addressed increasing supply variability due to renewables by using a Jacobi process to model demand in the German electricity market.

1.2 Main Contributions

In this paper, we focus on the first key aspect of the structural models mentioned in Subsection 1.1, namely the functional relationship between the spot price S_t and the fundamental market factors: electricity demand D_t, fuel prices F_t used in conventional generation, and renewable infeed W_t. These fundamental components $(D_t,\ F_t,\ W_t)$ capture the essential economic forces in price formation: demand reflects the load to be served, fuel prices determine the marginal cost of conventional generation, and renewable infeed–assumed to have zero marginal cost–displaces costlier supply from the merit order (see Carmona et al. [4], Kirschen and Strbac [15], Stoft [20]). Under the standard assumptions of price-inelastic demand and market equilibrium we derive a structural model in which prices are expressed as $S_t = b(R_t, F_t)$, where $R_t = D_t - W_t$ denotes the *residual demand* and the function b is the *market bid stack function without renewables*, which is described in full detail in Subsection 2.1. The resulting structural model, which is presented in Theorem 2.4, extends the framework of Carmona et al. [4, Sec. 3] by incorporating both renewable and conventional fuel sources.

Our primary objective is to evaluate the performance of various choices for b and the driving variables (D_t, F_t, W_t) (or equivalently (R_t, F_t)). By testing models presented in Theorem 2.4 on historical German data (for data details see Subsection 3.3), we conduct in-sample fitting and assess their out-of-sample performance. Notably, our findings indicate that models solely driven by gas price adjusted by the EUA price, and residual demand exhibit a robust and strong out-of-sample performance. This underscores the significance of considering specific factors in electricity prices effectively. More precisely, our results reveal the following four points. Let G_t and C_t represent the prices of gas and coal at time t, R_t represents the residual demand at time t, and $\overline{C_t}$, $\overline{G_t}$ denote prices of coal and gas respectively, adjusted for the price of EUAs. The functions u, v, and q are flexible and general in nature.

- Models of the form $S_t = \overline{G}_t q(R_t)$ and $S_t = \overline{C}_t q(R_t)$ consistently rank among the top performers, with $S_t = \overline{G}_t q(R_t)$ showing the highest robustness. By Theorem 2.4, these models reflect exactly electricity pricing in one-fuel efficient markets, or in multi-fuel markets with proportional fuel prices. In other words, theory-driven models are top performers.
- In recent years, gas-driven models exhibit the best out-of-sample performance.
- More general models, such as $S_t = v(C_t)u(G_t)q(R_t)$ or $S_t = u(G_t)q(R_t)$, are prone to overfitting.
- Structural models perform comparably to, and sometimes better than, the machine-learning benchmarks considered in this study, namely `AdaBoostRegressor` and.

While our focus is on modeling the structural relationship between electricity spot prices and driving variables, the forecasting of electricity spot prices has been a subject of extensive research. As summarized in reviews by Weron [22] and Petropoulos et al. [16], drivers, such as (residual) demand, EUA, and fuel prices, have been identified as key drivers of electricity spot prices. Our study is thus well aligned with these findings.

The paper is structured as follows. Section 2 introduces the fundamental mechanism by which electricity spot prices are determined, with a focus on the merit order principle–a cornerstone of structural models. Building on this foundation, we propose a theoretical framework that extends Carmona et al. [4] to include both renewable and conventional generation. A number of prominent models in the literature can be obtained as special cases. Section 3 presents our data analysis, including the methodology, data description, and results. A summary of our main findings is provided in Section 4. Additional details and a historical overview of price developments are included in Appendix.

2 Pricing of Electricity

This section lays the groundwork for a structural model of electricity prices. Our aim is to formalize how key market fundamentals–demand, fuel costs, and renewable infeed–jointly determine the day-ahead electricity price. We begin by describing the operational logic of price formation in wholesale electricity markets, and then derive an electricity price model based on the market bid stack function and the concept of residual demand. Specifically, the market bid stack function is described in detail in Subsection 2.1, and the pricing model is presented in Subsection 2.2.

In liberalized electricity markets, producers submit supply offers and consumers submit demand bids for each hour of the following day. These bids are collected in a day-ahead auction administered by a market operator. Supply bids are ordered from lowest to highest to construct the aggregate supply curve (or *bid curve*), while demand bids are ordered from highest to lowest to form the demand curve (or *ask curve*). The intersection of these curves determines the market-clearing price and quantity for each hour, following the merit order principle.

In efficient markets, producers' bids are assumed to reflect their marginal costs. These costs depend primarily on the fuel price, the power plant's efficiency (measured by its heat rate), and the cost of emissions. The heat rate indicates how much fuel is needed to produce one unit of electricity and varies across plants, even within the same fuel type. The fuel associated with the marginal unit–the last one needed to meet demand–is referred to as the *price-setting* fuel.

By ordering marginal costs within a fuel type, we obtain a *fuel bid curve*. If all plants using a given fuel had identical efficiencies, the curve would be flat. In practice, variation in plant characteristics leads to an upward-sloping bid curve. Figure 2.1 illustrates stylized bid curves for gas and coal in the German market.

2.1 The market bid stack function

We now formalize the preceding description into a mathematical model. Consider a market in which companies generate electricity using one of n conventional fuels (such as gas or coal) or renewable sources. Throughout this section, we make the following assumptions:

Assumption 2.1.

(1) Companies using conventional fuels submit bids based solely on their capacity and fuel costs. In addition, they always submit positive bids.
(2) There are no production costs associated with electricity generated from renewable sources (e.g., due to negligible fuel costs or priority dispatch). Accordingly, companies

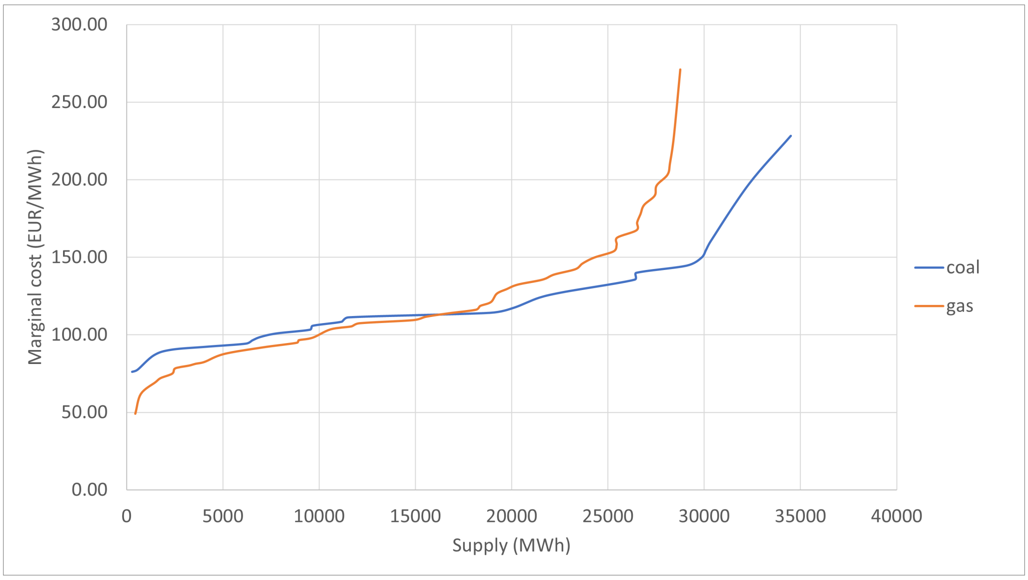


Figure 2.1: Empirical fuel bid curves for gas (orange) and coal (blue). The bid functions are based on Equation (2.3), where the heat function $q_i(\xi)$ (with i representing either gas or coal), corresponds to the inverse efficiency of an electricity-producing company using fuel i at rank ξ. Efficiency estimates were kindly provided by Danske Commodities. In Equation (2.3), we use as f_i the average fuel price per MWh at 12 pm in 2023, combined with the average emission certificate price per MWh of produced electricity.

that generate electricity from renewable sources will submit a bid of zero if renewable sources are available and will submit no bid if they are not.

(3) The merit order principle applies, meaning that bids are arranged from the least expensive to the most expensive.

We can now define *the market bid stack function*:

$$(\xi, f, w) \mapsto \overline{b}(\xi, f, w) \in [0, \infty),$$

where $\overline{b}(\xi, f, w)$ represents the marginal cost at supply level ξ given fuel prices $f = (f_1, \ldots, f_n) \in (0, \infty)^n$ and renewable generation $w \in [0, \infty)$. The supply level ξ is in the interval $(0, \overline{\xi} + w]$, where $\overline{\xi}$ is the maximum capacity of the companies that produce electricity using conventional fuels. By the merit order principle, $\xi \mapsto \overline{b}(\xi, f, w) \in [0, \infty)$ is non-decreasing for all f and w.

Denote by $b(\xi, f)$ the market bid stack function in the case where there is zero infeed from the renewables, that is, $b(\xi, f) := \overline{b}(\xi, f, 0)$. We call b *the market bid stack function without renewables.* By Assumption 2.1(1) it holds that

$$b(\xi, f) > 0 \text{ for all } f = (f_1, \ldots, f_n) \in (0, \infty)^n \text{ and } \xi \in (0, \overline{\xi}].$$

Moreover, by Assumption 2.1(2) and Assumption 2.1(3),

$$\overline{b}(\xi, f, w) = \begin{cases} 0 & \text{if } \xi \leqslant w \\ b(\xi - w, f) & \text{if } \xi > w. \end{cases} \tag{2.1}$$

The latter case in Equation (2.1) reflects the scenario where the total supply ξ is greater than the renewable generation w. In other words, the remaining supply $(\xi - w)$ denotes the supply obtained using conventional fuels.

Although renewable sources, such as wind farms, contribute to certain "plants", they operate differently from conventional fuels and power plants. In our model, renewable generation has zero marginal cost, whereas conventional generation has strictly positive marginal cost. Consequently, due to the merit order principle (Assumption 2.1(3)) renewables are prioritized to meet demand, which is reflected in the first case in Equation (2.1). However, unlike conventional power plants that provide consistent and predictable output, renewable energy production is variable. Conventional plants can operate at full capacity when needed, but renewable plants, such as wind farms, are subject to fluctuations in resource availability (e.g., wind speed).

Under the assumption that renewables have zero marginal cost, the key inputs are renewable generation capacity w and the vector of conventional fuel prices $f = (f_1, \ldots, f_n)$, as illustrated in Equation (2.1).

To further describe the market bid function, we introduce, as in Carmona et al. [4], *fuel bid stack functions* $b_i : [0, \overline{\xi}_i] \times (0, \infty) \to (0, \infty)$ for fuels $i = 1, \ldots, n$, where $\overline{\xi}_i \in (0, \infty)$ is the total capacity of companies using fuel i. We interpret $b_i(\xi, f_i)$ as the marginal cost of producing an additional unit of electricity using fuel i when the fuel price is f_i. The total maximum capacity of conventional fuels is then $\overline{\xi} = \sum_{i=1}^{n} \overline{\xi}_i$. Recall that $b(\xi, f)$ represents the market bid stack in a market without renewables. This function is determined by $b_1, \ldots, b_n$ as described in Proposition 1 in Carmona et al. [4] (see Proposition 1 below). Formally, the bid stack function without renewables $b(\xi, f)$ is given by the supremum of all possible prices $p \in \mathbb{R}$ where the sum of generalized inverses $b_i(\,\cdot\,, f_i)^{-1}$ evaluated at p is less than the supply level ξ, for $(\xi, f) \in [0, \overline{\xi}] \times (0, \infty)^n$.

For convenience, define

$$\underline{b}_i(f_i) := b_i(0, f_i) \quad \text{and} \quad \overline{b}_i(f_i) := b_i(\overline{\xi}_i, f_i),$$

as the minimum and maximum values attained by the fuel bid curve for fuel i, respectively. The following is an adaptation of Carmona et al. [4, Prop.1] to the current context.

Proposition 1. *Carmona et al. [4, Prop. 1] Given the above assumptions, the market bid stack function without renewables is given by*

$$b(\xi, f) = \min_{i=1,\ldots,n} \underline{b}_i(f_i) \vee \sup \left\{ p \in \mathbb{R} : \sum_{i=1}^{n} b_i(\cdot, f_i)^{-1}(p) < \xi \right\}, \quad \text{for } (\xi, f) \in [0, \overline{\xi}] \times (0, \infty)^n,$$

where we use the convention $\sup \varnothing = -\infty$, *and*

$$b_i(\cdot, f_i)^{-1}(p) := \overline{\xi}_i \wedge \inf \left\{ \xi \in (0, \overline{\xi}_i] : b_i(\xi, f_i) > p \right\} \tag{2.2}$$

is the generalized right-continuous inverse of the function $\xi \mapsto b_i(\xi, f_i)$.

Remark 2.2. Assume the market is efficient, in which companies bid their marginal costs. In this case, the fuel bid function $f_i \mapsto b_i(\xi, f_i)$ is linear for each fuel $i = 1, \ldots, n$. Specifically,

$$b_i(\xi, f_i) = f_i q_i(\xi), \tag{2.3}$$

where $q_i(\xi) = b_i(\xi, 1)$ represents the heat rate of the companies utilizing fuel i. A lower value of $q_i(\xi)$ indicates a more efficient company at rank ξ, while a higher value corresponds to less efficient companies. The linearity of b_i was already stressed by Pirrong and Jermakyan [17].

Given the importance of this linearity, we will revisit the argument here. First, consider a specific fuel price, say 1. The fuel bid stack $b_i(\xi, 1)$ is obtained by ranking the bids from cheapest to most expensive. By assumption, $b_i(\xi, 1)$ represents the marginal cost of a company at rank ξ. If the fuel price changes to f_i for some $f_i > 0$, the marginal cost of this company would also change proportionally to $f_i \cdot b_i(\xi, 1)$. Since the ranking of bids remains unchanged, the entire fuel bid stack with fuel price f_i becomes $\xi \mapsto f_i \cdot b_i(\xi, 1)$. Thus, $b_i(\xi, f_i) = f_i \cdot b_i(\xi, 1)$.

We now describe how the market bid stack function b (without renewables) is determined by the fuel heat rates, using Proposition 1 and assuming Equation (2.3). For simplicity, we consider the case $n = 2$, which means that f is two-dimensional, with $f = (f_1, f_2)$. Let $i = 1, 2$ and fix a supply level $\xi \in [0, \overline{\xi}]$. We use the following terminology

$$\text{fuel } i \text{ is } \begin{cases} \text{not used (n.u.)} & \text{if } b(\xi, f) \leqslant \underline{b}_i(f_i) \\ \text{a price setter (p.s.)} & \text{if } b(\xi, f) \in (\underline{b}_i(f_i), \overline{b}_i(f_i)] \\ \text{fully used (f.u.)} & \text{if } b(\xi, f) > \overline{b}_i(f_i). \end{cases}$$

Proposition 2. *Consider an efficient market with two conventional fuels, i.e. $n = 2$. That is, the fuel bid stack functions are given by Equation (2.3). Assume additionally that $q_i : [0, \overline{\xi}_i] \to [0, \infty)$ is strictly increasing and continuous for $i = 1, 2$. Then the following holds:*

(1) The market bid stack function without renewables for $\xi \in (0, \overline{\xi}]$ is as follows:

$b(\xi, f)$	Fuels	When
$f_1 q_1(\xi)$	1 is a p.s. and 2 is n.u.	$b_1(\xi, f_1) \leqslant \underline{b}_2(f_2)$
$f_2 q_2(\xi)$	2 is a p.s. and 1 is n.u.	$b_2(\xi, f_2) \leqslant \underline{b}_1(f_1)$
$f_1 q_1(\xi - \overline{\xi}_2)$	1 is a p.s. and 2 is f.u.	$b_1(\xi - \overline{\xi}_2, f_1) > \overline{b}_2(f_2)$
$f_2 q_2(\xi - \overline{\xi}_1)$	2 is a p.s. and 1 is f.u.	$b_2(\xi - \overline{\xi}_1, f_2) > \overline{b}_1(f_1)$
$[q_1^{-1}(\cdot/f_1) + q_2^{-1}(\cdot/f_2)]^{-1}(\xi)$	1 is a p.s. and 2 is a p.s.	otherwise.

(2) Assume that the two fuel prices are proportional, i.e. $f_1 = \alpha f_2$ for $\alpha > 0$. Then the market bid stack function is given by:

$$b(\xi, f) = f_2 \tilde{q}(\xi),$$

where

$$\tilde{q}(\xi) = \begin{cases} \alpha q_1(\xi) & \textit{if fuel 1 is a p.s. and fuel 2 is n.u.} \\ q_2(\xi) & \textit{if fuel 2 is a p.s. and fuel 1 is n.u.} \\ \alpha q_1(\xi - \overline{\xi}_2) & \textit{if fuel 1 is a p.s. and fuel 2 is f.u.} \\ q_2(\xi - \overline{\xi}_1) & \textit{if fuel 2 is a p.s. and fuel 1 is f.u.} \\ [q_1^{-1}(\cdot/\alpha) + q_2^{-1}(\cdot)]^{-1}(\xi) & \textit{if fuels 1 and 2 are both p.s.} \end{cases} \tag{2.4}$$

The proof of Proposition 2 is in 4.1.

Remark 2.3. Table 2.1 explicitly summarizes the market bid stack function without renewables $b(\xi, f)$, under different conditions.

Condition	Range for ξ	Fuels	$b(\xi, f)$
$\bar{b}_1(f_1) \leqslant \underline{b}_2(f_2)$	$(0, \bar{\xi}_1]$	1 is a p.s., 2 is n.u.	$b_1(\xi, f_1)$
	$(\bar{\xi}_1, \bar{\xi}]$	1 is f.u., 2 is a p.s.	$b_2(\xi - \bar{\xi}_1, f_2)$
$\underline{b}_1(f_1) \leqslant \underline{b}_2(f_2) \leqslant \bar{b}_1(f_1) \leqslant \bar{b}_2(f_2)$	$(0, \xi_{1,1}]$	1 is a p.s., 2 is n.u.	$b_1(\xi, f_1)$
	$(\xi_{1,1}, \bar{\xi}_1 + \xi_{2,2}]$	1 is a p.s., 2 is a p.s.	$[q_1^{-1}(\cdot/f_1) + q_2^{-1}(\cdot/f_2)]^{-1}(\xi)$
	$(\bar{\xi}_1 + \xi_{2,2}, \bar{\xi}]$	1 is f.u., 2 is a p.s.	$b_2(\xi - \bar{\xi}_1, f_2)$
$\underline{b}_1(f_1) \leqslant \underline{b}_2(f_2) < \bar{b}_2(f_2) \leqslant \bar{b}_1(f_1)$	$(0, \xi_{1,1}]$	1 is a p.s., 2 is n.u.	$b_1(\xi, f_1)$
	$(\xi_{1,1}, \bar{\xi}_2 + \xi_{1,2}]$	1 is a p.s., 2 is a p.s.	$[q_1^{-1}(\cdot/f_1) + q_2^{-1}(\cdot/f_2)]^{-1}(\xi)$
	$(\bar{\xi}_2 + \xi_{1,2}, \bar{\xi}]$	1 is a p.s., 2 is f.u.	$b_1(\xi - \bar{\xi}_2, f_1)$

Table 2.1: Assume $\underline{b}_1(f_1) \leqslant \underline{b}_2(f_2)$. The table specifies the conditions under which each of the two fuels acts as a price setter, is fully used, or is not used. Moreover, it specifies the market bid stack function without renewables, $b(\xi, f)$. Here, $\xi_{i,j}$ are defined in (4.2).

Having described the market bid stack function without renewables $\xi \mapsto b(\xi, f)$ in terms of the bid stack functions, Equation (2.1) then allows us to derive the market bid stack function with renewables.

2.2 A mathematical model for pricing of electricity

Let the vector of fuel prices at time t be denoted by $F_t = (F_t^1, \dots, F_t^n)$, and let W_t represent the amount of renewable energy produced at time t. To model electricity prices, we adopt standard assumptions commonly used in the literature (e.g., Carmona et al. [4], Barlow [1]). Specifically, we assume that the market is in equilibrium and that electricity demand is inelastic.

The equilibrium assumption means that, at each time t, electricity supply equals demand D_t. Inelasticity means that the quantity demanded remains constant regardless of price. Under these assumptions, the electricity price is determined by the requirement that supply–being a function of the price–matches the fixed demand at time t. Thus, the electricity price corresponds to the inverse of the supply function evaluated at D_t.

In Subsection 2.1, we introduced the market bid stack function, which precisely describes this inverse supply function at time t. Hence, under the equilibrium and demand inelasticity assumptions, we obtain:

$$S_t = \bar{b}(D_t, F_t, W_t). \tag{2.5}$$

Using Equation (2.1), this relation can be conveniently reformulated in terms of the market bid stack function without renewables, $b(\xi, f)$. Note that $b(\xi, f)$ is only defined for $\xi \geqslant 0$; however, we extend its definition by setting $b(\xi, f) = 0$ for $\xi \leqslant 0$. This extension is natural, as a negative value of ξ corresponds to negative residual demand, indicating that renewable generation exceeds demand. In such cases, the spot price is zero, consistent with Assumption 2.1(2), which states that renewables have zero marginal cost.

Combining Equations (2.1) and (2.5) yields Theorem 2.4(1). Parts (2) and (3) follow directly from Proposition 2.

Theorem 2.4. *Consider a market with n conventional fuels and assume that the demand for electricity is inelastic and that the market is in equilibrium. Then the following results hold:*

(1) The electricity price at time t is given by

$$S_t = b(R_t, F_t), \tag{2.6}$$

where b denotes the market bid stack function without renewables.

(2) Assume additionally that the market is efficient with a single conventional fuel with price $F_t = F_t^1$ at time t, i.e. $n = 1$. Then the electricity price at time t is

$$S_t = F_t q_1(R_t), \tag{2.7}$$

where q_1 denotes the heat rate.

(3) Consider the case with $n = 2$ fuels, such as gas and coal, with fuel prices $F_t = (G_t, C_t)$ at time t, where G_t is the price of gas and C_t is the price of coal. Assume additionally that G_t and C_t are proportional, that is, there is a constant α such that $C_t = \alpha G_t$ for all t. Then the price of electricity is

$$S_t = G_t \tilde{q}(R_t),$$

where $\tilde{q}$ is defined in Equation (2.4).

Remark 2.5. Note that all the mentioned variables, i.e. F_t, W_t, D_t and consequently R_t and S_t may be considered random. For the purpose of this paper, as we focus on the functional form, it is the same whether we consider the variables as deterministic or stochastic.

Example 2.6. Let us now consider some important examples related to the models that we will test on data.

(A) (Exponential fuel bid stack functions). Consider an efficient market with $n = 2$ fuels, such as gas and coal. We denote the fuel prices at time t by $F_t = (G_t, C_t)$, where G_t is the price of gas and C_t is the price of coal. As described in Carmona et al. [4], exponential bid stack functions enable the explicit determination of the electricity price and e.g. allow for the derivation of explicit formulas for the price of derivative assets. Let the fuel bid stack functions be given as

$$b_i(\xi, f_i) = f_i \exp(k_i + m_i \xi) \quad \text{for } i = g \text{ (gas) and } i = c \text{ (coal)}, \tag{2.8}$$

which is a special case of Equation (2.3). Proposition 2(2) describes the structure of the market bid stack function in this case – see also Carmona et al. [4, Corollary 1 and Table 1]. For example, in the case where both fuels are price setters, we have

$$S_t = b(R_t, F_t) = (C_t)^{\alpha_c} (G_t)^{\alpha_g} \exp(\beta + \gamma R_t), \tag{2.9}$$

where the constants are $\alpha_c = m_g/(m_c + m_g)$, $\alpha_g = 1 - \alpha_c$, $\beta = (k_c m_g + k_g m_c)/(m_c + m_g)$ and $\gamma = (m_c m_g)/(m_c + m_g)$. The explicit forms of S_t are obtained by plugging in the bid stack functions given in Equation (2.8) into Theorem 2.4 and Proposition 2(1). For example, Equation (2.9) is obtained by calculating the inverses in row 5 of the table in Proposition 2(1).

(B) (No conventional fuels with and without renewables). In his fundamental paper, Barlow [1] proposes the following model

$$S_t = \begin{cases} \left(\frac{a_0 - D_t}{b_0}\right)^{1/\alpha} & \text{when } D_t < a_0 - \varepsilon_0 b_0, \\ \varepsilon_0^{1/\alpha} & \text{when } D_t \geqslant a_0 - \varepsilon_0 b_0, \end{cases}$$

where $a_0 > 0$ is the maximal market supply, and $\epsilon_0, b_0 > 0$. As demand D_t approaches the maximum supply, the price S_t is capped at $\varepsilon_0^{1/\alpha}$, which is the maximal price when α is negative. Wagner [21] models S_t as a function of residual demand, that is $S_t = q(R_t)$ for some function q. To model peak hours, he uses a function similar to the one proposed by Barlow [1] with $\alpha = -1$. He also proposes models for the dynamics of renewable energy generation cf. also Grindel et al. [11].
Both Wagner [21] and Barlow [1] model the price of electricity, S_t, as a function of demand or residual demand, which aligns with our approach. However, they do not consider fuel inputs, meaning $n = 0$ in their cases. In other words, these are fundamental structural models where fuels are not incorporated explicitly. In contrast, Theorem 2.4 focuses on scenarios where $n \geqslant 1$, incorporating fuel effects into the analysis.

(C) (Fuels separated in terms of prioritization). Consider $n = 2$ fuels with prices $F_t = (F_t^1, F_t^2)$. Let the fuel bid stack functions b_1 and b_2 be given as in Equation (2.3), where q_1 and q_2 are strictly increasing and continuous. We also assume $\bar{b}_1(F_t^1) < \underline{b}_2(F_t^2)$ for all t. This implies that fuel 1 is always prioritized over fuel 2 when applying the merit order principle. For simplicity, assume that the fuel prices are proportional: $F_t^1 = \alpha F_t^2$ for all t for some $\alpha > 0$. By Proposition 2(2) and Equation (2.6), the electricity price is given by $S_t = F_t^2 \tilde{q}(R_t)$, where $\tilde{q}$ is defined in Equation (2.4). Explicitly, this formulation shows that

$$S_t = F_t^2 \left[\alpha q_1(R_t) \mathbb{1}_{\{R_t \in (0, \bar{\xi}_1]\}} + q_2(R_t - \bar{\xi}_1) \mathbb{1}_{\{R_t \in (\bar{\xi}_1, \bar{\xi}]\}}\right]. \tag{2.10}$$

This setup demonstrates that applying a single-fuel model to a scenario involving multiple fuels with clear prioritization can result in a discontinuous 'heat rate' function, $\xi \mapsto \tilde{q}(\xi)$, at points where ξ reaches the maximum capacity of a fuel.

Remark 2.7. Note that Equation (2.9) is multiplicative in fuels G_t and C_t. This multiplicative relationship depends on the choice of the fuel bid stack functions. For example, take the bid stack functions to be $b_i(\xi, f) = f\xi^\alpha$, for some $\alpha > 0$ and $i = 1, 2$. Then taking the necessary inverses, as per row 5 of Proposition 2((1)), one obtains that when both fuels are price setters, the price of electricity is given by $S_t = b(R_t, F_t) = (F_t^1 + F_t^2)^{-\alpha} R_t^\alpha$. Hence it is not multiplicative in the given fuels $F_t = (F_t^1, F_t^2)$.

3 Identification of Fundamental Variables and Functional Form

In this section, we evaluate the predictive performance of structural models derived from the theory in Section 2, using data from the German electricity market. We first describe the candidate models (Subsection 3.1) and then present the estimation and evaluation methods and the data (Subsections 3.2 and 3.3). In Subsection 3.4, we assess out-of-sample performance when each model is trained on one year of data and used to predict the subsequent year. Subsection 3.5 extends the analysis to multi-year training. Our results indicate that theory-driven, single-conventional-fuel models based on Theorem 2.4((2)) deliver the most robust predictive performance, especially under multi-year training.

Building on Theorem 2.4, we consider a model framework

$$S_t = b(R_t, F_t),$$

where the spot price of electricity S_t is in $\mathbb{R}$, R_t is the residual demand, and F_t is a $(d-1)$-dimensional vector consisting of $d-1$ fuel prices and $b : \mathbb{R}^d \to \mathbb{R}$. Within this framework, we specify 12 different models that differ in the choice of b and fuel variables F_t to identify the specification that best fits the data.

We evaluate performance using a rolling-origin approach, where models are trained on one or more consecutive past calendar years and used to predict the subsequent year. Out-of-sample performance is assessed using RMSE and MAE.

3.1 Candidate models

In Germany, gas-fired power plants typically serve as the marginal fuel in the electricity market. However, coal remains significant, particularly during periods of high demand. Historically, coal-fired plants were used more than gas, but this is changing with the green transition. We therefore consider models where the fuels used are gas and/or coal. Let G_t and C_t represent the market prices of natural gas and coal per MWh at time t, respectively. Additionally, let EUA_t denote the market price of European Union Allowances (EUAs) per ton at time t[4].

To incorporate CO_2 emissions, we define $\overline{G}_t$ and $\overline{C}_t$ as the fuel prices adjusted for EUA costs, given by:

$$\overline{G}_t = G_t + 0.189 \cdot EUA_t, \quad \text{and} \quad \overline{C}_t = C_t + 0.34 \cdot EUA_t.$$

The coefficients 0.189 and 0.34 represent the estimated CO_2 emissions (in tons per MWh) for gas- and coal-fired power plants, respectively, averaged over time. As noted in Section 2, a substantial component of the marginal cost of conventional fossil fuel power plants arises from the cost of EUA certificates. As illustrated in Figure 4.3, the prices of these certificates have increased significantly in recent years. Given this, we aim to investigate whether incorporating these fundamental variables enhances the accuracy of electricity price predictions.

Let α be a parameter, and let q, u, and v be real-valued functions. The working models are presented in Tables 3.2 and 3.3. Most of these models are theory-driven, following Section 2: under the assumptions of equilibrium and market efficiency, most working models describe electricity prices with one fuel or multiple fuels with proportional prices (see Theorem 2.4). Table 3.1 indicates that proportionality between coal and gas prices is a reasonable approximation.

[4] An EUA (European Union Allowance) is a tradable carbon asset issued to companies regulated under the EU Emissions Trading System. It grants the right to emit one metric ton of CO_2 equivalent, enabling emissions reductions through a cap-and-trade mechanism.

Year	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**
$\mathrm{Corr}(G_t, C_t)$	0.8016	0.5702	0.4761	0.8012	0.8138	0.728	0.6203	0.883
$\mathrm{Corr}(\overline{G}_t, \overline{C}_t)$	0.8392	0.6435	0.8612	0.5579	0.7539	0.8715	0.6139	0.6737

Table 3.1: Empirical correlations for gas and coal prices

Model	Formula
$M^{G,C}$	$S_t = v(C_t)u(G_t)q(R_t)$
M^G	$S_t = u(G_t)q(R_t)$
$M^G_{1.1}$	$S_t = \overline{G}_t q(R_t)$
$M^G_{2.1}$	$S_t = G_t^\alpha q(R_t)$
$M^G_{2.2}$	$S_t = G_t q(R_t)$
$M^G_{2.3}$	$S_t = G_t \exp(k + mR_t)$
M^C	$S_t = v(C_t)q(R_t)$
$M^C_{1.1}$	$S_t = \overline{C}_t q(R_t)$
$M^C_{2.1}$	$S_t = C_t^\alpha q(R_t)$
$M^C_{2.2}$	$S_t = C_t q(R_t)$
M	$S_t = q(R_t)$
$M_{1.1}$	$S_t = \left(\frac{a_0 - R_t}{b_0}\right)^{1/\alpha}$

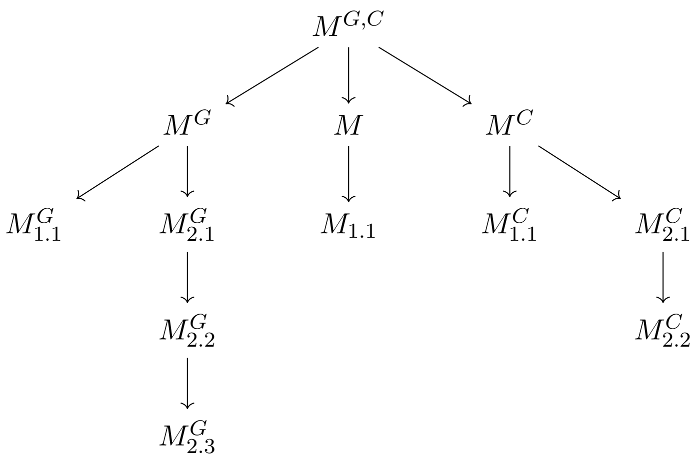


Table 3.2: Structural models library

Benchmark		Machine learning	
M^B	$S_t = S_{t-1}$	M_1^{ML}	`MLPRegressor`
		M_2^{ML}	`AdaBoostRegressor`

Table 3.3: The benchmark and the machine-learning models

We now take a closer look at the models in Table 3.2. Note that $M^G_{2.3}$ corresponds to a one-fuel exponential bid stack model with renewables cf. Example 2.6((A)) and Equation (2.7), where the heat rate is an exponential function. A slight generalization, where we use a general function q instead of an exponential one, leads to model $M^G_{2.2}$. Model $M^G_{1.1}$ is another alteration, in which we use the net gas price $\overline{G}_t$ instead of the gas price. In other words, models $M^G_{1.1}$ and $M^G_{2.2}$ are general one-fuel models in efficient markets (cf. Equation (2.7)). Going a step further and using $u(G_t)$ instead of $\overline{G}_t$ leads to M^G. Similarly, model $M^{G,C}$ is a generalisation of the 2-fuel exponential bid stack introduced in Carmona et al. [4], with renewables added (see Equation (2.9) in Example 2.6, in the case where both gas and coal are price setters). If we consider coal instead of gas, we obtain the models $M^C_{2.2}$, $M^C_{1.1}$, and M^C.

Furthermore, note that model $M_{1.1}$ is the Barlow [1] model (see Example 2.6(B)). Model M is a generalization of this model, where instead of a power function, we use a general increasing function q.

Moreover, Table 3.3 presents three additional benchmark models used to assess the predictive performance of the 12 models under consideration. The first is the naive benchmark $S_t = S_{t-1}$, where tomorrow's price is predicted to be equal to today's. This model relies on the lagged electricity price as an input, which is not included in any of the other models, as our focus is on understanding electricity prices based on its fundamental drivers and not the past electricity prices. Hence, this benchmark may have an informational advantage.

In addition, we include two machine-learning models: `AdaBoostRegressor` and `MLPRegressor` (see Scikit-learn [18], Scikit-learn [19]). These were selected using the Python package *LazyPredict*, which identified them as the top two performers out of more than 50 state-of-the-art machine-learning algorithms.

3.2 Estimating the parameters and measuring predictive performance

To estimate the parameters of the models in Table 3.2, we use ordinary least squares (OLS) on the in-sample data. That is, we estimate b by minimizing RSS, the sum of squared differences between observed and predicted values,

$$RSS = \sum_{t=1}^{n} \left(S_t - b(R_t, F_t)\right)^2 ,$$

where n is the length of the in-sample period and (R_t, F_t, S_t) denote the in-sample observations.

For the functions q, u, and v, we fit natural cubic splines with 4 knots, with knot locations chosen automatically at the 20%, 40%, 60%, and 80% quantiles. We do this to obtain a flexible class of functions and to test whether an exponential function, as suggested in Equation (2.8), or a power function, as in Example 2.6(B), fits the data. The use of natural cubic splines is recommended by James et al. [13] as a convenient way of modeling a nonlinear relationship and is also used in Burger et al. [2].

For performance metrics on out-of-sample data (one year), we use Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Our discussion emphasizes RMSE. Results for the alternative error metric are qualitatively similar and are reported in Appendix Tables 4.1 and 4.2 in the Appendix. Specifically, the RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{t=n+1}^{n+m} (S_t - \widehat{b}(R_t, F_t))^2},$$

where $\widehat{b}$ denotes the estimated function b trained on the in-sample data, and m is the length of the out-of-sample period (one year).

Implementation details for the benchmark machine-learning models: We use the Python implementations `AdaBoostRegressor` from `sklearn.ensemble` and `MLPRegressor` from `sklearn.neural_network`. Both models were trained on historical data incorporating residual demand and gas prices. We additionally considered augmenting the feature set with coal and EUA prices; in practice, this increased overfitting and worsened predictive performance, as reflected in higher RMSE. We therefore exclude these variables from the benchmark specifications.

3.3 Data description

We focus on the German electricity market using daily data from January 2016 to December 2023. All data used in this study are proprietary but can be obtained directly from established market sources, which are listed below. To ensure clarity, we fix the observation time at 12 each day. Preliminary checks indicate that using alternative hours produces equivalent results, so this restriction does not affect conclusions regarding the structural models' performance. After excluding weekends, major public holidays,[5] and two identified anomalies (one in 2018 and another in 2020), the final dataset consists of 1996 daily observations. The key components of our dataset are as follows:

- Electricity prices: German spot prices from EPEX Spot;
- Gas prices: Day-ahead THE (Trading Hub Europe) data from ICIS Heren;
- Coal prices: API2 Rotterdam coal prices from ICE (Intercontinental Exchange);
- Carbon prices: December EUA contracts from ICE;
- Demand and renewable generation: Actual hourly demand, wind generation, and solar generation from Volue.

Although forecast data are available and relevant–since market bids are based on forecasts–we use actual production data for simplicity and consistency. Preliminary comparisons indicate that using forecasted production yields similar predictive outcomes, supporting our choice for this study.

For a detailed description of the data, refer to Table 3.4. Also note that there are 13 instances of negative spot prices but no negative values for residual demand.

	R_t	G_t	S_t	C_t	EUA_t	$\overline{G}_t$	$\overline{C}_t$
mean	42613.33	36.16	78.06	12.49	37.07	42.77	25.10
std	13198.78	41.95	84.61	9.00	30.75	45.81	17.84
min	3140.56	3.71	-79.74	4.35	3.93	7.04	6.51
25%	33897.46	14.19	33.84	6.61	7.88	16.19	12.73
50%	42945.35	19.16	46.40	9.62	25.11	22.27	15.83
75%	51131.79	37.40	84.85	13.85	68.66	51.71	40.72
max	77120.97	307.88	702.33	49.61	100.34	324.32	75.28

Table 3.4: Summary statistics of daily data for the German electricity market, 2016-2023.

3.4 Predicting one year ahead

We train each of the 15 models on a single year of data and use them to predict the following year. The corresponding RMSE values are shown in Table 3.5. The lowest RMSE in each row is highlighted in green, and the next two lowest are marked in orange.

We comment on four aspects.

A. Remarkably, the performance of structural models is comparable to (if not better than) the two chosen machine-learning models M_1^{ML} and M_2^{ML}. See, for example,

[5]Source of holidays: [9]

Trained on	**Predicting**	M_2^{ML}	M_1^{ML}	$M_{2.1}^C$	$M_{1.1}^G$	$M_{1.1}^C$	$M_{2.3}^G$	$M_{1.1}$	M^B	$M_{2.2}^C$	M^C	M^G	$M_{2.2}^G$	$M_{2.1}^G$	$M^{G,C}$	M
2016	2017	10.09	9.83	8.39	7.68	10.25	40.97	9.85	12.71	11.00	8.45	7.06	7.55	7.25	7.58	10.44
2017	2018	16.69	12.44	17.52	8.11	8.39	48.48	17.07	13.38	13.73	22.39	13.19	8.34	16.93	21.73	16.85
2018	2019	8.35	9.15	16.99	10.88	6.19	40.56	8.60	11.15	12.45	23.90	17.71	15.07	17.89	27.61	7.74
2019	2020	9.34	14.64	7.29	8.99	5.72	34.41	10.73	12.39	7.10	9.58	7.59	12.41	7.48	7.51	8.68
2020	2021	98.54	79.86	61.04	43.48	58.48	130.45	103.87	42.92	59.74	γ_1	γ_2	73.65	90.61	γ_3	101.50
2021	2022	148.80	55.13	104.26	72.86	130.70	261.07	183.49	78.79	109.01	156.39	80.12	62.79	88.98	141.61	192.60
2022	2023	62.64	66.91	28.67	21.69	73.66	95.87	137.28	38.09	29.45	32.21	24.66	26.86	30.26	25.80	126.78

Table 3.5: RMSE values. The symbols $\gamma_1 = 1{,}092.79$, $\gamma_2 = 2{,}548{,}238{,}433.31$, and $\gamma_3 = 181{,}158.35$ are used in the table because the magnitudes are so large that they are uninformative.

Figure 3.1, which compares M_1^{ML} with $M_{1.1}^G$ when predicting 2018. The structural model is much better at predicting spikes, which are an important stylized fact of electricity prices. The machine-learning models are in top three only in two cases.

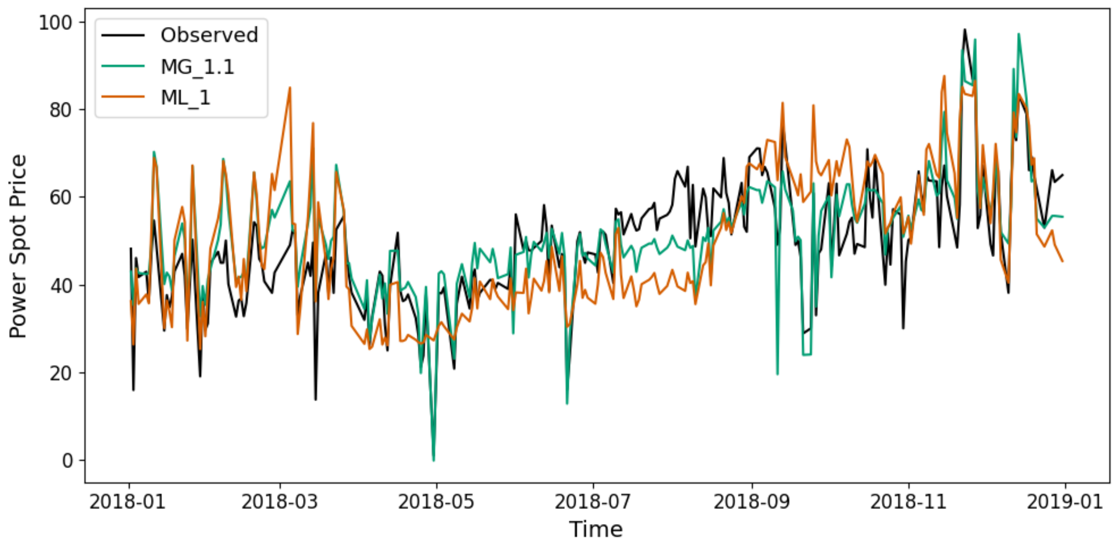


Figure 3.1: The black line represents observed power spot prices at 12:00 in 2018, while the green (orange) line represents predicted values for 2018 obtained from model $M_{1.1}^G$ (M_1^{ML}), trained on the 2017 dataset.

Specifically, when predicting 2019, M_2^{ML} performs well; however, it is outperformed by two of our structural models, $M_{1.1}^C$ and the simple M. Moreover, when predicting 2022, M_1^{ML} has the lowest RMSE, followed by $M_{2.2}^G$.

B. The structural models outperform the benchmark model M^B. Model M^B simply shifts the observed values by one day, leading to errors due to the volatile nature of electricity prices. If prices were constant at noon every day, this model would be error-free. However, due to volatility, structural models like $M_{1.1}^G$ generally perform better, except for 2021, where $M_{1.1}^G$ had an RMSE marginally larger than M^B.

C. Among the structural models, $M_{1.1}^G$ and $M_{1.1}^C$, both theory-driven, stand out when predicting a year based on the previous year's data. In earlier years, the coal-driven model $M_{1.1}^C$ performed better, while in later years, the gas-driven model $M_{1.1}^G$ performed best. Both models, which include the EUA price term, outperform their counterparts without EUA ($M_{2.2}^G$ and $M_{2.2}^C$).

Model $M^G_{2.2}$ only outperforms $M^G_{1.1}$ when predicting 2022, primarily due to the extreme gas price surges in 2021 and 2022, as $M^G_{2.2}$ is more sensitive to fluctuations in gas prices. This increased sensitivity arises because, unlike $M^G_{1.1}$, which uses the EUA-adjusted gas price $\overline{G}_t$ (i.e., $G_t + 0.189\, EUA_t$), $M^G_{2.2}$ uses the unadjusted gas price directly as the fuel price. The inclusion of EUA prices in $M^G_{1.1}$ effectively dampens its sensitivity to gas price variations. Consequently, the larger the price jumps, the higher the RMSE for $M^G_{1.1}$, as illustrated in Figure 3.2. It is also interesting to point

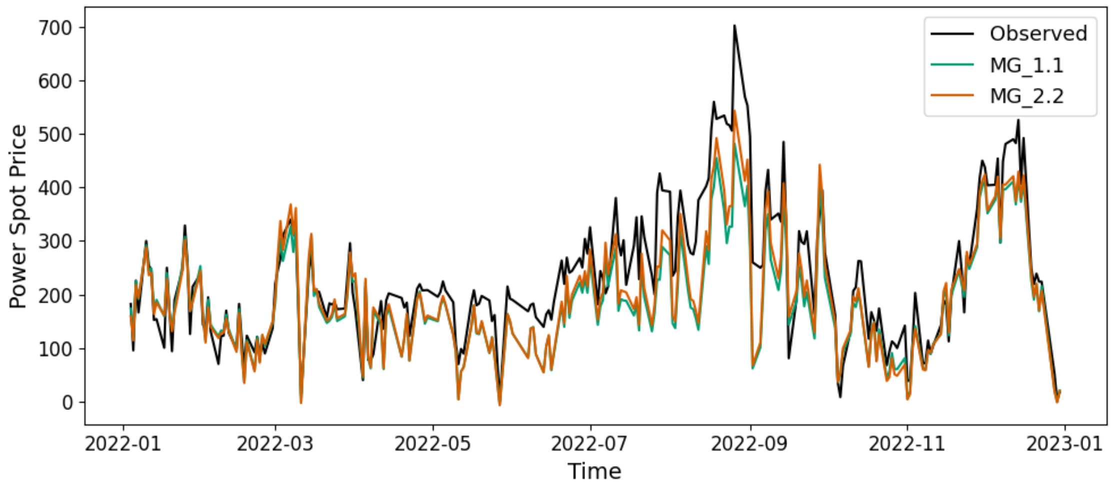


Figure 3.2: The black line represents observed power spot prices at 12:00 in 2022, while the green (orange) line represents predicted values for 2022 obtained from model $M^G_{1.1}$ ($M^G_{2.2}$), trained on the 2021 dataset.

out that more general models like M^G and $M^{G,C}$ perform well only when market dynamics remain stable. As they are prone to overfitting, they produce large errors when predicting years in which market trends change, such as in 2021. In 2021 and 2022, gas prices experienced significant volatility and price increases, driven by geopolitical tensions (the war in Ukraine), which disrupted global energy markets and sharply increased European gas prices.

D. A final note concerns the literature-based models, specifically $M^G_{2.3}$ (the single-fuel model with gas, renewables, and an exponential heat rate based on Carmona et al. [4]) and $M_{1.1}$ (the model from Barlow [1], adapted to use residual demand instead of demand). As expected, these simpler models are outperformed by their more general counterparts.

3.5 Stability over time

Next, we evaluate the stability of the predictions across different training time spans. Specifically, we analyze how prediction performance varies under these conditions. Again, we refer to the RMSE table (see Table 3.6), which summarizes the results across the different training ranges.

Among the top performing models $M^G_{1.1}$, $M^C_{1.1}$, $M^G_{2.2}$, and $M^{ML}_{1.1}$, the most robust and consistent performer is $M^G_{1.1}$. Structural models, in general, exhibit improved accuracy when

Trained on	**Predicting**	M_2^{ML}	M_1^{ML}	$M_{2.1}^C$	$M_{1.1}^G$	$M_{1.1}^C$	$M_{2.3}^G$	$M_{1.1}$	M^B	$M_{2.2}^C$	M^C	M^G	$M_{2.2}^G$	$M_{2.1}^G$	$M^{G,C}$	M
2016	2017	10.09	9.83	8.39	7.68	10.25	40.97	9.85	12.71	11.00	8.45	7.06	7.55	7.25	7.58	10.44
2017	2018	16.69	12.44	17.52	8.11	8.39	48.48	17.07	13.38	13.73	22.39	13.19	8.34	16.93	21.73	16.85
2016-2017	2018	17.80	11.11	15.67	8.16	8.39	48.50	19.49	13.38	12.37	18.96	11.58	8.22	10.35	16.97	19.12
2018	2019	8.35	9.15	16.99	10.88	6.19	40.56	8.60	11.15	12.45	23.90	17.71	15.07	17.89	27.61	7.74
2016-2018	2019	9.63	12.79	12.67	10.38	7.57	40.53	7.20	11.15	14.24	12.38	14.90	15.08	15.16	14.40	7.71
2019	2020	9.34	14.64	7.29	8.99	5.72	34.41	10.73	12.39	7.10	9.58	7.59	12.41	7.48	7.51	8.68
2016-2019	2020	8.41	12.29	8.41	10.75	7.53	34.80	9.49	12.39	12.77	7.78	31.32	16.13	11.73	11.84	7.38
2020	2021	98.54	79.86	61.04	43.48	58.48	130.45	103.87	42.92	59.74	1,092.79	2,548,238,433.31	73.65	90.61	181,158.35	101.50
2016-2020	2021	93.68	32.26	92.95	23.59	52.52	131.25	102.92	42.92	77.03	1,988.45	1,221.94	24.96	79.53	396.57	102.16
2018-2020	2021	93.22	37.00	89.28	24.37	55.48	131.16	100.21	42.92	71.81	247.34	89.94	27.91	81.97	99.70	99.13
2019-2020	2021	97.88	37.58	83.45	36.04	57.15	130.76	101.85	42.92	63.27	84.20	563.88	52.90	87.27	6,963,196,320.70	100.07
2021	2022	148.80	55.13	104.26	72.86	130.70	261.07	183.49	78.79	109.01	156.39	80.12	62.79	88.98	141.61	192.60
2016-2021	2022	141.95	63.53	92.69	63.99	135.02	261.05	224.30	78.79	129.26	202.59	79.81	56.04	72.61	149.64	227.08
2018-2021	2022	149.97	52.69	93.29	63.50	134.96	261.04	217.79	78.79	118.42	198.56	80.06	54.74	78.72	147.61	221.79
2019-2021	2022	147.79	56.59	97.24	61.92	134.33	261.04	214.87	78.79	110.20	183.26	78.33	52.70	84.15	131.63	219.12
2020-2021	2022	146.39	53.10	101.69	66.29	133.42	261.05	207.48	78.79	109.46	163.19	76.01	56.46	88.20	106.47	213.69
2022	2023	62.64	66.91	28.67	21.69	73.66	95.87	137.28	38.09	29.45	32.21	24.66	26.86	30.26	25.80	126.78
2016-2022	2023	26.88	33.72	31.28	21.31	55.68	95.91	39.83	38.09	22.21	27.86	23.61	27.36	26.53	25.23	38.83
2018-2022	2023	27.30	33.12	29.38	21.52	57.54	95.90	31.96	38.09	24.23	33.42	26.85	27.41	25.84	27.87	33.64
2019-2022	2023	28.18	34.87	28.89	21.66	59.80	95.90	29.31	38.09	27.84	33.75	26.14	27.40	24.86	26.97	32.76
2020-2022	2023	27.79	30.19	31.61	21.95	60.50	95.92	52.14	38.09	31.92	33.73	22.39	28.14	27.35	23.43	55.40
2021-2022	2023	26.49	30.33	30.50	21.63	65.22	95.91	72.77	38.09	31.29	36.27	22.31	27.59	28.27	23.93	72.03

Table 3.6: RMSE values.

trained on longer time spans. This trend is clearly illustrated by $M_{1.1}^G$: when predicting the year 2022, its RMSE decreases by 10 when trained on data from 2016-2021, compared to using data from 2021 alone (see Figure 3.3). We also observe that the benchmark model

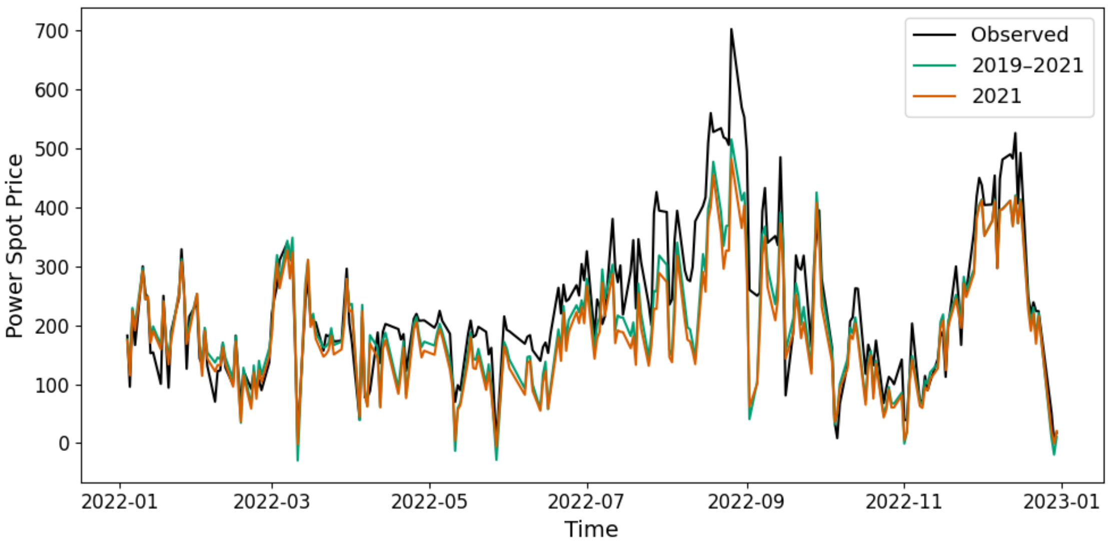


Figure 3.3: The black line represents observed power spot prices at 12:00 in 2022, while the green (orange) line represents predicted values for 2022 obtained from model $M_{1.1}^G$ trained on the 2019-2021 (2021) dataset.

$M_{1.1}$ from Barlow [1] achieves the best performance in one scenario. Specifically, when predicting the year 2019 using training data from 2016-2018, $M_{1.1}$ outperforms the other models, although $M_{1.1}^C$ demonstrates a comparably strong performance. For comparison, see Figure 3.4, which shows that $M_{1.1}^C$ captures spikes notably better.

Taken together, these findings confirm and strengthen the conclusions drawn in the previous subsection: the structural models—particularly $M_{1.1}^G$ in the later years and $M_{1.1}^C$ in earlier years—consistently deliver the most accurate and robust predictions. Moreover, their predictive performance improves further when trained on extended historical datasets,

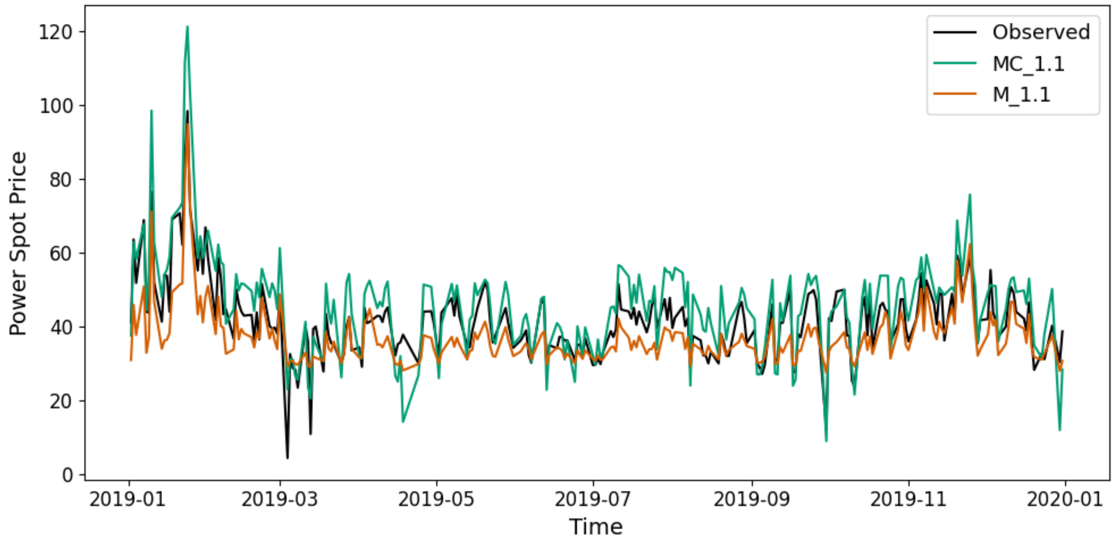
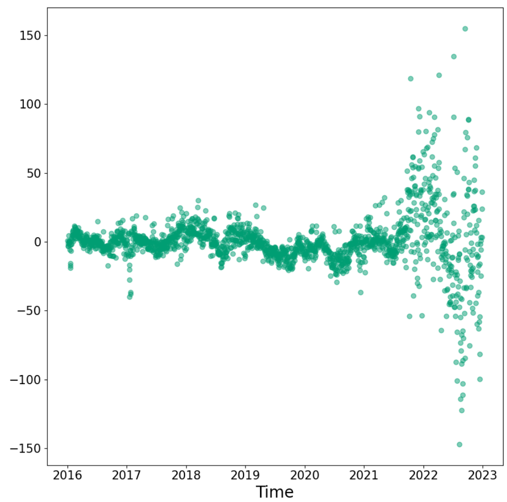


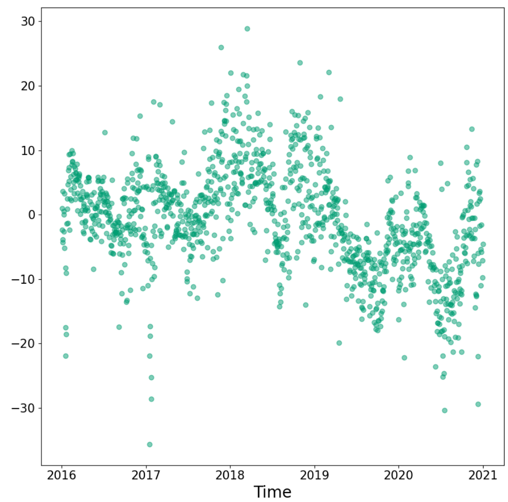
Figure 3.4: The black line represents observed power spot prices at 12:00 in 2019, while the green (orange) line represents predicted values for 2019 obtained from model $M^C_{1.1}$ ($M_{1.1}$), trained on the 2016–2018 dataset.

underscoring the value of incorporating multiple years of data in model estimation.

Remark 3.1. We examine the residuals from model $M^G_{1.1}$ under two training windows–2016-2022 and 2016-2020 (See Figure 3.5). The residuals display some seasonality, and in the 2016-2022 window the residual variance appears to rise with fuel prices. Recall that in 2022 (and adjacent months) fuel prices spiked unusually, coinciding with the emergence of new energy-market patterns. We do not pursue a detailed residual analysis here, as the in-sample estimates appear reasonably robust to moderate heteroskedasticity (i.e., inhomogeneous variance).



(a) Trained on 2016-2022.



(b) Trained on 2016-2020.

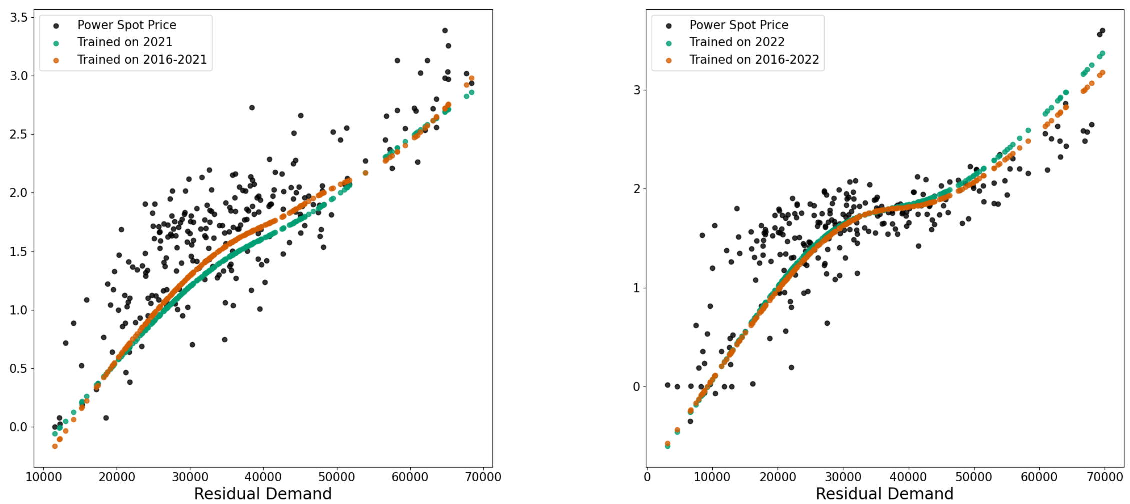
Figure 3.5: Residuals for $M^G_{1.1}$.

4 Summary of Data Analysis Results

Our analysis demonstrates that the choice of model type and the incorporation of domain knowledge are critical for accurately predicting electricity prices. Key observations include:

- When trained on at least two years of data, the theory-driven models $M^G_{1.1}$ and $M^C_{1.1}$ consistently rank among the top three performers.
- In recent years, gas-driven models have achieved the best performance.
- Models with a large number of parameters, such as M^C, M^G, and $M^{G,C}$, require caution. They are prone to overfitting and therefore struggle to adapt to changes in price patterns - precisely the dynamics we aim to capture.
- Structural models perform comparably to, and in some cases slightly better than, machine-learning models on this type of dataset, which is a noteworthy result.
- Historically, models without explicit fuel inputs, such as the Barlow [1] model, have performed well. However, as the energy mix evolves (e.g., with increased renewable supply), these models tend to underperform relative to more structured models like $M^G_{1.1}$. Nonetheless, the significance of Barlow [1] lies in introducing the structural modeling paradigm for electricity markets.

Among the structural models tested, the simple one-fuel model $M^G_{1.1}$ exhibits the most robust overall performance and effectively captures key information about spot price movements and market trends. Figure 4.1 illustrates the estimated heat rate function for $M^G_{1.1}$, i.e., the estimated q in $S_t = \overline{G}_t q(R_t)$. The plot indicates that using a purely exponential or power function for the heat rate does not provide the best fit to the data.



(a) Heat function for $M^G_{1.1}$ trained on 2021 (green) vs. 2016-2021 (orange).

(b) Heat function for $M^G_{1.1}$ trained on 2022 (green) vs. 2016-2022 (orange).

Figure 4.1: Estimated heat function for $M^G_{1.1}$, trained on either a single year or multiple years. The underlying power spot price data is scaled by the gas price to match the scale of the heat functions.

The sharp slope observed at the start of the graph may be explained by the remarks in Example 2.6(C). Specifically, another fuel, perhaps lignite, could induce a jump in the theoretical heat rate function (as in Equation (2.10)). The estimated curve in Figure 4.1 can therefore be interpreted as a smoothed representation of that jump.

Appendix

4.1 Proof of Proposition 2

Proof. The result in Part (2) is a direct consequence of Part (1) by substituting $f_1 = \alpha f_2$ into the formulas for $b(\xi, f)$ in each of the cases in Part (1). Therefore, we need only prove Part (1). Without loss of generality, we assume $\underline{b}_1(f_1) \leqslant \underline{b}_2(f_2)$.

Since q_i is assumed non-negative, strictly increasing, and continuous and $f_i > 0$, it is straightforward that for $i = 1, 2$, the fuel bid stack function $\xi \mapsto b_i(\xi, f_i) = f_i q_i(\xi)$ is strictly increasing and continuous from $[0, \overline{\xi}_i]$ onto $[\underline{b}_i(f_i), \overline{b}_i(f_i)]$, and hence an inverse exists and is given by $s \mapsto q_i^{-1}(s/f_i)$. Note that for $i = 1, 2$, the generalized inverse in Equation (2.2) is

$$b_i(\cdot, f_i)^{-1}(p) = \begin{cases} 0 & \text{if } p < \underline{b}_i(f_i) \\ q_i^{-1}(p/f_i) & \text{if } p \in [\underline{b}_i(f_i), \overline{b}_i(f_i)) \\ \overline{\xi}_i & \text{if } p \geqslant \overline{b}_i(f_i) \end{cases} \tag{4.1}$$

Let us first consider the case $\underline{b}_1(f_1) < \overline{b}_1(f_1) \leqslant \underline{b}_2(f_2) < \overline{b}_2(f_2)$. Then

$$b_1(\cdot, f_1)^{-1}(p) + b_2(\cdot, f_2)^{-1}(p) = \begin{cases} 0 & \text{if } p < \underline{b}_1(f_1) \\ b_1(\cdot, f_1)^{-1}(p) & \text{if } p \in [\underline{b}_1(f_1), \overline{b}_1(f_1)) \\ \overline{\xi}_1 & \text{if } p \in [\overline{b}_1(f_1), \underline{b}_2(f_2)) \\ \overline{\xi}_1 + b_2(\cdot, f_2)^{-1}(p) & \text{if } p \in [\underline{b}_2(f_2), \overline{b}_2(f_2)]. \end{cases}$$

As the fuel bid stack functions are increasing and continuous from $[0, \overline{\xi}_i]$ onto $[\underline{b}_i(f_i), \overline{b}_i(f_i)]$, the function $p \mapsto b_i(\cdot, f_i)^{-1}(p)$ restricted to $[\underline{b}_i(f_i), \overline{b}_i(f_i)]$ is strictly increasing and onto $[0, \overline{\xi}_i]$. It follows from Proposition 1 that for $\xi \in (0, \overline{\xi}]$,

$$\begin{aligned} b(\xi, f) &= \sup\{p \in \mathbb{R} : b_1(\cdot, f_1)^{-1}(p) + b_2(\cdot, f_2)^{-1}(p) < \xi\} \\ &= \begin{cases} b_1(\xi, f_1) & \text{if } \xi \in (0, \overline{\xi}_1] \\ b_2(\xi - \overline{\xi}_1, f_2) & \text{if } \xi \in (\overline{\xi}_1, \overline{\xi}]. \end{cases} \end{aligned}$$

Thus, for $\xi \in (0, \overline{\xi}_1]$ (which corresponds to $b_1(\xi, f_1) \leqslant \underline{b}_2(f_2)$), fuel 1 is the price setter while fuel 2 is not used, and for $\xi \in (\overline{\xi}_1, \overline{\xi}]$ (corresponding to $b_2(\xi - \overline{\xi}_1, f_2) > \overline{b}_1(f_1)$), fuel 2 is the price setter and fuel 1 is fully used. These conclusions are consistent with the information provided in Part (1) and Table 2.1.

To help with the proof in the next two cases, define

$$\xi_{i,1} = b_i(\cdot, f_i)^{-1}(\underline{b}_j(f_j)) \quad \text{and} \quad \xi_{i,2} = b_i(\cdot, f_i)^{-1}(\overline{b}_j(f_j)) \quad \text{for } i \neq j. \tag{4.2}$$

If fuel 1 initially serves as the price setter, then $\xi_{1,1}$ represents the quantity of fuel 1 used when fuel 2 enters as a price setter. Similarly, if fuel j (where $j = 1$ or 2) exits as a price setter first, then $\xi_{i,2}$ (where i is not j) indicates how much of fuel i is used at the point when fuel j exits.

Now assume that $\underline{b}_1(f_1) \leqslant \underline{b}_2(f_2) \leqslant \overline{b}_1(f_1) \leqslant \overline{b}_2(f_2)$. By Equation (4.1),

$$b_1(\cdot, f_1)^{-1}(p) + b_2(\cdot, f_2)^{-1}(p) = \begin{cases} 0 & \text{if } p < \underline{b}_1(f_1) \\ b_1(\cdot, f_1)^{-1}(p) & \text{if } p \in [\underline{b}_1(f_1), \underline{b}_2(f_2)) \\ q_1^{-1}(p/f_1) + q_2^{-1}(p/f_2) & \text{if } p \in [\underline{b}_2(f_2), \overline{b}_1(f_1)) \\ \overline{\xi}_1 + b_2(\cdot, f_2)^{-1}(p) & \text{if } p \in [\overline{b}_1(f_1), \overline{b}_2(f_2)]. \end{cases}$$

Restricted to $[\underline{b}_1(f_1), \overline{b}_2(f_2)]$, the function $p \mapsto b_1(\cdot, f_1)^{-1}(p) + b_2(\cdot, f_2)^{-1}(p)$ is continuous, strictly increasing and onto $[0, \overline{\xi}]$, which means that the generalized inverse reduces to the ordinary inverse, and hence by Proposition 1,

$$b(\xi, f) = \begin{cases} b_1(\xi, f_1) & \text{if } \xi \in \ (0, \xi_{1,1}] \\ [q_1^{-1}(\cdot/f_1) + q_2^{-1}(\cdot/f_2)]^{-1}(\xi) & \text{if } \xi \in (\xi_{1,1}, \overline{\xi}_1 + \xi_{2,2}] \\ b_2(\xi - \overline{\xi}_1, f_2) & \text{if } \xi \in (\overline{\xi}_1 + \xi_{2,2}, \overline{\xi}]. \end{cases}$$

In the first of these cases (corresponding to $b_1(\xi, f_1) \leqslant \underline{b}_2(f_2)$), fuel 1 is the price setter and fuel 2 is not used. In the second case, where none of the restrictions in the first four rows of the table in Part (1) are satisfied, both fuels are price setters. In the third case (corresponding to $b_2(\xi - \overline{\xi}_1, f_2) > \overline{b}_1(f_1)$), fuel 2 is the price setter while fuel 1 is fully used. Hence, this second case of $\underline{b}_1(f_1)$, $\overline{b}_1(f_1)$, $\underline{b}_2(f_2)$ and $\overline{b}_2(f_2)$ aligns with the data presented in Part (1) and Table 2.1.

Finally assume that $\underline{b}_1(f_1) \leqslant \underline{b}_2(f_2) < \overline{b}_2(f_2) \leqslant \overline{b}_1(f_1)$. By arguments as above, it follows that

$$b(\xi, f) = \begin{cases} b_1(\xi, f_1) & \text{if } \xi \in \ (0, \xi_{1,1}] \\ [q_1^{-1}(\cdot/f_1) + q_2^{-1}(\cdot/f_2)]^{-1}(\xi) & \text{if } \xi \in (\xi_{1,1}, \overline{\xi}_2 + \xi_{1,2}] \\ b_1(\xi - \overline{\xi}_2, f_1) & \text{if } \xi \in (\overline{\xi}_2 + \xi_{1,2}, \overline{\xi}]. \end{cases}$$

As above, this case of $\underline{b}_1(f_1)$, $\overline{b}_1(f_1)$, $\underline{b}_2(f_2)$ and $\overline{b}_2(f_2)$ is also consistent with Part (1) and Table 2.1. This covers all the cases and hence the desired result follows. □

4.2 Market development throughout the years

For completeness, Figure 4.2 presents the evolution of power spot prices and marginal costs of gas and coal power plants at 12:00 for selected years. Under the competitive-market assumption, a commonly used consensus formula for these costs is:

$$mc_t^G = \frac{G_t + 0.184EUA_t}{0.5} \quad \text{and} \quad mc_t^C = \frac{C_t + 0.34EUA_t}{0.38}.$$

All the prices are presented as EUR/MWh. Overall, the consensus-based marginal cost for gas-fired plants tracks power-price fluctuations better than the corresponding coal-based measure. Figure 4.3 shows the evolution of EUA certificate prices, also converted to EUR/MWh for comparability. The figure also overlays the day-ahead gas price (blue), illustrating the importance of EUA prices.

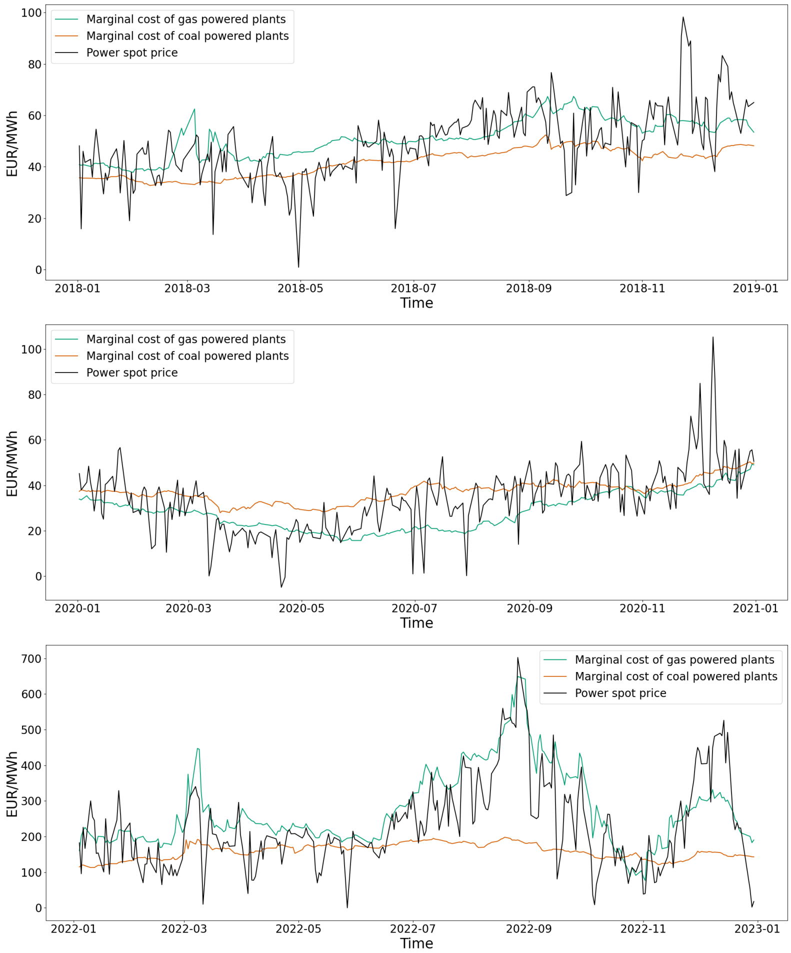


Figure 4.2: Marginal cost of coal (orange) and gas (green) powered plants together with spot power price (black) development in the years 2018, 2020, and 2022.

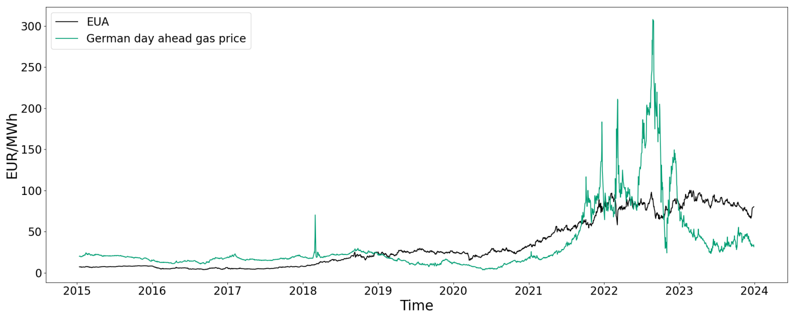


Figure 4.3: Development of the EUA certificate (black) and the German gas (THE) prices (green) from 2015 to 2024.

4.3 MAE Tables

The main measure of performance in our data analysis are the RMSE values for different models and time horizons. However, here we present the MAE values, to show that they yield similar results.

Trained on	**Predicting**	M_2^{ML}	M_1^{ML}	$M_{2.1}^C$	$M_{1.1}^G$	$M_{1.1}^C$	$M_{2.3}^G$	$M_{1.1}$	M^B	$M_{2.2}^C$	M^C	M^G	$M_{2.2}^G$	$M_{2.1}^G$	$M^{G,C}$	M
2016	2017	6.48	6.95	5.15	5.42	7.24	37.69	7.78	8.92	8.30	5.47	5.08	5.34	5.14	5.56	7.97
2017	2018	14.04	9.65	14.74	6.38	6.18	46.34	14.35	9.16	11.46	18.10	10.95	6.47	14.27	17.40	14.15
2018	2019	6.64	7.22	15.92	9.50	5.00	39.12	6.99	7.65	11.37	20.73	16.35	13.52	16.28	25.86	6.48
2019	2020	7.53	11.72	5.84	6.94	4.45	31.39	8.37	8.77	5.35	7.49	6.02	9.61	6.01	6.03	7.01
2020	2021	63.04	50.77	28.48	31.83	29.32	102.46	70.42	25.36	27.88	379.24	182,849,459.04	51.28	58.71	24,463.51	68.13
2021	2022	106.00	43.41	70.73	53.94	89.80	228.66	145.02	57.47	71.95	113.77	62.13	47.75	65.45	106.40	156.13
2022	2023	55.47	47.34	19.81	16.38	64.37	87.18	132.76	26.60	20.13	21.54	19.81	22.57	26.25	21.22	114.94

Table 4.1: MAE values corresponding to Table 3.5.

Trained on	**Predicting**	M_2^{ML}	M_1^{ML}	$M_{2.1}^C$	$M_{1.1}^G$	$M_{1.1}^C$	$M_{2.3}^G$	$M_{1.1}$	M^B	$M_{2.2}^C$	M^C	M^G	$M_{2.2}^G$	$M_{2.1}^G$	$M^{G,C}$	M
2016	2017	6.48	6.95	5.15	5.42	7.24	37.69	7.78	8.92	8.30	5.47	5.08	5.34	5.14	5.56	7.97
2017	2018	14.04	9.65	14.74	6.38	6.18	46.34	14.35	9.16	11.46	18.10	10.95	6.47	14.27	17.40	14.15
2016-2017	2018	15.13	8.60	13.14	6.47	6.50	46.37	16.58	9.16	10.08	15.70	9.36	6.45	8.17	13.35	16.25
2018	2019	6.64	7.22	15.92	9.50	5.00	39.12	6.99	7.65	11.37	20.73	16.35	13.52	16.28	25.86	6.48
2016-2018	2019	8.53	11.56	11.75	9.05	5.94	39.10	5.60	7.65	13.33	11.65	13.45	13.57	13.65	13.23	6.10
2019	2020	7.53	11.72	5.84	6.94	4.45	31.39	8.37	8.77	5.35	7.49	6.02	9.61	6.01	6.03	7.01
2016-2019	2020	6.57	9.87	6.67	8.91	5.79	31.74	7.37	8.77	10.83	6.29	22.54	13.90	9.74	9.41	5.75
2020	2021	63.04	50.77	28.48	31.83	29.32	102.46	70.42	25.36	27.88	379.24	182,849,459.04	51.28	58.71	24,463.51	68.13
2016-2020	2021	59.72	23.08	60.00	14.81	27.29	103.30	68.64	25.36	45.10	547.81	254.13	17.30	51.04	155.86	68.11
2018-2020	2021	58.82	24.82	55.50	15.24	27.79	103.20	65.20	25.36	38.75	97.51	56.31	18.01	51.16	49.37	64.35
2019-2020	2021	62.72	23.65	50.24	25.26	28.36	102.78	67.60	25.36	29.67	50.82	133.18	33.98	55.46	479,598,076.20	66.02
2021	2022	106.00	43.41	70.73	53.94	89.80	228.66	145.02	57.47	71.95	113.77	62.13	47.75	65.45	106.40	156.13
2016-2021	2022	100.97	51.89	69.46	47.23	94.41	228.64	189.54	57.47	93.28	154.97	58.15	42.31	54.53	113.32	193.40
2018-2021	2022	107.52	42.55	68.18	47.02	94.46	228.64	182.72	57.47	82.25	151.02	58.99	41.33	58.37	111.99	188.18
2019-2021	2022	104.23	45.97	68.47	45.96	93.81	228.63	179.70	57.47	73.70	136.68	58.55	39.84	61.43	100.54	185.65
2020-2021	2022	104.87	41.76	70.52	49.22	92.71	228.64	171.75	57.47	72.92	119.21	59.53	42.95	64.37	82.64	179.99
2022	2023	55.47	47.34	19.81	16.38	64.37	87.18	132.76	26.60	20.13	21.54	19.81	22.57	26.25	21.22	114.94
2016-2022	2023	19.50	28.22	27.40	16.06	48.96	87.20	30.81	26.60	17.03	20.45	19.76	22.99	22.44	21.52	28.89
2018-2022	2023	18.83	27.56	24.41	16.14	50.72	87.20	23.51	26.60	17.98	22.74	22.64	23.00	21.56	23.76	26.73
2019-2022	2023	21.00	30.08	22.01	16.22	52.70	87.20	20.92	26.60	19.30	22.07	22.03	22.99	20.35	22.98	26.04
2020-2022	2023	20.17	26.13	22.53	16.36	52.31	87.22	46.14	26.60	21.36	23.19	18.29	23.74	23.10	19.70	46.31
2021-2022	2023	18.22	27.25	21.43	16.16	57.35	87.21	68.34	26.60	20.46	23.19	17.84	23.25	24.09	20.02	63.74

Table 4.2: MAE values corresponding to Table 3.6.

Acknowledgements

The authors would like to thank Kun Zhu, Asbjørn Meinhardt, Jacob Hald Hansen, Julie Harbo Arleth and other colleagues from Danske Commodities for valuable discussions and insights during the development of this work. This research was supported by Innovation Fund Denmark (Ref.no. 2052-00039B) and Danske Commodities. The authors thank Danske Commodities for providing access to the data used in this study.

References

[1] Martin T Barlow. A diffusion model for electricity prices. *Mathematical finance*, 12 (4):287–298, 2002.

[2] Markus Burger, Bernhard Klar, Alfred Müller, and Gero Schindlmayr. A spot market model for pricing derivatives in electricity markets. *Quantitative finance*, 4(1):109, 2003.

[3] René Carmona and Michael Coulon. A survey of commodity markets and structural models for electricity prices. quantitative energy finance: modeling, pricing, and hedging in energy and commodity markets, 2014.

[4] René Carmona, Michael Coulon, and Daniel Schwarz. Electricity price modeling and asset valuation: a multi-fuel structural approach. *Mathematics and Financial Economics*, 7:167–202, 2013.

[5] Álvaro Cartea and Pablo Villaplana. Spot price modeling and the valuation of electricity forward contracts: The role of demand and capacity. *Journal of Banking & Finance*, 32(12):2502–2519, 2008.

[6] Sema Coskun and Ralf Korn. Modeling the intraday electricity demand in germany. *Mathematical modeling, simulation and optimization for power engineering and management*, pages 3–23, 2021.

[7] Gauthier de Maere d'Aertrycke and Yves Smeers. The valuation of power futures based on optimal dispatch. *The Journal of Energy Markets*, 3(3):27, 2010.

[8] Thomas Deschatre, Olivier Féron, and Pierre Gruet. A survey of electricity spot and futures price models for risk management applications. *Energy Economics*, 102:105504, 2021.

[9] DigiDates. German public holidays — digidates, 2024. URL https://digidates.de/en/germanpublicholidays. Accessed: 31 March 2025.

[10] Roland Füss, Steffen Mahringer, and Marcel Prokopczuk. Electricity derivatives pricing with forward-looking information. *Journal of Economic dynamics and control*, 58:34–57, 2015.

[11] Ria Grindel, Wieger Hinderks, and Andreas Wagner. Application of continuous stochastic processes in energy market models. *Mathematical Modeling, Simulation and Optimization for Power Engineering and Management*, pages 25–50, 2021.

[12] Sam Howison and Michael Coulon. Stochastic behaviour of the electricity bid stack: from fundamental drivers to power prices. *The Journal of Energy Markets*, 2:29–69, 2009.

[13] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. *An introduction to statistical learning: With applications in python*. Springer Nature, 2023.

[14] Takashi Kanamura and Kazuhiko Ōhashi. A structural model for electricity prices with spikes: Measurement of spike risk and optimal policies for hydropower plant operation. *Energy economics*, 29(5):1010–1032, 2007.

[15] Daniel S Kirschen and Goran Strbac. *Fundamentals of Power System Economics*. Wiley, 2004.

[16] Fotios Petropoulos, Daniele Apiletti, Vassilios Assimakopoulos, Mohamed Zied Babai, Devon K Barrow, Souhaib Ben Taieb, Christoph Bergmeir, Ricardo J Bessa, Jakub Bijak, John E Boylan, et al. Forecasting: theory and practice. *International Journal of Forecasting*, 38(3):705–871, 2022.

[17] Craig Pirrong and Martin Jermakyan. The price of power: The valuation of power and weather derivatives. *Journal of Banking & Finance*, 32(12):2520–2529, 2008.

[18] Scikit-learn. Adaboost regressor — scikit-learn documentation, 2024. URL https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostRegressor.html. Accessed: 31 March 2025.

[19] Scikit-learn. Mlp regressor — scikit-learn documentation, 2024. URL https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html. Accessed: 31 March 2025.

[20] Steven Stoft. *Power System Economics: Designing Markets for Electricity*. Wiley-IEEE Press, 2002.

[21] Andreas Wagner. Residual demand modeling and application to electricity pricing. *The Energy Journal*, 35(2):45–74, 2014.

[22] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International journal of forecasting*, 30(4):1030–1081, 2014.