Topics in Probability Theory

Adam J.Y. Nie

February, 2021

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the Australian National University

Dedicated to my late father and grandfathers,
for being my mentors, my dearest friends and my heroes.

Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or equivalent institution, and that, to the best of my knowledge, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Chapter 2 and Chapter 3 of the thesis were completed under the supervision and guidance of my PhD advisor Prof. Boris Buchmann. Chapter 4 of the thesis documents my contributions to an ongoing collaboration between Daning Bi, Dr. Xiao Han, Dr. Yanrong Yang and myself. All expositions and results in the thesis are my own work, except for where explicitly stated otherwise.

Adam Junyu Nie
February, 2021

Acknowledgements

As a PhD student at the Australian National University I was privileged with the opportunity to explore many fascinating areas of modern mathematics and statistics. These explorations and the resulting completion of this thesis would not have been possible without the help and support of my advisors, colleges, friends and family. I owe a great debt of gratitude to my advisor Prof. Boris Buchmann for inspiring me with his unrivalled passion towards his craft and his fine tastes in mathematics. On top of the vast amount of knowledge he passed onto me, often in his uniquely enigmatic style, he has taught me through his action the importance of taking the correct attitude towards research and teaching. Boris has made my experience as a PhD student pleasant and memorable, often through the many amusing stories he told over dinners, fueled by his vast knowledge in history and a few glasses of wine.

Another major influence on how I developed as a PhD student is my advisor Prof. Ross Maller, who has been incredibly generous in his support, both academically and in life. I have enjoyed his presence as a mentor and a friend, as well as our weekly dinners and bi-daily coffee breaks, which Ross organized with unfailing regularity. Ross has always been a role model for me to look up to, and has continued to inspire me through his knowledge, enthusiasm, his genuine care for people around him, and a healthy dose of his signature cynicism. I am also grateful towards Ross for making me feel included, both in the school and in the wider academic community.

I would like to thank Dr. Yanrong Yang, who introduced me to the wonders and horrors of random matrix theory. Yanrong had been patient and supportive since the beginning of our collaboration and has spent considerable effort reading through parts of my thesis. I am grateful for her trust in me to take on a topic I knew next to nothing about, and her continuing support along the way. I am also indebted to our collaborator Dr. Xiao Han, who has patiently answered my questions in great detail. In fact, I suspect that a large portion of my current knowledge on random matrices can be traced back, one way or another, to one of our many lengthy conversations. Without his help, my journey into random matrix theory would have been much less pleasant.

As a PhD student under Ross and Boris I was privileged to have met many visitors from different parts of the world. I am thankful towards Prof. Sid Resnick for his advices and encouragement, as well as all the social commentaries he brought during his annual visits. I have enjoyed talking to Prof. David M. Mason about mountain hiking and the ambiance of a certain Italian restaurant. I would also like to thank Prof. Alex Szimayer, Prof. Ana Ferreira, Prof. Jianfeng Yao, Prof. Eckhard Platen, Prof. Uwe Einmahl and Dr. Soudabeh Shemehsavar for their comments and discussions during their visits.

I am very thankful towards Prof. Ben Goldys, Dr. Pierre Portal and Prof. William Dunsmuir for their advices and encouragement during the last five years. I am also grateful towards my colleges as well as the academic and administrative staffs of ANU. I would like to thank Dr. Yuguang Ipsen, Dr. Fei Huang, Dr. Kevin Lu, Dr. Tanya Schindler and Prof. Dale Roberts for their help and support. I would like to thank Prof. Tim Higgins especially for being an awesome PhD coordinator and tirelessly answering all of our questions. I would like to thank my friends and colleagues Yang Yang, Chen Tang, Daning Bi, Dr. Le Chang, Dr. Jeremy Wang and Dr. Yuan Gao for making my experience as a student at ANU throughly enjoyable. I am also very thankful towards Yang Yang and Chen Tang for being the best officemates I could have asked for.

None of my achievements would be possible without the unconditional love and support from my parents. Moving to a new city and living on my own has definitely made me more appreciative of their constant support. Finally, no words could fully capture my gratitude towards my girlfriend Lu Wang. She has brought me joy, comforted me when I was hurt, and patiently supported me all the way to the completion of my PhD. I thank her for being in my life and accompanying me through the highs and lows of the last few years. Lastly, an honorable mention goes to the latest addition to our household - KuChar, who has decided to take over my pillow for its new bed. KuChar has expedited the completion of this thesis significantly by sitting on my face and hitting me with its tail until I surrender the pillow, get up and continue writing this thesis.

Abstract

This thesis focuses on the theoretical aspects of several classes of continuous time, continuous state space stochastic processes. Chapter 2 and Chapter 3 consider Lévy processes and Lévy driven stochastic functional differential equations. Chapter 4 studies a high dimensional factor model and the eigenvalues of the sample auto-covariance matrix.

In Chapter 2 we extend the construction of the so-called weak subordination of multivariate Lévy processes in [49] to an infinite dimensional setting. More specifically, we give sufficient conditions for the existence of the weak subordination between a Lévy processes defined on an arbitrary Hilbert space and a sequence-valued Lévy subordinator defined on suitable Banach spaces. As by-products of our main results, we obtain a characterization of subordinators on certain sequence spaces, as well as a characterization of Lévy measures on direct sums of Banach spaces with different geometries.

Chapter 3 focuses on a Lévy driven stochastic delayed differential equation (SDDE) which arises as a continuous time analogue to the discrete time GARCH process. The SDDE was obtained in the recent works [65, 66, 160] as a weak limit in the Skorokhod topology of a sequence of suitably scaled discrete GARCH processes, as the time between observations tends to zero. In our work, we give sufficient conditions for the existence, uniqueness and regularity of the solution to the SDDE. We show that the SDDE can be reformulated as a stochastic functional differential equation and investigate the behaviour of its sample paths. The mean process and the covariance process of the solution are computed and are shown to exhibit similar behaviours to the discrete GARCH process.

Chapter 4 focuses on a high dimensional factor model proposed by [104, 105] and subsequently studied in [111] to model time series data. We investigate the asymptotic distribution of leading eigenvalues of the (product symmetrized) sample auto-covariance matrix under a high dimensional regime where the dimension and sample size tend to infinity simultaneously. Utilizing some new developments [52, 110, 111] in high dimensional random matrix theory, we obtain a central limit theorem for the empirical eigenvalues after suitable centering and scaling. It is shown that the correct centerings for the eigenvalues are in general not equal to the population eigenvalues.

Contents

Chapter 1

Introduction

This thesis consists of three chapters, each of which focuses on establishing the theoretical foundations of a class of stochastic processes appearing in the recent literature. The stochastic processes considered in this thesis share a common origin in the field of finance and econometrics, and were originally proposed to model financial time series such as asset returns. The goal of our work is to investigate the theoretical properties of these processes, thereby gaining a better understanding of the behaviours of these processes and how they should be used in practical situations.

The common theme between Chapter 2 and Chapter 3 is the study of Lévy processes, which are introduced in Section 1.1. Chapter 2 focuses on the concept of subordination, which can be interpreted intuitively as a stochastic time change of a Lévy process. The main goal of the chapter is to extend the construction of the weak subordination proposed in [49] to an infinite dimensional setting. In Chapter 3 we consider a Lévy driven stochastic delayed differential equation obtain in [160] as a continuous time limit of the GARCH process. We formulate conditions for the existence, uniqueness and regularity of the solution, and show that the solution has a covariance structure similar to that of the GARCH process. In Chapter 4 we study the asymptotic properties of a high dimensional factor model through the perspective of random matrix theory. Under a setting where the dimension and the sample size diverge simultaneously, we establish the asymptotic normality of the spiked eigenvalues of the product-symmetrized sample auto-covariance matrix.

We now give an overview of each topic, focusing mainly on the backgrounds and the motivations of our work. More details on the technical aspects of these topics can be found in the introduction section of each chapter. After a brief introduction to Lévy processes in Section 1.1, we introduce the concept of subordination in Section 1.2. In Section 1.3 we give an overview of the history and recent development of continuous time approximations to the GARCH process. Section 1.4 introduces high dimensional factor models and their connections to high dimensional random matrix theory.

1.1 Lévy Processes

Lévy processes, named after the French mathematician Paul Lévy, are often described as a continuous time analogue to discrete time random walks. More precisely, Lévy processes are defined as continuous time stochastic processes with independent, stationary increments and stochastic continuity. A formal definition of Lévy processes suitable to the context of our work as well as their basic properties will be given in Section 2.2 after we introduce the necessary notations. Common examples of Lévy processes include Brownian motions, compound Poisson processes, Gamma processes and stable processes.

Classical treatises of Lévy processes defined on Euclidean spaces include [6, 145, 152]. Observe that the definition of Lévy processes only requires the formation of increments and some notion of continuity. Therefore, in addition to real-valued and vector-valued Lévy processes, Lévy processes can be generalized and defined in very general topological vector spaces. Towards this direction, we will mainly refer to [3, 5, 8, 78, 114, 134] for treatments of Lévy processes in Hilbert and Banach spaces, which is the setting for our current work. For even more general settings, we refer to [112, 113] for the treatment of Lévy processes on compact Lie groups, [73] for Lévy processes defined on the dual of a nuclear space and [8] for cylindrical Lévy processes on the dual of a Banach space.

Lévy processes and equations driven by Lévy processes have been studied extensively in the recent years, often as an extension to Gaussian type models. Applications of Lévy processes appear in numerous and diverse fields, including mathematical finance [122, 41, 60, 121], financial economics [20, 100], insurance [59, 99], and physics [72, 24]. In these works, Lévy processes are used directly to model some stochastic, time varying quantity of interest, for instance stock returns and insurance claims. The choices of Lévy processes used in these models are usually tailored to the problem of interest, often by constructing a Lévy processes whose features closely resemble the stylized features of the quantities of interest. This is feasible due to the great flexibility of the class of Lévy processes.

At the same time, instead of being applied directly to model various types of data, Lévy processes are also used as driving noises for stochastic partial differential equations (SPDE). Since these SPDEs are usually formulated in general Hilbert or Banach spaces, this necessitates the study of infinite dimensional Lévy processes. Works in this direction typically consider the existence, uniqueness, spatial/temporal regularity and ergodicity of the solution to an SPDE driven by Lévy processes. Examples of such SPDEs include the Langevin equation and its solution the Ornstein-Uhlenbeck process [4, 45, 107, 148] and the Heath-Jarrow-Morton-Musiela (HJMM) equation frequently used in term structure modelling [25, 125]. The study of Lévy driven SPDEs is spanned over multiple fields of science and an extensive review of the relevant literature is beyond the scope of this thesis. We refer the interested reader to the following list of recent works in various areas related to Lévy driven SPDEs [7, 25, 44, 45, 55, 63, 107, 117, 125, 142, 144, 148].

1.2 Subordination of Lévy Processes

A commonly utilized concept in many of the aforementioned works is the subordination of Lévy processes. Let $X = (X_1(t), \ldots, X_n(t))_{t\geq 0}$ be a Lévy process taking values in $\mathbb{R}^n$ and $T = (T_1(t), \ldots, T_n(t))_{t\geq 0}$ be a subordinator on $\mathbb{R}^n$, i.e. a Lévy process whose components are all non-decreasing. Since $T_i(t) \geq 0$ for all $i = 1, \ldots, n$ and $t \geq 0$, we can form a path-wise composition between X and T and define a process $X \circ T$ by setting

$$(X \circ T)(t) := \Big(X_1(T_1(t)), \ldots, X_n(T_n(t))\Big), \quad t \geq 0. \tag{1.2.1}$$

The process $X \circ T$ is known as the (strong) subordination of X and T. Here, the subordinator T represents a stochastic flow of time, thus the subordination of Lévy processes is commonly referred to as a stochastic time change. Here for illustration purposes we confine our discussion to the finite dimensional case; the infinite dimensional setting, which is the focus of our work, will be discussed in more detail in Chapter 2.

Subordination of Lévy processes on Euclidean spaces has been used extensively in mathematical finance to model stock returns. In this context, typically the Lévy process X models the prices/returns of n assets and the subordinator T models a random time change applied to each coordinate of X. A large amount of research was inspired by the pioneering work [122] who constructed the (univariate) variance-gamma (VG) process as a standard Wiener process subordinated by a gamma subordinator. The subordinator acts as a time change from real time to “business time”, which is a measure of the volumes of trading and flow of information. The “business time” is inherently stochastic and is a more appropriate way to conceptualize the trading of assets. Since its inception, the VG model has been adopted in the industry by many financial institutions as an alternative to the more traditional Black-Scholes model with geometric Brownian motion. The success of VG type models is largely due to the fact that the VG model offers a better fit to financial data and can capture many stylized features of the data, see [119, 123, 129].

In order to model multiple assets simultaneously and capture the correlation between them, various attempts had been made to construct a multivariate version of the VG model, see for instance [120, 121, 153]; see also [47] for some overarching theory on this subject. We remark that generalizing the VG model to a multivariate setting is a deceptively difficult task, mainly due to the theory of multivariate subordination being much more complex and involved than the univariate case. More specifically, in practice it is desirable to stay within the framework of Lévy processes when constructing the subordinated process $X \circ T$. That is, we would like the process $X \circ T$ to be a Lévy process itself due to all the nice and convenient properties Lévy processes have. This requirement is trivial in the univariate case where $X \circ T$ is guaranteed to be a Lévy process for arbitrary Lévy process X and subordinator T. In the multivariate setting however, the process $X \circ T$ is in general not a Lévy process. Indeed, for an immediate counter-example, take $X = (X_1, X_2)$ where $X_1 = X_2$ are the same Wiener processes and set $T_1(t) = t$ and $T_2(t) = 2t$ which are deterministic. Simple computations show that $X \circ T$ does not have independent increments and hence cannot be a Lévy process.

To the best extents of our knowledge, it is unclear if it is possible to formulate a set of necessary and sufficient conditions on X and T to ensure $X \circ T$ is again a Lévy process. Under the overarching assumption that T and X are independent, the literature mainly focuses on two types of sufficient conditions on X and T under which $X \circ T$ remains a Lévy process. In the first case, it can be shown that when the subordinator T is univariate, i.e. $T_1, \ldots, T_n$ are indistinguishable processes (where indistinguishability is as defined in page 3 of [145]), the process $X \circ T$ is always a Lévy process for any choice of the Lévy process X. This result is well-known (see for instance [152]) and we will refer to this setting as univariate subordination. On the other hand, [21] showed that if the Lévy process X has independent components, i.e. $X_1, \ldots, X_n$ are independent Lévy processes, then $X \circ T$ is a Lévy process for any choice of (multivariate) subordinator T. We will refer to this setting as multivariate (strong) subordination.

Returning briefly to our earlier discussions on models, we remark that most attempts to construct a multivariate version of the VG model utilize some mixture or superposition of the two types of (strong) subordination described above in order to stay within the framework of Lévy processes. However, these two types of sufficient condition are quite restrictive in practice. We observe that in the univariate setting, whenever the subordinator T has a jump, all coordinates of $X \circ T$ will have a jump simultaneously. This feature is clearly restrictive from a practical perspective, as the prices of multiple assets do not necessarily change simultaneously. On the other hand, in the multivariate setting, since the subordinators $T_1, \ldots, T_n$ can have jumps at distinct times, so can the coordinates of the process $X \circ T$. This is an obvious improvement upon the univariate case, however, multivariate subordination requires the independence of the coordinates of X, which is again restrictive in practice since asset prices are usually correlated.

These restrictions are fundamentally due to the fact that the (strong) subordination $X \circ T$ is not necessarily a Lévy process. To overcome these restrictions, [49] introduced another type of operation $X \odot T$ between X and T, called the weak subordination, which always produces a Lévy process for arbitrary choices of X and T. The formal definition of $X \odot T$, along with some intuitive constructions, will be given in Chapter 2 after the necessary notations are introduced. Here we motivate the weak subordination by observing that $X \odot T$ is a direct generalization of $X \circ T$ in the following sense.

Whenever X and T satisfy one of the two sufficient conditions discussed above, in which case $X \circ T$ is a Lévy process, the weak subordination $X \odot T$ can be shown to be equal in distribution to $X \circ T$. On the other hand, if X and T do not satisfy these conditions, we remark that in general there may be no Lévy process whose distribution matches that of $X \circ T$ (see Proposition 1.3.6 and Proposition 2.3.29 of [119] and the discussion therein). Nevertheless, in this case the weak subordination $X \odot T$ is still a Lévy process, but relates to $X \circ T$ in a more complex and subtle way.

Using weak subordination, [49] constructed a new multivariate version of the VG process, which the authors named weak variance generalised gamma convolutions (WVGG). The self-decomposability of the WVGG class, which is a question of theoretical importance, is investigated in a more recent work [50]. The authors also constructed a generalization of the variance-α-gamma model of [153], which the authors named the weak variance-α-gamma model (WVαG). The WVαG model is calibrated to real data and applied to option pricing in the recent works [48], [128] and [129].

The main focus of our work is to extend the construction of the weak subordination to an infinite dimensional setting. In particular we will show that for an arbitrary Lévy process on a separable Hilbert space and a uniformly bounded sequence of real subordinators $T = (T_1, T_2, \ldots)$, the weak subordination $X \odot T$ always exists as a Lévy process on a suitable Banach space. As a by-product, we develop the theory of Lévy measures and Lévy subordinators in some non-standard settings.

The motivations behind our work are twofold. On one hand, we remark again that the technique of (univariate) subordination has been utilized in the SPDE literature to define Lévy noises in function spaces, see [45, 107] and the references therein. This approach allows one to establish controls on the constructed noise in terms of the properties of X and T. However, to the best of our knowledge, all existing literature considers cylindrical Wiener processes on a Hilbert space H, which by definition have independent coordinates. The existing literature, apart from [140] which we will discuss again in Chapter 2, also only considers univariate subordinators T, which implies that all coordinates of the constructed noise of $X \circ T$ will have simultaneous jumps. From this perspective, it is natural to extend the theory of multivariate and weak subordination into an infinite dimensional setting as a method to construct Lévy noises with more flexible dependence/jump structure, while maintaining explicit control over the processes constructed.

On the other hand, our work is also motivated by theoretical curiosity. The proofs behind the constructions of the weak subordination in [49] rely heavily on the finite dimensional setting of the problem. In particular, many of their arguments rely on key properties of Euclidean space such as the equivalence of all norms on $\mathbb{R}^d$, the compactness of the unit ball and the trivial observation that any finite sequence of real numbers can be ordered. It is therefore not clear from [49] if the existence of weak subordination is an inherent feature of Lévy processes (and not just an artifact of working in Euclidean spaces). Our work answers this question in the positive by showing that the construction of weak subordination is possible in a much more general setting.

1.3 Continuous Time GARCH processes

An important aspect of statistical analysis in finance and financial econometrics is finding models that capture the so-called stylized features of these types of data. Many financial time series such as asset prices, exchange rates and various macroeconomic series are heteroscedastic, i.e. the volatility of these series tends to be time varying. For instance, it is known that asset returns and their volatility series are heavily interdependent - a large fluctuation in the asset price typically causes a large fluctuation in the volatility, which persists for a period of time before reverting to a baseline level, see [57, 81]. These features fit our intuitions about the financial market - a large fluctuation in the price of an asset leads to investor uncertainty, which drives up the volatility of the price. The volatility of these series also tends to exhibit long-range dependence, which often manifests in the heavy-tailedness of the series, another feature commonly observed in financial time series.

The ARCH (autoregressive conditionally heteroscedastic) and GARCH (generalized ARCH) processes, introduced by [68] and [37], are designed specifically to capture some of these stylized features. In the GARCH model, the asset return process is temporally uncorrelated but not independent, and the volatility of asset returns is modeled directly by modelling the the conditional variance process of asset returns given past information. The conditional variance process follows an auto-regressive moving-average type of structure, and is driven by the same noise as the return process. For a review of the properties and stationarity of GARCH processes, we refer to [40] and [116].

While financial data are often observed at discrete times, the modelling of these data is sometimes carried out in continuous time for practical concerns as well as theoretical convenience. In practice, financial time series are often observed at irregular frequencies, often tied to trading hours of stock exchanges or days of the week. Modelling irregularly spaced data directly is often challenging, see [69, 77, 127] for examples of discrete time GARCH models adapted to irregularly spaced data; it is often fruitful to consider continuous time models instead. The increasing popularity of high frequency data in recent years also motivated the developments of continuous time models such as [109, 159]. On the other hand, from a theoretical perspective, discrete time models are expressed via recurrence equations while continuous time models are often written as differential equations. Consequently, the asymptotic theory for discrete time models is often more involved and less explicit than for the continuous time models.

The preceding paragraph may seem to suggest that there is a large discrepancy between continuous and discrete time modelling; however, the two paradigms are often quite compatible with each other. In fact, as the time between observations tends to zero, many discrete time models (after appropriate scaling) can be shown to converge in some sense to continuous time models, see [102, 103, 118, 158, 160] and the references therein. Hence continuous time models can be used as approximations to discrete time models when the time between observations is relatively small. On the other hand, discrete time models can obviously be obtained from continuous time ones by simply sampling at discrete times, in which case the former can be interpreted as approximations to the latter.

As an interesting and perhaps informal side note, we remark that while financial data is observed at discrete times, whether the data is *inherently* discrete or continuous is a separate and more philosophical question. Some argue that despite having discrete time observations, the evolution of stock prices actually happens continuously in time. In this case, the discrete time observations are simply samples from this underlying, unobserved continuous time process. Others oppose this view and argue that the listed price of a stock is simply the price at which it is being traded. Therefore unless a trade happens, the price of a stock should remain constant, implying that continuous time models especially diffusions are technically convenient approximations to the truth. A debate on which view is more sensible is beyond the scope of the current work. We do remark however that under either view, the study of continuous time models is needed.

Motivated by the above observations, various attempts have been made to construct a continuous time version of the GARCH process. Clearly such a continuous time extension should capture some aspects of the discrete time GARCH process, such as the recurrence relationship of the volatility process. Many approaches start with discrete time GARCH processes and allow the time between observations to tend to zero, and with proper scaling, a continuous limit may be obtained. The first notable attempt in this direction can be found in [132], where the author obtained a diffusion in the limit. The nature of this construction is very similar to taking a scaling limit of discrete time random walks in the Skorokhod topology to obtain a Brownian motion, see [35].

Although the diffusion limit of [132] is intuitive to understand, the resulting diffusion process loses many characteristic properties of the GARCH process. In particular, the diffusion limit is driven by two independent Wiener processes whereas the GARCH process is driven by a single sequence of noise. The feedback and interactions between the price and volatility processes is therefore not captured by the diffusion limit. This discrepancy is investigated further in [166], which established that the statistical inferences for the discrete time GARCH process and the diffusion limit are not equivalent asymptotically (in the sense of Le Cam's deficiency distance, see [46]). An attempt to resolve this issue can be found in [58], where the author modified the construction of the diffusion limit in [132] to obtain a limit which is driven by a single Brownian motion. However, the volatility process obtained in [58] is deterministic, which is clearly undesirable. Despite its limitations, the diffusion limits of [132] have been studied and applied in numerous works. We refer to [64, 95] and the references within.

A very different approach to constructing a continuous time GARCH process was put forward by [100]. In the GARCH process, the price process is driven by a sequence of white noise, and the volatility process is driven by the square of that sequence. In [100] the authors replaced these sequences of driving noises with increments of a Lévy process and the squares of those increments respectively, obtaining a pair of continuous time processes named the Continuous Time GARCH (COGARCH) process. In doing so, the price process and the volatility process are by design driven by the same Lévy process, circumventing the main issues of the diffusion limit in [132]. The volatility process in the COGARCH model is an example of the so-called generalized Ornstein-Uhlenbeck process, which is the subject of many recent works, see [27, 28, 29, 115].

The COGARCH process retains a lot of the unique features of the GARCH process and is thus considered a more useful extension of the GARCH process than the diffusion limit of [132]. For properties and stylized features of the COGARCH process, we refer to the surveys [102, 116] and the references therein. It is shown in [96] that similar to the construction of the diffusion limit in [132], the COGARCH process can also be obtained as a weak limit of a sequence of GARCH(1,1) processes. Unlike [132] which scaled the noise of the discrete GARCH process to obtain a diffusion, the construction in [96] "randomly thins" the noise to obtain the driving noise in the COGARCH process. The authors of [96] also argued heuristically that the diffusion limit and the COGARCH limit are the only possible limits of sequences of GARCH(1,1) processes. Parallel to the results of [96], a different construction of the COGARCH process from discrete time GARCH processes was considered in [124]. The authors obtain the COGARCH process as a limit of a sequence of irregularly spaced GARCH processes using the so-called first jump approximation. As a by-product [124] constructed a pseudo-maximum-likelihood estimator for parameters in the COGARCH process which can be applied to irregularly spaced data. The parameter estimation of the COGARCH process was also considered in [82, 124, 131] and the COGARCH process was applied in practical situations such as option pricing in [101, 102]. An analogous result to [166] for the diffusion limit was obtained in [46]. Multivariate versions of the COGARCH process have been studied in for instance [155].

It is worth noting that both the diffusion limit and the COGARCH process are Markovian processes, while the original GARCH(p, q) process is only Markovian when $p = q = 1$. This suggests that the diffusion limit and the COGARCH process are rather simplistic in their serial dependence structure in comparison to the original GARCH process. The first attempt at a non-Markovian continuous time GARCH process can be found in [118]. In contrast to the diffusion limit and the COGARCH process which are obtained as limits of GARCH(1,1) processes, in [118] the order of the approximating GARCH processes changes as the time between observations decreases. In particular, for each $n \in \mathbb{N}$ and fixed $p > 0$, a GARCH$(pn + 1, 1)$ process is defined on a uniform grid of mesh size n^{-1}. As $n \to \infty$, the sequence of GARCH processes is shown to converge to the solution of a stochastic delayed differential equation (SDDE) driven by a Wiener process, which is a direct, non-Markovian generalization of the diffusion limit. We note that in [118], as the time n^{-1} between observations tends to zero, the "GARCH order" of the approximation GARCH$(pn+1,1)$ process tends to infinity while the "ARCH order" remains at one. Allowing the "ARCH order" to diverge as well turns out to be significantly more difficult, since it involves the analysis of stochastic integrals with delays. Furthermore, the idea of [118] is developed strictly in the context of the resulting diffusion limit, and consequently suffers the same issues, namely the two independent sources of noises discussed above.

A similar idea was explored in greater depth and generality in the recent PhD thesis [160] and the working papers [65, 66]. Unlike [118] which generalized only the diffusion limit to a non-Markovian setting, the results in [65, 66, 160] cover both the diffusion case and the COGARCH case. The authors considered a GARCH$(pn+1,qn+1)$ process embedded into continuous time on a grid of mesh size n^{-1}, where $p,q>0$ are fixed real numbers. As the time between observations n^{-1} tends to zero, both the ARCH and the GARCH orders diverge. The sequence of GARCH$(pn+1,qn+1)$ processes is shown to converge weakly to the solution of the pair of stochastic delayed equations

$$\begin{aligned} Y_t &= Y_0 + \int_0^t \sqrt{X_{s-}}dL_s, \\ X_t &= \theta_t + \int_{-p}^0 \int_u^{t+u} X_s ds\mu(du) + \int_{-q}^0 \int_{u+}^{t+u} X_{s-}d[L,L]_s\nu(du), \end{aligned} \tag{1.3.1}$$

where Y and X are the return and variance processes respectively. The solution to this equation is named the continuous time GARCH process with delays $p,q\geq 0$, or the CDGARCH(p,q) process for short. Here μ and ν are Borel measures capturing the effects of higher order lags, and θ is a semimartingale representing some form of baseline for the volatility process, which we will make precise in our work. The driving noise L is a semimartingale and $[L,L]$ is the quadratic variation process of L.

Depending on how the driving noise of the discrete approximating sequence is constructed, the limiting driving noise L could either be a Brownian motion, in which case the CDGARCH process generalizes the construction of [118], or a Lévy process, in which case the CDGARCH process is a generalization of the COGARCH process. In fact, the limit in [118] is a special case of the CDGARCH(p,q) process when $q=0$, and the COGARCH process of [100] is a special case of the CDGARCH(p,q) process when $p=q=0$.

We remark that the main focus of [160] is on establishing the convergence of GARCH processes to a continuous time limit, the CDGARCH process. In particular [160] did not further explore the behaviours of this limit and how it compares to discrete time GARCH processes or the COGARCH process. In our work we take equation (1.3.1) as a starting point and investigate whether it is a good continuous time analogue to the GARCH process. We pose conditions for the existence, uniqueness and regularity of the solutions to (1.3.1). We study the second order behaviour of the solution and show its resemblance to that of the GARCH process.

1.4 High Dimensional Factor Models

The analysis of multivariate time series, especially ones of high dimensionality, is a topic of increasing importance in the modern age. Besides its theoretical importance as a classic topic in the theory of statistics and probability, multivariate time series has been successfully utilized in many empirical fields including finance, economics, medical science and many others. The wide study of high dimensional time series is very easily understood; indeed, many forms of data in real life are time varying and exhibit autocorrelation from which useful information can be extracted. Moreover, the increasing abundance of computing power and capacity in the modern age has made the analysis of large datasets possible and routine. Many high dimensional stochastic processes exhibit behaviours fundamentally different from what classic asymptotic theory predicts, resulting in the rise of many new fields and bodies of theory devoted to the study of these behaviours. A complete survey of these different fields is beyond the scope of our work, we refer to [90] and [70] for surveys on two such fields for the interested reader.

A prominent feature of high dimensional datasets and stochastic processes is the so-called 'curse of dimensionality'. One way this 'curse' manifests itself is through the fact that the number of parameters in a model often explodes quickly as the dimension of the model increases. This directly affects how multivariate time series are analyzed and used in practice, as we will now explain. Many conventional time series models (e.g. vector ARMA models) have been thoroughly analyzed through Fourier methods in the time/frequency domain, and the asymptotic theories of these models were known for decades, see for instance [43, 80, 154] for some classical treatises of these topics. However, in practice, these are rarely used to model datasets of high dimensions since the number of parameters of interest very quickly becomes intractable. For instance, even for the simple case of a vector autoregressive process of order one, the number of parameters increases at a rate of p^2. In practice, when dealing with high dimensional time series, most methods rely on either regularization to control the number of non-zero parameters, or some form of dimension-reduction technique such as principal component analysis (PCA) to reduce the complexity of the model before analysis is carried out. We refer to [90] and [70] for surveys on recent developments on these approaches.

Our work is motivated by one such example of dimension-reduction technique applicable to high dimensional time series - the use of factor models. Factor models have enjoyed much recent success in various empirical fields, especially finance, economics and econometrics. The literature is vast and we refer to [34, 74, 75, 76, 157, 156] and the references therein for some notable examples of factor models used to analyze time series data in these fields. These models exploit the idea that for many types of multivariate data, a large number of variables are in fact driven by a comparatively small number of common factors. Most models allow each variable to have an idiosyncratic part as well, but assume that the common factors account for most of the information in the model. By identifying these factors (which can often be modeled using more traditional methods) as well as the relationship between the original variables and the common factors, we are effectively reducing the dimension of the problem.

In addition to being widely applied in empirical fields, the theoretical aspect of these models, especially the large sample asymptotic theory, has attracted significant attention as well. The asymptotic theory surrounding factor models as a dimension reduction technique was first explored more than half a century ago. Early works such as [2, 42, 137] build on the assumption that the dimension p of the time series is large but fixed, and developed the asymptotic theory as the sample size T tends to infinity. Since then, the literature on factor models has diverged and branched into many interesting directions. Models of varying complexity under various settings have been proposed, and corresponding asymptotic theories were established under ever more general conditions. We refer to [156] and the references therein for a survey on some recent variants of factor models that attracted the attention of the statistical and mathematical community.

One recent direction in factor modeling that has attracted significant attention from econometricians, statisticians and probabilists is the so called high-dimensional setting where both the dimension p and the sample size T tend to infinity simultaneously. This is certainly not an unreasonable assumption, as for many datasets encountered in real life, the dimension p can be much bigger than T. As we will illustrate in Chapter 4, the theory of factor models under this setting is significantly different to the traditional setting where p is fixed. The asymptotic theory of factor models under the high dimensional setting is considered in a number of recent works including [10, 11, 12, 104, 105, 110, 111, 133]. Typical goals of these works include determining the correct number of factors, estimating the factor loading space and making predictions. Several of these results are directly related to our current work, we will give a more detailed survey in Chapter 4.

Amongst these works, we note that [10, 11, 12] focus on estimating the factors and factor loadings directly, while [10, 11, 12, 104, 105, 110, 111, 133] study the factor structure through the sample auto-covariance matrices of the time series. The latter idea is based on the observation that when strong serial correlation is exhibited by the data, a substantial amount of information is contained in the eigenvalues and eigenvectors of the matrix $M := \sum_{\tau=1}^{\tau_0} \Sigma_\tau \Sigma_\tau^\top$, where Σ_τ is the lag-τ (population) auto-covariance matrix of the time series and τ_0 is some chosen constant. Most of these works rely on estimating the eigenvalues (μ_i) of the matrix M using the eigenvalues (λ_i) of some estimate $\widehat{M}$ of M. This naturally brings the discussion to the corresponding asymptotic theory of empirical eigenvalues (λ_i) of M, which is the main focus of our current work.

Towards this direction, we note that [104] established the asymptotic rate of $|\lambda_i - \mu_i|$ under the assumption that each μ_i diverges at a rate close to p. A more recent work [111] considers the same model under the assumption that all (μ_i) are finite. Using techniques from random matrix theory, the authors in [111] established the phase transition behaviour of (λ_i) and identified their asymptotic locations as the sample size T tends to infinity. A more detailed discussion of these results can be found in Chapter 4.

To the best extent of our knowledge, the existing literature concentrates mainly on obtaining the asymptotic limits of the eigenvalues (λ_i) and very little is known about the asymptotic distributions of (λ_i). Our work contributes to the literature by initiating some studies along this direction. More specifically, we establish the asymptotic normality of spiked eigenvalues of M under quite general conditions. Similar to the settings of [104] we assume the spiked eigenvalues (μ_i) tend to infinity, but do not impose any conditions on their speed as [104] does. We show that under this more general setting, although the spiked empirical eigenvalues (λ_i) are close to their population counterparts (μ_i) asymptotically, the difference $\lambda_i - \mu_i$ typically does not decay fast enough to obtain a central limit theorem. Using recent techniques in random matrix theory we construct a more accurate centering for λ_i in order to obtain its limiting asymptotic distribution.

Chapter 2

Weak Subordination of Infinite Dimensional Lévy Processes

2.1 Introduction

Our exposition in Section 1.2 so far only focused on the strong subordination of finite dimensional Lévy processes. We will now discuss two important topics to set context to our work - the subordination of infinite dimensional Lévy processes and the weak subordination of Lévy processes as constructed in [49]. In 2.1.1 we give an overview of the theory of subordination of infinite dimensional Lévy processes, drawing connections to our discussions in Section 1.2. We will then give an intuitive construction of the weak subordination in Section 2.1.2 that illustrates the motivations behind it. The content of the rest of the chapter will be summarized in Section 2.1.3.

2.1.1 Subordination of Infinite Dimensional Lévy Processes

We recall from Section 1.2 that by univariate subordination of Lévy processes we refer to the situation where the Lévy process X can be arbitrary, but the subordinator T is required to be univariate, that is, all coordinates of T are indistinguishable processes, where indistinguishability of stochastic processes is defined on for instance page 3 of [145].

Univariate subordination can be easily extended to the case where X lives on a separable Banach space, see for example [139], where the process $X \circ T$ is defined by specifying its characteristic triplet. The univariate subordination is sometimes used in the SPDE literature as a method to construct Lévy noises with certain properties. We remark that there are other ways of defining Lévy type noises in infinite dimensional spaces, for instance see [8] for a direct construction of a cylindrical Lévy process on Banach spaces. One convenient feature of using subordination to define Lévy processes is that we often have explicit control on the process $X \circ T$ using the characteristic triplet of X and T. As we will see, this is a feature shared by the weak subordination as well.

To give an example, we first recall that a symmetric α-stable Lévy process on $\mathbb{R}^n$ can be constructed as a standard Wiener process on $\mathbb{R}^n$ subordinated by a univariate $\alpha/2$-stable subordinator. This observation is used in [107] to define an infinite dimensional α-stable white noise, using a cylindrical Wiener process X on a separable Hilbert space subordinated by a univariate $\alpha/2$-stable subordinator T. The authors studied the Langevin equation driven by this process and properties of the semigroup generated by its solutions.

Similarly, [45] defined Lévy white noise as a cylindrical Wiener process subordinated by an arbitrary univariate subordinator T. The authors considered the Langevin equation driven by this noise, and characterised spatial and temporal regularities of the solution in terms of the properties of the subordinator T. In particular, stochastic integrals with respect to the process $X \circ T$ can be controlled using the integrability of the Lévy measure of T near the origin. Similar usages of subordination can be found in [164].

On the other hand, the extension of multivariate subordination as constructed by [21] to an infinite dimensional setting is non-trivial. Indeed, it is not immediately clear how the definition (1.2.1) should be extended to a an infinite dimensional setting or if the process $X \circ T$ lives on the same space X lives on. To our best knowledge, multivariate subordination has not been studied in an infinite dimensional setting apart from [140]. The authors of [140] considered the setting where X takes values in the space of nuclear operators $L_1(H)$ on a separable Hilbert space H and the subordinator T takes values in the positive cone of the same space $L_1(H)$. The construction and results of [140] are indeed comparable to the original work [21], but the setting of [21] is perhaps rather unnatural for the following reasons.

First of all, if the motivation behind extending multivariate subordination to function spaces is to construct Lévy noises suitable to the study of SPDEs, then it is most natural to consider cases where X lives on some Hilbert or Banach space. If on the other hand we wish to construct an infinite dimensional model for the prices of a large number of assets, then the most natural setting is to define X on sequence spaces. The space of trace-class operators $L_1(H)$ used in [140] does not fit under either of the cases. Secondly, the assumption [140] that $T(t)$ is a positive trace-class operator on H is very difficult to motivate for the following reasons. Indeed, in the finite dimensional case, the subordinator T can be understood as a sequence of stochastic time changes for the coordinates of X. We argue that it is natural to keep this interpretation when generalizing the construction, in which case T should be defined as a sequence of subordinators.

As a final side note, we remark that there is a deep and fascinating connection between certain geometric aspects of Banach spaces and the behaviours of probabilistic objects defined on these spaces. Unfortunately, a comprehensive survey of the many surprising results in this direction is beyond the scope of this work. We refer the interested reader to [9, 61, 83, 84, 114] for some classical results in the theory of Banach spaces and its connections to probability theory, and to [1, 86, 106, 108] for some modern treatises on the topic. These connections are especially important to our work, since the definition and characterisation of Lévy measures on Banach spaces depend on the geometry of the Banach space itself. We will summarize some relevant results in Section 2.2.1.

2.1.2 Weak Subordination of Lévy Processes

Before giving a rigorous definition of the weak subordination, we illustrate the intuitions behind it with the following heuristic construction. We first note that any subordinator T in $\mathbb{R}^n$ can be written uniquely as $T(t) = t\eta + S(t)$, where $\eta \in \mathbb{R}^n_+$ is a deterministic vector and S is a pure-jump subordinator on $\mathbb{R}^n$. For the purposes of our work it is helpful to consider the processes $t \mapsto t\eta$ and $t \mapsto S(t)$ separately.

We discuss the $t \mapsto t\eta$ part of the subordinator T first. Suppose X is an arbitrary Lévy process in $\mathbb{R}^n$ and T is the deterministic subordinator $T(t) = t\eta$, where $\eta \in \mathbb{R}^n_+$. Then it is not difficult to show that there always exists a Lévy process Y, unique in law, satisfying $Y(t) \overset{\mathcal{D}}{=} (X \circ T)(t)$ for all $t \geq 0$. That is, even though the strong subordination $X \circ T$ is in general not a Lévy process as discussed in Section 1.2, there always exists a Lévy process Y with the same marginal distributions as $X \circ T$. It is therefore very natural to regard Y as a generalization of strong subordination $X \circ T$. Indeed, in the case where $T(t) = t\eta$, the weak subordination $X \odot T$ is defined to be this Lévy processes Y.

We now consider the pure-jump part of T. Unlike the previous case, the existence of a Lévy process Y with marginal distributions matching those of $X \circ T$ is in general not guaranteed, see Proposition 2.3.29 of [119]. The weak subordination in this case generalizes another key feature of the strong subordination, namely the distribution of jumps. We remark here that since T is a pure jump process, the strong subordination $X \circ T$ is clearly constant in between jumps of T and is hence a pure jump process as well.

The jumps of $X \circ T$ are given by $\Delta(X \circ T)(t) = X(T(t)) - X(T(t-))$, where $T(t-) := \lim_{s \uparrow t} T(s)$. Note that whenever T is univariate or X has independent components, in which case $X \circ T$ is a Lévy process, by stationarity of increments we have $\Delta(X \circ T)(t) \overset{\mathcal{D}}{=} X(\Delta T(t))$. This feature can be regarded as the "natural" behaviour of jumps when $X \circ T$ is a Lévy process. However, outside of these cases, the strong subordination $X \circ T$ is not necessarily a Lévy process and $\Delta(X \circ T)(t)$ is in general not equal to $X(\Delta T(t))$ in distribution.

The weak subordination aims to define a Lévy process that preserves a version of this feature even when $X \circ T$ is no longer a Lévy process. More specifically, the idea of weak subordination in this case is to construct a (pure jump) Lévy process Z whose jumps $\Delta Z(t)$ at time t, when conditioned on the subordinator $T(t)$, has the same distribution as the random variable $X(\Delta T(t))$. The requirement of Z being a pure jump process is very natural since $X \circ T$ is always a pure jump process itself. It is shown in [49] and [119] that this construction is always possible on some augmented probability space. See for example page 24 of [119] for a construction of Z using marked Poisson point processes. In this case, the weak subordination $X \odot T$ is defined to be such a Lévy process Z.

For the general case where $T(t) = t\eta + S$, the weak subordination $X \odot T$ is defined by combining the processes Y and Z in the previous two cases. Specifically, $X \odot T$ is defined as the unique (in distribution) Lévy process, whose distribution is equal to the convolution of the distributions of Y and Z described above. This feature is a direct generalization of univariate and multivariate subordination as well, see Proposition 4.3 of [47].

2.1.3 Overview of the Chapter

We extend the construction of [49] to the setting where X is an arbitrary Lévy process taking values in a separable Hilbert space with a chosen orthonormal basis (e_n), and T is a subordinator defined on a suitable infinite dimensional sequence space. Such an extension is nontrivial since the arguments in [49] do not generalize to non-Euclidean spaces.

Recall from Section 2.1.2 that to define the weak subordination, we need to condition on $T(t)$ and consider the random variable $X(\Delta T(t))$. This motivates a preliminary step in our analysis; in Section 2.3 we focus on the evaluation of X at a multivariate time parameter $\tau \in \mathbb{R}_+^{\mathbb{N}}$, given by $\mathbf{X}(\tau) := \sum_n \langle X(\tau_n), e_n \rangle e_n$. We show that this sum defines a random variable in H as long as τ is a uniformly bounded sequence, i.e. $\tau \in \ell_\infty^+$. We show that the distribution of the resulting random variable $\mathbf{X}(\tau)$ is infinitely divisible and specify its characteristic triplet as the limit of a certain sequence of triplets.

From these results we conclude that, in order to define the random variable $X(\Delta T(t))$ on the same space as X, the sequence $\Delta T(t)$ needs to be uniformly bounded almost surely. It is easy to show that this implies T has to be uniformly bounded as well. As a consequence, the natural choice of space to define the subordinator T on is ℓ_∞. This immediately presents a difficulty since ℓ_∞ is not a separable space and the analysis becomes much more involved. To overcome this complication, in Section 2.4 we consider ℓ_∞ as a subspace of a bigger, suitable weighted sequence space which is separable. We characterize subordinators on such a space and then establish conditions for T to concentrate on the subspace ℓ_∞. This way we avoid directly doing analysis on ℓ_∞.

Another complication we encountered in our setting is that fact that in [49], the weak subordination of X and T is in fact defined as the pair of process $(T, X \odot T)$ on $\mathbb{R}^{2n}$. Beside increased generality, a reason for this choice is that the weak subordination $X \odot T$ is only defined up to distribution. Defining the pair $(T, X \odot T)$ instead of just $X \odot T$ therefore allows us to consider the joint distribution of T and $X \odot T$ and match up the jumps of the two processes. We refer to [119] for more details.

Keeping this desirable structure in our work requires some extra efforts, since the pair $(T, X \odot T)$ lives on the direct sum of a Hilbert space and ℓ_∞, which is a very specific and non-standard setting. This issue is tackled in Section 2.5, where we characterize Lévy measures on direct sums of Banach spaces with different geometries. Finally, in Section

2.6 we give a formal definition of the weak subordination as well as some explanation on how this definition corresponds to our intuitive construction in Section 2.1.2. We then establish the existence of the weak subordination $X \odot T$, which is the main result of our work, in Theorem 2.24.

2.2 Notations and preliminaries

We first collect some preliminary results and setup the necessary notations for our expositions and proofs. For more details we refer the reader to [6, 94, 145, 152] for classic treatises of probability theory and Lévy processes, [1, 146] for topics in functional analysis and Banach spaces, and [1, 5, 85, 108, 114, 142] for topics in probability theory, Lévy processes and their connections to the theory of Banach spaces.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space where the filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual hypotheses of right continuity and completeness (see [145]). Let $(E, |\cdot|_E)$ be a separable Banach space equipped with its Borel σ-algebra $\mathcal{B}(E)$. We will usually omit the subscript and write $|\cdot|$ when the context is clear. Write E' for the topological dual of E and $\langle \cdot, \cdot \rangle$ for the duality between E and E'. When E happens to be a Hilbert space, we identify E' with E via Riesz's theorem and use the notation $\langle \cdot, \cdot \rangle$ for the inner product on E. Write $B_r(x)$ for the open ball in E with radius r and center x, and $B := B_1(0)$.

We first give a formal definition Lévy processes taking values in a Banach space E, which is the the central object of our study. Note that this definition is a straight-forward generalization of the usual definition of Lévy processes taking values in $\mathbb{R}^n$.

Definition 2.1. Let $X = (X(t))_{t\geq 0}$ be a càdlàg stochastic process adapted to $(\mathcal{F}_t)_t$ taking values in a separable Banach space E. Then X is called a Lévy process if

a) $X(0) = 0$, $\mathbb{P}$-almost surely,

b) X has independent and stationary increments, i.e. for any $0 \leq s < t$, the increment $X(t) - X(s)$ is independent from $\mathcal{F}_s$ and is equal in distribution to $X(t - s)$,

c) X is stochastically continuous (in the norm of E), i.e. $|X(t) - X(s)|_E \to 0$ in probability whenever $|t - s| \to 0$.

Write $\Delta X(t) := X(t) - X(t-)$ for the jump of X at time t, where $X(t-) := \lim_{s\uparrow t} X(s)$. Let $N : [0, \infty) \times \mathcal{B}(E \setminus \{0\}) \to L^0(\Omega, \mathbb{N})$ be a family of random measures defined by

$$N_t(dx) := \#\Big\{0 \leq s \leq t, \Delta X(s) \in dx\Big\} = \sum_{0\leq s\leq t} 1_{\Delta X(s)\in dx},$$

where $L^0(\Omega, \mathbb{N})$ is the space of random variables taking values in the natural numbers. Then $N_t(\cdot)$ is a Poisson random measure for any $t \geq 0$ and we have $N_t(A) < \infty$ a.s. for any $t \geq 0$ whenever A is bounded away from zero. Furthermore, there exists a σ-finite measure ν on $\mathcal{B}(E \setminus \{0\})$ such that for any $t \geq 0$, we have

$$\mathbb{E}[N_t(A)] = t\nu(A), \quad A \in \mathcal{B}(E \setminus \{0\}).$$

The measure ν is known as the Lévy measure of the Lévy process X.

In the finite dimensional case, Lévy measures are completely characterized by integrability conditions. It is well known that a σ-finite measure ν on $\mathcal{B}(\mathbb{R}^n \setminus \{0\})$ is the Lévy measure of some Lévy process on $\mathbb{R}^n$ if and only if

$$\int_{\mathbb{R}^n} 1 \wedge |x|^2 \nu(dx) < \infty.$$

However, in the Banach space setting, this condition is no longer necessary, nor sufficient. In fact, it is known that the above condition is necessary and sufficient for a measure on a Banach space E to be a Lévy measure, if and only if E is isometrically isomorphic to a Hilbert space (see Theorem 2.4 in Section 2.2.1), i.e. E is a Hilbert space itself.

In finite dimensions, we usually take the above integrability condition to be the definition of a Lévy measure. On the other hand, the definition of Lévy measures is not as straight-forward in the Banach space setting. In fact, the properties of Lévy measures on a Banach space E depends heavily on certain geometric aspects of E. We will introduce some elements of Banach space theory in Section 2.2.1 which allows us to give a useful definition for Lévy measures on Banach spaces and state the Lévy-Khintchine formula.

2.2.1 Lévy measures on Banach spaces

The characterisation of Lévy measures on Banach spaces is intimately connected to the geometry of the underlying space, specifically to the notions of type and cotype. Let $(\epsilon_n)_n$ be a Rademacher sequence, i.e. a sequence of i.i.d. random variables uniformly distributed on $\{-1, 1\}$. We recall from [86] that a Banach space E is said to have type $p \in [1, 2]$ iff there exists a constant $K_p \geq 0$, depending only on the choice of E, such that

$$\mathbb{E}\left|\sum_{i=1}^{n} \epsilon_i x_i\right|^p \leq K_p \sum_{i=1}^{n} |x_i|^p$$

holds for any finite sequence $x_1, \dots, x_n$ of elements in E. Similarly E is said to have cotype $q \in [2, \infty]$ if and only if there exists $C_q \geq 0$ such that

$$\mathbb{E}\left|\sum_{i=1}^{n} \epsilon_i x_i\right|^q \geq C_q \sum_{i=1}^{n} |x_i|^q$$

for any finite sequence $x_1, \dots, x_n$ in E and $n \in \mathbb{N}$.

To compare with the finite dimensional setting, we note that if x_i are vectors in $\mathbb{R}^n$, then it is easy to show that $\mathbb{E}|\sum_{i=1}^n \epsilon_i x_i|^2 = \sum_{i=1}^n |x_i|^2$, which is essentially a restatement of the parallelogram identity. The above inequalities and the notion of type and cotype can therefore be interpreted as a measure of how much (and in which direction) the parallelogram identity is distorted in a particular Banach space.

We remark that if E has type p and cotype q, by Hölder's inequality E automatically has type p' and q' for all $1 \leq p' < p$ and $q < q' \leq \infty$. In particular, by the triangle inequality we have $\max_i |x_i| \leq \mathbb{E}|\sum_i \epsilon_i x_i| \leq \sum_i |x_i|$, which implies that every Banach space has type 1 and cotype ∞. A Banach space is therefore said to have non-trivial type if it has some type $p > 1$, and non-trivial cotype if it has some cotype $q < \infty$. For a more detailed treatment on the type and cotype of Banach spaces, we refer the readers to [1].

Furthermore, the type and cotype of a Banach space are isomorphic invariants and therefore are often used to classify Banach spaces. An important result is that a Banach space E is isomorphic to a Hilbert space if and only if it is of both type 2 and cotype 2, which follows from the parallelogram identity. For some common examples, we note that every L^p space over some σ-finite measure space is of type $p \wedge 2$ and cotype $p \vee 2$. We will frequently use the following characterisation of types from Theorem 2.1 of [83]:

Theorem 2.2. *A Banach space E has type $p \in [1, 2]$ if and only if there exists a constant $K'_p \geq 0$ depending only on E such that the estimate*

$$\mathbb{E}\Big|\sum_{i=1}^n X_i\Big|^p \leq K'_p \sum_{i=1}^n \mathbb{E}|X_i|^p$$

holds for all finite sequence $X_1, \dots, X_n$ of independent E-valued random variables with mean zero and finite p-th moment.

Recall $B := B_1(0)$ denotes the unit ball on E. Finally, to define Lévy measures on E, we let $K : E \times E' \to \mathbb{C}$ be the function

$$K(x, u) = e^{i\langle x,u\rangle} - 1 - i\langle x, u\rangle 1_B(x), \quad x \in E, u \in E', \tag{2.2.1}$$

which is a generalization of the function appearing in the characteristic exponent of a Lévy process in finite dimensions. We follow Theorem 5.4.8 of [114] and define:

Definition 2.3. Let μ be a σ-finite measure μ on $\mathcal{B}(E)$ with $\mu(\{0\}) = 0$. Then μ is said to be a Lévy measure if and only if

$$\int_E |K(x, u)|\mu(dx) < \infty$$

for all $u \in E'$ and the mapping

$$u \mapsto \exp\Big(\int_E K(x, u)\mu(dx)\Big)$$

is the characteristic function of a probability measure on E.

Recall that when E is a Hilbert space (in particular any Euclidean space $\mathbb{R}^n$), a σ-finite measure μ on E is a Lévy measure if and only if $\int 1 \wedge |x|^2 \mu(dx) < \infty$ (see for instance [134]). For a general Banach space E, this condition is not necessary nor sufficient, instead, we have some partial characterisation from [9] and [61]:

Theorem 2.4. *Let E be a Banach space. Then*

a) E if of type p $\iff$ every Borel measure μ on E with $\mu(\{0\}) = 0$ satisfying the condition $\int 1 \wedge |x|^p \mu(dx)$ is a Lévy measure.

b) E if of cotype q $\iff$ every Lévy measure μ satisfies $\int 1 \wedge |x|^q \mu(dx)$.

We are finally ready to state a generalization of the Lévy-Khintchine decomposition to Banach space valued Lévy process (see for example [3] and [78]) :

Proposition 2.5. *Let X be a Lévy process taking values on a Banach space E. Then the law of $X(t)$ is infinitely divisible for all $t \geq 0$ and there exists $\gamma \in E$, a covariance operator $Q : E' \to E$ and a Lévy measure $\mathcal{X}$ on E, such that the characteristic function of X admits the decomposition*

$$\mathbb{E}[e^{i\langle X(t),u\rangle}] = e^{t\Psi_X(u)}, \quad u \in E',$$

where the characteristic exponent Ψ_X of X is given by

$$\Psi_X(u) = i\langle \gamma, u\rangle - \frac{1}{2}\langle u, Qu\rangle + \int_E K(x,u)\mathcal{X}(dx), \tag{2.2.2}$$

where recall that K is defined in (2.2.1).

The triplet $(\gamma, Q, \mathcal{X})$ is called the characteristic triplet of X and uniquely determines the process X up to distribution.

2.3 Cone-valued time parameters

As discussed in Section 2.1.3, to define the weak subordination we need to first consider the evaluation of a Lévy process X on a Hilbert space H at a multivariate time index τ, which we write as $\mathbf{X}(\tau)$. This object is defined in (2.3.1) and will be the focus of this section. First, in Proposition 2.6 and Corollary 2.7 we give sufficient conditions for $\mathbf{X}(\tau)$ to be well-defined as a random element of H. The distribution of $\mathbf{X}(\tau)$ is studied in Section 2.3.1. In Proposition 2.9 and Theorem 2.10 we show that the distribution of $\mathbf{X}(\tau)$ is infinitely divisible and give its triplet in terms of the triplets of X and the time parameter

τ. Some technical lemmas commonly seen in the subordination literature are presented in Lemma 2.11, and the stochastic continuity of $\tau \mapsto \mathbf{X}(\tau)$ is shown in Corollary 2.12.

Let H be a separable Hilbert space with a chosen orthonormal basis $(e_n)_n$. Suppose $(X(t))_{t\geq 0}$ is a Lévy process on H adapted to the filtration $(\mathcal{F}_t)_t$, then each $X_n(t) := \langle X(t), e_n\rangle$ defines a real valued Lévy process and the series $\sum_n X_n(t)e_n$ converges to $X(t)$ in probability on H (see for instance [142]).

Write $\mathbb{R}_+^{\mathbb{N}}$ for the set of all non-negative real numbers $\tau = (\tau_n)_{n\in\mathbb{N}}$, which we will refer to as cone-valued time parameters. Using $\mathbb{R}_+^N$ as an index set, we may define a stochastic process $\{\mathbf{X}(\tau), \tau \in \mathbb{R}_+^{\mathbb{N}}\}$ via the (formal) expression:

$$\mathbf{X}(\tau) := \sum_{n=1}^{\infty} X_n(\tau_n)e_n. \tag{2.3.1}$$

That is, we are evaluating each coordinate of X (with respect to the basis $(e_n)_n$) at a different time τ_n. It is not hard to see that the series is not necessarily convergent in H for every choice of $\tau \in \mathbb{R}_+^N$, unless X concentrates on a finite dimensional subspace of H. We therefore first give a sufficient condition for the convergence of the above series, namely we require the sequence $\tau = (\tau_n)_n$ to be uniformly bounded.

Proposition 2.6. *Let $T > 0$ and write $K_T := \{(\tau_n)_n \in \ell_\infty, |\tau_n|_\infty \leq T\}$.*

a) Suppose X is a square integrable Lévy process, then the series (2.3.1) is convergent in $L^2(\Omega, H)$ uniformly on K_T for all $T > 0$.

b) Suppose X is a compound Poisson process, then the series (2.3.1) is convergent $\mathbb{P}$-almost surely uniformly on K_T for all $T > 0$.

Proof. Write $\mathbf{X}^N(\tau) := \sum_{n\leq N} X_n(\tau_n)e_n$ for $N \in \mathbb{N}$, then clearly each $\mathbf{X}^N$ is a random variable taking values in H since it is defined by a finite sum. Let $T > 0$ be arbitrary. For any $N, M \in \mathbb{N}$ with $N > M$, we define

$$D^2_{N,M} : = \sup_{|\tau|\leq T} |\mathbf{X}^N(\tau) - \mathbf{X}^M(\tau)|^2 \leq \sum_{n=M+1}^{N} \sup_{|\tau|\leq T} X_n(\tau_n)^2. \tag{2.3.2}$$

Clearly each $D^2_{N,M}$ is a random element of H as well. To show the convergence of (2.3.1), it is sufficient to show that $D_{N,M} \to 0$ in L^2 or almost surely as $N, M \to \infty$.
(a) We assume first that X is a mean zero square integrable Lévy martingale, then $\mathbb{E}|X(T)|^2 = \sum_n \mathbb{E}[X_n(T)]^2 < \infty$ for $T \geq 0$. By Doob's inequality,

$$\mathbb{E}[D^2_{N,M}] \leq \sum_{n=M+1}^{N} \mathbb{E}\Big[\sup_{\tau_n\leq T} X_n(\tau_n)^2\Big] \leq 4 \sum_{n=M+1}^{N} \mathbb{E}\Big[X_n(T)^2\Big] \overset{M,N\to\infty}{\to} 0. \tag{2.3.3}$$

In general, suppose X is a square integrable Lévy process with mean $\mathbb{E}X(1) = \mu \in H$. Then $t \mapsto \widetilde{X}(t) := X(t) - t\mu$ defines a square integrable Lévy martingale. By the elementary inequality $(a+b)^2 \leq 2a^2 + 2b^2$, we have

$$\sup_{\tau_n \leq T} X_n(\tau_n)^2 \leq 2 \sup_{\tau_n \leq T} \widetilde{X}_n(\tau_n)^2 + 2T^2 \langle \mu, e_n \rangle^2$$

from which we can obtain

$$\mathbb{E}[D^2_{N,M}] \leq 8 \sum_{n=M+1}^{N} \mathbb{E}\Big[\widetilde{X}_n(T)^2\Big] + 2T^2 \sum_{n=M+1}^{N} \langle \mu, e_n \rangle^2 \to 0, \quad M, N \to \infty,$$

where the first sum converges to zero by (2.3.3) and the second sum converges to zero since $\sum \langle \mu, e_n \rangle^2 = |\mu|^2 < \infty$ by Parseval's identity.

(b) Suppose X is a compound Poisson process. For $\mathbb{P}$-almost every $\omega \in \Omega$, the function $t \mapsto X(\omega, t)$ is right continuous and takes only a finite number (say k) of values on each interval $[0, T]$, which we will denote as $X(\omega, r_1), \dots, X(\omega, r_k)$ for some $(r_1, \dots, r_k) \subset [0, T]$ depending on the choice of ω. Then

$$D^2_{N,M} \leq \sum_{n=M+1}^{N} \sup_{|\tau| \leq T} X_n(\tau_n)^2 = \sum_{n=M+1}^{N} \max_{1 \leq i \leq k} X_n(r_i)^2.$$

Since X takes values in H, for each $i \in \{1, \dots, k\}$, the sequence $(X_n(r_i))_n$ is in ℓ_2, $\mathbb{P}$-almost surely. Therefore $(\max_{1 \leq i \leq k} X_n(r_i))_n \in \ell_2$ as well and so $D_{N,M} \to 0$ almost surely. $\square$

As an easy consequence of the Lévy-Itô decomposition (see [5, 78]), we can combine the above two results and state the following.

Corollary 2.7. *Let X be a Lévy process in H and $\tau \in \ell^+_\infty$. Then*

$$\mathbf{X}(\tau) := \sum_{n=1}^{\infty} X_n(\tau_n) e_n$$

is convergent in probability to a random variable in H.

We remark here that clearly the random variable $\mathbf{X}(\tau)$ is almost surely equal to $\sum_{n \in \mathrm{supp}(\tau)} X_n(\tau_n) e_n$, where $\mathrm{supp}(\tau) := \{\mathrm{n} \in \mathbb{N}, \tau_\mathrm{n} \neq 0\}$.

2.3.1 Distribution and Lévy triplet

To state the distribution of the random variable $\mathbf{X}(\tau)$, we first introduce some notations. For $N \in \mathbb{N}$ and a sequence $\tau = (\tau_n)_n \in \mathbb{R}^{\mathbb{N}}_+$, we write τ^N for the truncated sequence $\tau^N := (\tau_n 1_{n \leq N})_n$. Let $\{(1)_N, \dots, (N)_N\}$ be a permutation of $\{1, \dots, N\}$ such that the sequence $(\tau_{(n)_N})_{n \leq N}$ is non-increasing, i.e. $\tau_{(1)_N} \geq \dots \geq \tau_{(N)_N}$. Write $\Delta\tau_{(n)_N} := \tau_{(n)_N} - \tau_{(n+1)_N}$, $n \in \{1, \dots, N\}$, with the convention $\tau_{(N+1)_N} := 0$. Finally let $\pi_{(1,\dots,n)_N}$ be the orthogonal projection onto the linear subspace of H spanned by $\{e_{(1)_N}, \dots, e_{(n)_N}\}$.

To state the distribution of $\mathbf{X}(\tau)$, we first introduce a outer product operation on the characteristic triplet $(\gamma, Q, \mathcal{X})$ of X. For $N \in \mathbb{N}$, define

$$\begin{aligned}
\tau^N \diamond \gamma &:= \sum_{n=1}^{N} \tau_n \langle \gamma, e_n \rangle, \\
\tau^N \diamond Q &:= \sum_{i=1}^{N} \sum_{j=1}^{N} (\tau_i \wedge \tau_j) \langle Q e_i, e_j \rangle e_i \otimes e_j, \\
\tau^N \diamond \mathcal{X} &:= \sum_{n=1}^{N} \Delta\tau_{(n)_N} \Big(\mathcal{X} \circ \pi^{-1}_{(1,\dots,n)_N} \Big|_{H \setminus \{0\}} \Big), \\
C_{\tau^N \diamond \mathcal{X}} &:= \sum_{n=1}^{N} \Delta\tau_{(n)_N} \int_{B^c} \Big(\pi^{-1}_{(1,\dots,n)_N} x \Big) 1_B \Big(\pi^{-1}_{(1,\dots,n)_N} x \Big) \mathcal{X}(dx),
\end{aligned}$$

where $e_i \otimes e_j$ is the operator $x \mapsto \langle x, e_i \rangle e_j$. It is routine to check that $\tau^N \diamond Q$ defines a covariance operator on H, $\tau^N \diamond \mathcal{X}$ defines a Lévy measure on H and $C_{\tau^N \diamond \mathcal{X}}$ is well defined as a sum of Bochner integrals taking values in H.

Proposition 2.8. *The law of $\mathbf{X}(\tau^N) = \sum_{n=1}^{N} X_n(\tau_n) e_n$ is infinitely divisible with characteristic triplet $(\tau^N \diamond \gamma + C_{\tau^N \diamond \mathcal{X}},\ \tau^N \diamond Q,\ \tau^N \diamond \mathcal{X})$.*

Proof. Let $u \in H$, then $\langle \mathbf{X}(\tau^N), u \rangle = \sum_{n=1}^{N} u_{(n)_N} X_{(n)_N}(\tau_{(n)_N})$ and we may write

$$\begin{aligned}
\langle \mathbf{X}(\tau^N), u \rangle &= \sum_{n=1}^{N} u_{(n)_N} \sum_{k=n}^{N} \Big(X_{(n)_N}(\tau_{(k)_N}) - X_{(n)_N}(\tau_{(k+1)_N}) \Big) \\
&= \sum_{k=1}^{N} \sum_{n=1}^{k} u_{(n)_N} \Big(X_{(n)_N}(\tau_{(k)_N}) - X_{(n)_N}(\tau_{(k+1)_N}) \Big) =: \sum_{k=1}^{N} Z_{N,k}.
\end{aligned}$$

By stationarity and independence of the increments of X, $(Z_{N,k})_{k \le N}$ is an independent sequence of infinitely divisible random variables with

$$Z_{N,k} = \sum_{n=1}^{k} u_{(n)_N} \Big(X_{(n)_N}(\tau_{(k)_N}) - X_{(n)_N}(\tau_{(k+1)_N}) \Big) \stackrel{\mathcal{D}}{=} \Big\langle \pi_{(1,\dots,k)_N} u, X(\Delta\tau_{(k)_N}) \Big\rangle.$$

Therefore $\mathbf{X}(\tau^N)$ is infinitely divisible as well with characteristic exponent given by

$$\Psi_{\mathbf{X}(\tau^N)}(u) = \sum_{k=1}^{N} \Psi_{Z_{N,k}}(1) = \sum_{k=1}^{N} \Delta\tau_{(k)_N} \Psi_{X(1)}(\pi_{(1,\dots,k)_N} u). \tag{2.3.4}$$

Write $|u|_Q := \langle u, Qu \rangle$. Expanding (2.3.4) using (2.2.2), we have

$$\begin{aligned}
\Psi_{\mathbf{X}(\tau^N)}(u) = i \sum_{k=1}^{N} \Delta\tau_{(k)_N} \langle \gamma, \pi_{(1,\dots,k)_N} u \rangle - \frac{1}{2} \sum_{k=1}^{N} \Delta\tau_{(k)_N} |\pi_{(1,\dots,k)_N} u|_Q \\
+ \sum_{k=1}^{N} \Delta\tau_{(k)_N} \int_H K(x, \pi_{(1,\dots,k)_N} u) \mathcal{X}(dx).
\end{aligned} \tag{2.3.5}$$

Writing $\gamma_n := \langle \gamma, e_n \rangle$ the first term in (2.3.5) is given by

$$\begin{aligned}\sum_{k=1}^{N} \Delta\tau_{(k)_N} \langle \gamma, \pi_{(1,\ldots,k)_N} u \rangle &= \sum_{k=1}^{N} \Delta\tau_{(k)_N} \sum_{n=1}^{k} \gamma_{(n)_N} u_{(n)_N} \\ &= \sum_{n=1}^{N} \gamma_{(n)_N} u_{(n)_N} \sum_{k=n}^{N} \Delta\tau_{(k)_N} = \sum_{n=1}^{N} \tau_{(n)_N} \gamma_{(n)_N} u_{(n)_N} = \langle \tau^N \diamond \gamma, u \rangle.\end{aligned}$$

Similarly, writing $Q_{ij} := \langle Q e_i, e_j \rangle$, the covariance term in (2.3.5) becomes

$$\begin{aligned}\sum_{k=1}^{N} \Delta\tau_{(k)_N} \langle \pi_{(1,\ldots,k)_N} u, Q \pi_{(1,\ldots,k)_N} u \rangle &= \sum_{i,j=1}^{N} Q_{(i)_N (j)_N} u_{(i)_N} u_{(j)_N} \sum_{k=i\vee j}^{N} \Delta\tau_{(k)_N} \\ &= \sum_{i,j=1}^{N} (\tau_{(i)_N} \wedge \tau_{(j)_N}) Q_{(i)_N (j)_N} u_{(i)_N} u_{(j)_N} = \langle u, (\tau^N \diamond Q) u \rangle.\end{aligned}$$

Finally, observe that $\{x : x \in B\} \subseteq \{x : \pi_{(1,\ldots,k)_N} x \in B\}$ so we have $1_B(x) = 1_B(\pi_{(1,\ldots,k)_N} x) - 1_B(\pi_{(1,\ldots,k)_N} x) 1_{B^c}(x)$. Furthermore, we have

$$\begin{aligned}K(x, \pi_{(1,\ldots,k)_N} u) &= e^{i\langle \pi_{(1,\ldots,k)_N} x, u \rangle} - 1 - i \langle \pi_{(1,\ldots,k)_N} x, u \rangle 1_B(x) \\ &= K(\pi_{(1,\ldots,k)_N} x, u) + i \langle \pi_{(1,\ldots,k)_N} x, u \rangle (1_B(\pi_{(1,\ldots,k)_N} x) - 1_B(x)).\end{aligned}$$

Therefore the last sum in (2.3.5) is given by

$$\begin{aligned}&\sum_{k=1}^{N} \Delta\tau_{(k)_N} \int_H K(x, \pi_{(1,\ldots,k)_N} u) \mathcal{X}(dx) \\ &= \int_H K(x, u) \left(\sum_{k=1}^{N} \Delta\tau_{(k)_N} \mathcal{X} \circ \pi^{-1}_{(1,\ldots,k)_N} \right)(dx) \\ &\quad + i \sum_{k=1}^{N} \Delta\tau_{(k)_N} \int_H \langle x, \pi_{(1,\ldots,k)_N} u \rangle 1_B(\pi_{(1,\ldots,k)_N} x) 1_{B^c}(x) \mathcal{X}(dx) \\ &= \int_H K(x, u)(\tau^N \diamond \mathcal{X})(dx) + C_{\tau^N \diamond \mathcal{X}}.\end{aligned}$$

The characteristic triplet of $\mathbf{X}(\tau^N)$ follows immediately. □

By Proposition 2.6 we have $\mathbf{X}(\tau^N) \to \mathbf{X}(\tau)$ in probability as $N \to \infty$ for all $\tau \in \ell_\infty$. Then the limit $\mathbf{X}(\tau)$ is infinitely divisible as well and the characteristic triplet $(\tau^N \diamond \gamma + C_{\tau^N \diamond \mathcal{X}}, \tau^N \diamond Q, \tau^N \diamond \mathcal{X})$ of $\mathbf{X}(\tau^N)$ should converge to the characteristic triplet of $\mathbf{X}(\tau)$ in suitable topologies, which is made precise by the following proposition. We note that the expressions for $\tau^N \diamond \mathcal{X}$ and $C_{\tau^N \diamond \mathcal{X}}$ involve series whose summands themselves depend on N and consequently the limits do not have an explicit expression. Nevertheless, we can derive estimates on the limits as functions of τ and $\mathcal{X}$.

Proposition 2.9. *Let $(\gamma, Q, \mathcal{X})$ be a characteristic triplet and $\tau \in \ell_\infty$.*

a) *The sequence* $(\tau^N \diamond \gamma)_N$ *is convergent in the norm of* H *to*

$$\tau \diamond \gamma := \lim_{N\to\infty} \tau^N \diamond \gamma = \sum_{n=1}^{\infty} \tau_n \langle \gamma, e_n \rangle,$$

and the limit satisfies $|\tau \diamond \gamma|_H \leq |\tau||\gamma|_H$.

b) *The sequence* $(C_{\tau^N \diamond \mathcal{X}})_N$ *is convergent in the norm of* H *to*

$$C_{\tau\diamond\mathcal{X}} := \lim_{N\to\infty} C_{\tau^N\diamond\mathcal{X}}$$

and the limit satisfies $|C_{\tau\diamond\mathcal{X}}|_H \leq |\tau|\mathcal{X}(B^c)$.

c) *The sequence* $(\tau^N \diamond Q)_N$ *is convergent in the trace norm of* $L_1(H)$ *to*

$$\tau \diamond Q := \lim_{N\to\infty} \tau^N \diamond Q = \sum_{n=1}^{\infty} (\tau_i \wedge \tau_j)\langle Qe_i, e_j\rangle e_i \otimes e_j.$$

The limit is a covariance operator satisfying $|\tau \diamond Q|_{L_1(H)} \leq |\tau||Q|_{L_1(H)}$.

d) *There exists a Lévy measure* $\tau \diamond \mathcal{X}$ *on* H *such that*

$$\tau^N \diamond \mathcal{X}|_{U^c} \Rightarrow \tau \diamond \mathcal{X}|_{U^c}$$

holds for any neighborhood U *of zero, where* $\Rightarrow$ *denotes convergence in the topology of weak convergence of (finite) measures. Furthermore, the limit satisfies*

$$(\tau \diamond \mathcal{X})(B^c_\epsilon) \leq |\tau|\mathcal{X}(B^c_\epsilon), \quad \forall \epsilon > 0.$$

e) *Each* $1 \wedge |x|^2(\tau^N \diamond \mathcal{X})(dx)$ *is a finite Borel measure and*

$$1 \wedge |x|^2(\tau^N \diamond \mathcal{X})(dx) \Rightarrow 1 \wedge |x|^2(\tau \diamond \mathcal{X})(dx).$$

Furthermore for all $r > 0$,

$$\int_{B_r} 1 \wedge |x|^2(\tau \diamond \mathcal{X})(dx) \leq |\tau|\Big((1 \wedge r^2)\mathcal{X}(B^c_r) + \int_{B_r} 1 \wedge |x|^2\mathcal{X}(dx)\Big).$$

Proof. (a) is obvious since $\gamma \mapsto \sum_n \tau_n \langle \gamma, e_n\rangle e_n$ is the multiplication operator M_τ, diagonal with respect to $(e_n)_n$ with eigenvalues $(\tau_n)_n \in \ell_\infty$.

(b) By Theorem VI 5.5 of [134], it holds that $\tau^N \diamond \gamma + C_{\tau^N\diamond\mathcal{X}}$ converges in the norm of H to some limit. Since the sequence $(\tau^N \diamond \gamma)_N$ is convergent by (a), we conclude that $(C_{\tau^N\diamond\mathcal{X}})_N$ must converge as well. Moreover, for $N \in \mathbb{N}$, by the triangle inequality we have

$$\begin{aligned}|C_{\tau^N\diamond\mathcal{X}}|_H &\le \sum_{n=1}^N \Delta\tau_{(n)_N}\int_{|x|\ge 1}|\pi^{-1}_{(1,\dots,n)_N}x|1_B(\pi^{-1}_{(1,\dots,n)_N}x)\mathcal{X}(dx)\\ &\le \sum_{n=1}^N \Delta\tau_{(n)_N}\mathcal{X}(B^c) = \tau_{(1)_N}\mathcal{X}(B^c) \le |\tau|\mathcal{X}(B^c),\end{aligned}$$

which gives the estimate $|C_{\tau\diamond\mathcal{X}}|_H = \lim_N |C_{\tau^N\diamond\mathcal{X}}|_H \le |\tau|\mathcal{X}(B^c)$.

(c) Let $M, N \in \mathbb{N}$ with $M > N$. Then by the definition of $\tau^N \diamond Q$, we have

$$|\tau^M \diamond Q - \tau^N \diamond Q|_{L_1(H)} = \sum_{n=N+1}^M \tau_n\langle Qe_n, e_n\rangle \le |\tau| \sum_{n=N+1}^M \langle Qe_n, e_n\rangle \to 0,$$

since Q is of trace class. It is routine to check that the limit $\tau \diamond Q$ is symmetric and positive definite. Furthermore, we have $|\tau \diamond Q|_{L_1(H)} = \sum_n \tau_n Q_{nn} \le |\tau||Q|_{L_1(H)} < \infty$.

(d) By Theorem VI 5.5 of [134], for each $\epsilon > 0$, the measure $(\tau^N \diamond \mathcal{X})|_{B^c_\epsilon}$ is a finite Lévy measure and converges weakly to a Lévy measure on H, which we call $\tau \diamond \mathcal{X}$. Since $|\pi_{(1,\dots,n)_N}x| \le |x|$, for $\epsilon > 0$, we have

$$\mathcal{X} \circ \pi^{-1}_{(1,\dots,n)_N}(B^c_\epsilon) = \mathcal{X}(\{|\pi_{(1,\dots,n)_N}x| > \epsilon\}) \le \mathcal{X}(B^c_\epsilon).$$

We therefore have the bound

$$(\tau^N \diamond \mathcal{X})(B^c_\epsilon) = \sum_{n=1}^N \Delta\tau_{(n)_N}\mathcal{X} \circ \pi^{-1}_{(1,\dots,n)_N}(B^c_\epsilon) \le \tau_{(1)_N}\mathcal{X}(B^c_\epsilon) \le |\tau|\mathcal{X}(B^c_\epsilon),$$

which holds uniformly over $N \in \mathbb{N}$, which gives our claimed estimate on $\tau \diamond \mathcal{X}$.

(e) For an arbitrary $E \in \mathcal{B}(H)$, $N \in \mathbb{N}$ and $\epsilon > 0$, write

$$a_{N,\epsilon} := \int_{E\cap B^c_\epsilon} 1 \wedge |x|^2 (\tau^N \diamond \mathcal{X})(dx),$$

so that $\lim_{\epsilon\to 0}\lim_{N\to\infty} a_{N,\epsilon}$ exists and is equal to

$$\begin{aligned}\lim_{\epsilon\to 0}\lim_{N\to\infty} a_{N,\epsilon} &= \lim_{\epsilon\to 0}\lim_{N\to\infty}\int_{E\cap B^c_\epsilon} 1 \wedge |x|^2(\tau^N \diamond \mathcal{X})(dx)\\ &= \lim_{\epsilon\to 0}\int_{E\cap B^c_\epsilon} 1 \wedge |x|^2(\tau \diamond \mathcal{X})(dx)\\ &= \int_E 1 \wedge |x|^2(\tau \diamond \mathcal{X})(dx),\end{aligned}$$

where the second equality holds by the weak convergence of $\tau^N \diamond \mathcal{X}$ to $\tau \diamond \mathcal{X}$ on sets bounded away from zero and the last expression is finite since $\tau \diamond \mathcal{X}$ is a Lévy measure on H. For $N \in \mathbb{N}$, since $E - E \cap B^c_\epsilon = E \cap B_\epsilon \subseteq B_\epsilon$,

$$D_{N,\epsilon} := \left|a_{N,\epsilon} - \int_E 1 \wedge |x|^2(\tau^N \diamond \mathcal{X})(dx)\right| \le \int_H 1_{B_\epsilon}(x) 1 \wedge |x|^2(\tau^N \diamond \mathcal{X})(dx)$$

$$= \sum_{n=1}^{N} \Delta\tau_{(n)_N} \int_H 1_{B_\epsilon}(x) 1 \wedge |x|^2 (\mathcal{X} \circ \pi_{(1,\ldots,n)_N}^{-1})(dx)$$
$$= \sum_{n=1}^{N} \Delta\tau_{(n)_N} \int_H 1_{B_\epsilon}(\pi_{(1,\ldots,n)_N} x) 1 \wedge |\pi_{(1,\ldots,n)_N} x|^2 \mathcal{X}(dx). \tag{2.3.6}$$

Since $|\pi_{(1,\ldots,n)_N} x| \leq |x|$, we can easily derive the identity

$$1_{B_\epsilon}(\pi_{(1,\ldots,n)_N} x) = 1_{B_\epsilon}(x) + 1_{B_\epsilon^c}(x) 1_{B_\epsilon}(\pi_{(1,\ldots,n)_N} x).$$

The integrand of (2.3.6) is then bounded pointwise on H by a function f_ϵ on H given by

$$f_\epsilon(x) := (1 \wedge |x|^2) 1_{B_\epsilon}(x) + (1 \wedge \epsilon^2) 1_{B_\epsilon^c}(x)$$

uniformly in N, and we have the bound

$$D_{N,\epsilon} \leq \sum_{n=1}^{N} \Delta\tau_{(n)_N} \int_H f_\epsilon(x) \mathcal{X}(dx) \leq |\tau| \int_H f_\epsilon(x) \mathcal{X}(dx).$$

As $\epsilon \to 0$, $f_\epsilon(x)$ converges pointwise on H to zero (uniformly in N). Since

$$f_\epsilon(x) \leq 1 \wedge |x|^2 \in L^1(H, \mathcal{B}(H), \mathcal{X}),$$

by the dominated convergence theorem, $D_{N,\epsilon} \to 0$ uniformly in N. Equivalently, $a_{N,\epsilon}$ is uniformly convergent to the limit $\int_E 1 \wedge |x|^2 (\tau^N \diamond \mathcal{X})$ as $\epsilon \to 0$. Together with the existence of $\lim_{N\to\infty} a_{N,\epsilon}$ for each $\epsilon > 0$, we conclude that the two limits commute and therefore

$$\int_E 1 \wedge |x|^2 (\tau \diamond \mathcal{X})(dx) = \lim_{\epsilon\to 0} \lim_{N\to\infty} a_{N,\epsilon}$$
$$= \lim_{N\to\infty} \lim_{\epsilon\to 0} a_{N,\epsilon} = \lim_{N\to\infty} \int_E 1 \wedge |x|^2 (\tau^N \diamond \mathcal{X})(dx),$$

which implies the weak convergence of $1 \wedge |x|^2 (\tau^N \diamond \mathcal{X})(dx) \Rightarrow 1 \wedge |x|^2 (\tau \diamond \mathcal{X})(dx)$ as $N \to \infty$, since $E \in \mathcal{B}(H)$ is arbitrary.

By the same arguments as above, we see that for all $N \in \mathbb{N}$ and $r > 0$,

$$\int_{B_r} 1 \wedge |x|^2 (\tau^N \diamond \mathcal{X})(dx) \leq \int_H f_r(x) \mathcal{X}(dx)$$
$$\leq |\tau| \Big(\int_{B_r} 1 \wedge |x|^2 \mathcal{X}(dx) + \mathcal{X}(B_r^c)(1 \wedge r^2) \Big)$$

which is uniform in N. By the weak convergence of the sequence $1 \wedge |x|^2 (\tau^N \diamond \mathcal{X})(dx)$, the bound also applies to the limit $1 \wedge |x|^2 (\tau \diamond \mathcal{X})(dx)$ and the inequality in (e) follows. □

We summarize the above results into the following theorem:

Theorem 2.10. *Fix $\tau \in \ell_\infty$ and let X be a Lévy process on H with characteristic triplet*

$(\gamma, Q, \mathcal{X})$. *Then the H-valued random variable $\mathbf{X}(\tau)$ is infinitely divisible with characteristic triplet $(\tau \diamond \gamma + C_{\tau\diamond\mathcal{X}}, \tau \diamond Q, \tau \diamond \mathcal{X})$ as defined in Proposition 2.9.*

The following technical lemma is analogous to Lemma 30.3 of Sato [152], equations (3.17)-(3.20) of Barndorff-Nielsen et al. [21], Lemma 27 of Pérez-Abreu and Rocha-Arteaga [141] and Lemma 5.1 of Buchmann et al. [49]. These estimates will be used later on in the Chapter to control certain integrals against the Lévy measure $\tau \diamond \mathcal{X}$.

Lemma 2.11. *The exists finite constants C_1, C_2, C_3, only dependent on the triplet $(\gamma, Q, \mathcal{X})$, such that for all $\tau \in \ell_\infty$,*

$$\mathbb{P}\Big(|\mathbf{X}(\tau)|^2 > 1 - |\tau|\Big) \leq C_1(|\tau| + |\tau|^2), \tag{2.3.7}$$

$$\mathbb{E}\Big[|\mathbf{X}(\tau)|^2 1_{|\mathbf{X}(\tau)|\leq 1}\Big] \leq C_2(|\tau| + |\tau|^2), \tag{2.3.8}$$

$$\Big|\mathbb{E}\Big[\mathbf{X}(\tau) 1_{|\mathbf{X}(\tau)|\leq 1}\Big]\Big| \leq C_3|\tau|. \tag{2.3.9}$$

Proof. Let Y_0 and Y_1 be Lévy processes on H with characteristic triplets $(0, 0, (\tau \diamond \mathcal{X})|_{B^c})$ and $(\tau \diamond \gamma + C_{\tau\diamond\mathcal{X}}, \tau \diamond Q, (\tau \diamond \mathcal{X})|_B)$ respectively, then Y_0 is a compound Poisson process with jumps of norm larger than 1, and Y_1 has bounded jumps and hence moments of all orders. By Theorem 2.10, We have $Y_0(1) + Y_1(1) \overset{\mathcal{D}}{=} \mathbf{X}(\tau)$. Observe that

$$\mathbb{P}(|\mathbf{X}(\tau)|^2 > 1 - |\tau|) \leq \mathbb{P}(Y_0(1) \neq 0) + \mathbb{P}(|Y_1(1)|^2 > 1 - |\tau|). \tag{2.3.10}$$

Since $1 - e^{-x} \leq x$, we have

$$\mathbb{P}(Y_0(1) \neq 0) \leq 1 - e^{-(\tau\diamond\mathcal{X})(B^c)} \leq (\tau \diamond \mathcal{X})(B^c) \leq |\tau|\mathcal{X}(B^c),$$

where the last inequality holds by Proposition 2.9 (d). For the second term in (2.3.10), Markov's inequality gives $\mathbb{P}(|\tau| + |Y_1(1)|^2 > 1) \leq |\tau| + \mathbb{E}[|Y_1(1)|^2]$, where the second term is finite since Y_1 has moments of all orders. By the monotone convergence theorem

$$\begin{aligned}
\mathbb{E}[|Y_1(1)|^2] &= \sum_{n=1}^{\infty} (\mathbb{E}\langle Y_1(1), e_n\rangle)^2 + \sum_{n=1}^{\infty} \mathrm{Var}\langle Y_1(1), e_n\rangle \\
&= \sum_{n=1}^{\infty} \langle \mathbb{E}[Y_1(1)], e_n\rangle^2 + \sum_{n=1}^{\infty} \langle \mathrm{Cov}(Y_1(1))e_n, e_n\rangle \\
&= |\mathbb{E}[Y_1(1)]|^2 + |\mathrm{Cov}(Y_1(1))|_{L_1(H)},
\end{aligned}$$

where, by the definition of $Y_1(1)$, $\mathbb{E}[Y_1(1)] = \tau \diamond \gamma + C_{\tau\diamond\mathcal{X}}$ and $\mathrm{Cov}(Y_1(1)) = \tau \diamond Q + \int_B x\langle x, \cdot\rangle(\tau \diamond \mathcal{X})(dx)$. By the estimates in Proposition 2.9 (a), (b) and the inequality $(a + b)^2 \leq 2a^2 + 2b^2$, the term $|\mathbb{E}[Y_1(1)]|^2$ is bounded by

$$|\mathbb{E}[Y_1(1)]|^2 = 2|\tau \diamond \gamma|^2 + 2|C_{\tau\diamond\gamma}|^2 \leq \Big(2|\gamma|^2 + 2\mathcal{X}(B^c)^2\Big)|\tau|^2.$$

Similarly, by the triangle inequality and Proposition 2.9 (c) and (e), we have

$$\begin{aligned}|\mathrm{Cov}(Y_1(1))|_{L_1} &\leq |\tau \diamond Q|_{L_1} + \int_B |x|^2 (\tau \diamond \mathcal{X})(dx) \\ &\leq |\tau| \Big(|Q|_1 + \mathcal{X}(B^c) + \int_B |x|^2 \mathcal{X}(dx) \Big).\end{aligned}$$

Collecting the above terms, we see that (2.3.7) holds with

$$C_1 = 1 + 2|\gamma|^2 + |Q|_{L_1} + 2\mathcal{X}(B^c) + 2\mathcal{X}(B^c)^2 + \int_B |x|^2 \mathcal{X}(dx).$$

Next, writing $1_{|\mathbf{X}(\tau)| \leq 1} = (1_{Y_0 \neq 0} + 1_{Y_0 = 0}) 1_{|\mathbf{X}(\tau)| \leq 1}$, we get

$$\mathbb{E}[|\mathbf{X}(\tau)|^2 1_{|\mathbf{X}(\tau) \leq 1|}] \leq \mathbb{P}(Y_0(1) \neq 0) + \mathbb{E}[|Y_1(1)|^2], \tag{2.3.11}$$

from which (2.3.8) follows with $C_2 = C_1 - 1$. Similar to the above, we have

$$\Big| \mathbb{E}[\mathbf{X}(\tau) 1_{|\mathbf{X}(\tau)| \leq 1}] \Big| \leq \mathbb{P}(Y_0(1) \neq 0) + \Big| \mathbb{E}[Y_1(1) 1_{|Y_1(1)| \leq 1}] \Big|,$$

where $|\mathbb{E}[Y_1(1) 1_{|Y_1(1)| \leq 1}]| \leq |\tau \diamond \gamma| + |C_{\tau \diamond \mathcal{X}}| \leq |\tau|(|\gamma| + \mathcal{X}(B^c))$. This gives (2.3.9) with $C_3 = |\gamma| + 2\mathcal{X}(B^c)$. □

As a consequence of the above estimates, we can show that $\{\mathbf{X}(\tau), \tau \in \ell_\infty^+\}$ is stochastically continuous when considered as a process indexed by ℓ_∞^+.

Corollary 2.12. *The process $\{\mathbf{X}(\tau), \tau \in \ell_\infty\}$ is stochastically continuous, i.e. whenever $\tau^k \to \tau$ in norm as $k \to \infty$, $\mathbf{X}(\tau^k) \to \mathbf{X}(\tau)$ in probability.*

Proof. Note that since each X_n is a Lévy process on $\mathbb{R}$, we have

$$|\mathbf{X}(\tau^k) - \mathbf{X}(\tau)|^2 = \sum_n |X_n(\tau_n^k) - X_n(\tau_n)|^2 \stackrel{\mathcal{D}}{=} \sum_n |X_n(\tau_n^k - \tau_n)|^2$$

so it suffices to show that the ℓ_∞^+-indexed process $\{\mathbf{X}(\tau), \tau \in \ell_\infty^+\}$ is stochastically continuous at zero. Let $\tau^k \to 0$ as $k \to \infty$. Then

$$\begin{aligned}\mathbb{E}\Big[1 \wedge |\mathbf{X}(\tau^k)|^2\Big] &= \mathbb{P}\Big(|\mathbf{X}(\tau^k)| > 1\Big) + \mathbb{E}\Big[|\mathbf{X}(\tau^k)|^2 1_{|\mathbf{X}(\tau^k)| \leq 1}\Big] \\ &\leq \mathbb{P}\Big(|\mathbf{X}(\tau^k)| > 1 - |\tau^k|_\infty\Big) + \mathbb{E}\Big[|\mathbf{X}(\tau^k)|^2 1_{|\mathbf{X}(\tau^k)| \leq 1}\Big] \\ &\leq (C_1 + C_2)(|\tau^k|_\infty + |\tau^k|_\infty^2),\end{aligned}$$

where the last estimate follows from Lemma 2.11 and the constants C_1 and C_2 depend only on the triplet of X. Taking $k \to \infty$ shows that $|X(\tau^k)|$ converges to zero in probability. □

2.4 Lévy subordinators on ℓ_∞^+

Based on Proposition 2.6 we wish to choose subordinators taking values in the space of bounded, positive sequences. However, since ℓ_∞^+ is not separable, we begin our analysis in a weighted ℓ_1 space that contains ℓ_∞^+ as a subset. This is the focus of Section 2.4.1. In Definition 2.13 we define a subordinator T on the positive cone of this weighted space. Proposition 2.14 then gives the Lévy-Khintchine decomposition of T. Once this is completed, Theorem 2.15 in Section 2.4.1 gives sufficient conditions for T to admit a Lévy-Itô decomposition in ℓ_∞, which is the space we wanted originally.

We remark that our choice of weighted ℓ_1 over other weighted ℓ_p spaces (in particular the Hilbert space ℓ_2) is necessary for a nice characterisation of Lévy subordinators in Banach spaces. More precise results and discussions on Banach space valued subordinators we refer to [138, 141, 150].

Let $w = (w_n)_n$ be a summable sequence of positive numbers normalized to $|\omega|_1 = 1$. Define the w-weighted sequence space $\ell_{1,w}$:

$$\ell_{1,w} := \left\{ x \in \mathbb{R}^{\mathbb{N}}, \sum_{n=1}^{\infty} |x_n| w_n < \infty \right\}$$

endowed with the weighted ℓ_1 norm $|x|_{1,w} := \sum_{n=1}^{\infty} |x_n| w_n$. Then $\ell_{1,w}$ is a separable Banach space of type 1. We first state some preliminary facts and introduce the necessary notations on such sequence spaces. More details can be found in for instance [1].

Since w is a positive sequence, the counting measure on $\mathbb{N}$ is equivalent to the w-weighted counting measure, so trivially $\ell_{\infty,w} = \ell_\infty$. It is then clear that the topological dual $(\ell_{1,w})^*$ of $\ell_{1,w}$ is isomorphic to ℓ_∞, and we write $\langle x, y \rangle := \sum_{n=1}^{\infty} x_n y_n w_n$ for the duality pairing between $\ell_{1,w}$ and ℓ_∞. The Borel σ-algebra $\mathcal{B}(\ell_{1,w})$ coincides with the cylindrical σ-algebra generated by sets of the form $\{x \in \ell_{1,w} : (\langle x, u_1 \rangle, \dots, \langle x, u_n \rangle) \in A\}$, where $u_1, \dots, u_n$ are elements of ℓ_∞ and $A \in \mathcal{B}(\mathbb{R}^n)$. In particular, sets of the form $\{x \in \ell_{1,w}, x_n \geq 0, \forall n \in J\}$ where $J \subseteq \mathbb{N}$ are Borel measurable.

The standard basis $(e_n)_n$ of ℓ_1 is a Schauder basis for $\ell_{1,w}$ as well, and the corresponding biorthogonal functionals $(e_n^*)_n \subseteq (\ell_{1,w})^*$ can be identified isometrically with the collection of bounded sequences $(w_n^{-1} e_n)_n \subseteq \ell_\infty$. That is, we have the identity $e_n^*(x) = x_n = \langle x, w_n^{-1} e_n \rangle$ for any $x \in \ell_{1,w}$. Therefore each $e_n^* : \ell_{1,w} \to \mathbb{R}$ is a Borel function, and the function $|\cdot|_\infty = \sup_n \langle \,\cdot\, , e_n^* \rangle : \ell_{1,w} \to \mathbb{R}$ is Borel measurable as well. This implies that sets of the form $\{|x|_\infty \in A, A \in \mathcal{B}(\mathbb{R})\}$ are elements of $\mathcal{B}(\ell_{1,w})$. The space ℓ_∞ is a subspace of $\ell_{1,w}$ and the embedding $\ell_\infty \hookrightarrow \ell_{p,w}$ is continuous, since $|x|_{1,w} \leq |x|_\infty |w|_1$. On the other hand, ℓ_∞ is not a closed subspace of $\ell_{1,w}$ under the weighted norm.

Let $\ell_{1,w}^+ := \{x \in \ell_{1,w}, x_n \geq 0, \forall n \geq 1\}$ be the cone of non-negative sequences in $\ell_{1,w}$. Then $\ell_{1,w}^+$ is a proper cone in the sense that $x \in \ell_{1,w}^+$ and $-x \in \ell_{1,w}^+$ implies $x = 0$. Since $|x|_{1,w} = \sum_n w_n x_n$ on the cone $\ell_{1,w}^+$, the norm $|\cdot|_{1,w}$ on $\ell_{1,w}^+$ agrees with the linear functional

$w = (w_n)_n \in \ell_\infty$. Cones with this property are called Birkhoff-Kakutani cones, see [150]. The results in [150] imply that subordinators can be defined on Birkhoff-Kakutani cones in a similar way to the finite dimensional setting. That is, we can define:

Definition 2.13. A Lévy process process $(T(t))_{t\geq 0}$ in the sequence space $\ell_{1,w}$ is said to be an $\ell_{1,w}^+$-subordinator if and only if $T(t)$ takes values in $\ell_{1,w}^+$ almost surely for all $t \geq 0$.

It can be shown that $(T(t))_{t\geq 0}$ takes values in $\ell_{1,w}^+$ a.s. for each t if T is $\ell_{1,w}^+$-increasing, i.e. we have $T(t) - T(s) \in \ell_{1,w}^+$ for any $s \leq t$. The results of [150] imply the following characterization of subordinators defined on $\ell_{1,w}$ which is a Birkhoff-Kakutani cone.

Proposition 2.14 (Proposition 8, [150])**.** *Let $(T(t))_{t\geq 0}$ be a Lévy process on $\ell_{1,w}$. Then T is an $\ell_{1,w}^+$-subordinator iff there exists $\eta \in \ell_{1,w}^+$ and a Lévy measure $\mathcal{T}$ on $\ell_{1,w}$ concentrated on $\ell_{1,w}^+ \setminus \{0\}$ and satisfying*

$$\int_{\ell_{1,w}^+} 1 \wedge |x|_{1,w} \mathcal{T}(dx) < \infty, \tag{2.4.1}$$

such that T admits a Lévy-Khintchine decomposition of the form

$$\mathbb{E}[e^{i\langle T(t),u\rangle}] = \exp\left\{ it\langle \eta, u\rangle + t\int_{\ell_{1,w}^+} \left(e^{i\langle x,u\rangle} - 1\right)\mathcal{T}(dx)\right\}, \quad u \in \ell_\infty.$$

In this case, the Laplace transform of T is given by

$$\mathbb{E}[e^{-\langle T(t),u\rangle}] = \exp\left\{ -t\langle \eta, u\rangle - t\int_{\ell_{1,w}^+} \left(1 - e^{-\langle x,u\rangle}\right)\mathcal{T}(dx)\right\}$$

and the associated norm process $(|T(t)|_{1,w})_{t\geq 0}$ is a real-valued subordinator.

2.4.1 Lévy-Itô decomposition on ℓ_∞^+

Using the results of [150], especially the Lévy-Khintchine decomposition in Proposition 2.14, we have defined a subordinator T on the weighted sequence space $\ell_{1,w}^+$. Recall from earlier discussions that for the purpose of constructing the weak subordination, it is necessary to define T on the space ℓ_∞.

Let $(T(t))_{t\geq 0}$ be a subordinator on $\ell_{1,w}^+$ with drift $\eta \in \ell_{1,w}^+$ and Lévy measure $\mathcal{T}$. We formulate a sufficient condition for the law of $T(t)$ to be concentrated on the sub-cone ℓ_∞^+. Let $N_t(dx)$ be the Poisson random measure associated with T, i.e.

$$N_t(A) := \sum_{0<s\leq t} 1_A(\Delta T(s)), \quad A \in \mathcal{B}(\ell_{1,w}^+).$$

Theorem 2.15. *Suppose $\eta \in \ell_{1,w}^+$ and the Lévy measure $\mathcal{T}$ of T concentrates on the subcone ℓ_∞^+ and is finite on the set $\{|x|_\infty > 1\}$. Then* *a) The $\ell_{1,w}^+$-subordinator T admits a Lévy-Itô decomposition of the form*

$$T(t) = t\eta + \int_{|x|_\infty > 1} xN_t(dx) + \lim_{\epsilon \downarrow 0} \int_{|x|_\infty \in (\epsilon,1]} xN_t(dx), \tag{2.4.2}$$

where the first integral is almost surely equal to a sum of finitely many ℓ_∞^+-valued terms, and the limit is taken with respect to $|\cdot|_{1,w}$.

b) If furthermore $\eta \in \ell_\infty^+$ and the Lévy measure $\mathcal{T}$ satisfies

$$\int_{|x|_\infty \in (0,1]} |x|_\infty \mathcal{T}(dx) < \infty, \tag{2.4.3}$$

then $T(t)$ takes values in ℓ_∞^+ almost surely for every $t \geq 0$.

Proof. We precede the proof with the following technical lemmas.

Lemma 2.16. *The first integral in (2.4.2) is almost surely equal to a sum of finitely many terms in ℓ_∞^+. Furthermore, it is a.s. equal to the limit*

$$\int_{|x|_\infty > 1} xN_t(dx) = \lim_{\epsilon \downarrow 0} \int_{|x|_\infty > 1} x1_{|x|_{1,w} > \epsilon} N_t(dx)$$

Proof. Since $\mathcal{T}$ is finite on $\{|x|_\infty > 1\}$, so is $N_t(\cdot)$ almost surely for each t, and the first integral in (2.4.2) is a finite sum taking values in ℓ_∞^+.

Since $|x|_{1,w} > 0$ whenever $|x|_\infty > 1$, the function $x \mapsto x1_{|x|_{1,w} > \epsilon} 1_{|x|_\infty > 1}$ converges pointwise to $x \mapsto x1_{|x|_\infty > 1}$ as $\epsilon \to 0$. Furthermore, $x1_{|x|_{1,w} > \epsilon} 1_{|x|_\infty > 1}$ has $|\cdot|_{1,w}$ norm bounded by $|x|_{1,w} 1_{|x|_\infty > 1}$ which is integrable with respect to $N_t(dx)$ a.s. since N_t is a finite measure on $\{|x|_\infty > 1\}$ almost surely. By dominated convergence the limit exists and is equal to the left hand side. □

Lemma 2.17. *The limit in equation (2.4.2) exists and is equal to*

$$\lim_{\epsilon \downarrow 0} \int_{|x|_\infty \in (\epsilon,1]} xN_t(dx) = \lim_{\epsilon \downarrow 0} \int_{|x|_\infty \in (\epsilon,1]} x1_{|x|_{1,w} > \epsilon} \, N_t(dx).$$

Proof. Since $|x|_{1,w} \leq |x|_\infty$, we have $1_{|x|_\infty \in (\epsilon,1]} \leq 1_{|x|_{1,w} \in (0,1]}$. Therefore we have

$$|x|_{1,w} 1_{|x|_{1,w} > \epsilon} 1_{|x|_\infty \in (\epsilon,1]} \leq |x|_{1,w} 1_{|x|_{1,w} \in (0,1]},$$

which is integrable with respect to $N_t(dx)$ almost surely by (2.4.1). Since $|x|_{1,w} = 0$ iff $x = 0$, the functions $x \mapsto x1_{|x|_\infty \in (\epsilon,1]}$ and $x \mapsto x1_{|x|_\infty \in (\epsilon,1]} 1_{|x|_{1,w} > \epsilon}$ both converge pointwise to the limit $x \mapsto x1_{|x|_\infty \in (0,1]}$. By the dominated convergence theorem both limits exists and are therefore equal. □

Proof of Theorem 2.15. (a) Since T is a subordinator on $\ell_{1,w}^+$, by the Lévy-Itô decomposition of T on $\ell_{1,w}^+$ we have

$$T(t) = t\eta + \lim_{\epsilon \downarrow 0} \int_{|x|_{1,w} > \epsilon} x N_t(dx), \tag{2.4.4}$$

where the limit is taken with respect to $|\cdot|_{1,w}$. Since $|x|_{1,w} \leq |x|_\infty$, we have $|x|_\infty \geq \epsilon$ whenever $|x|_{1,w} > \epsilon$ and therefore we have the equality

$$1_{|x|_{1,w} > \epsilon} = 1_{|x|_{1,w} > \epsilon} 1_{|x|_\infty > 1} + 1_{|x|_{1,w} > \epsilon} 1_{\epsilon < |x|_\infty \leq 1}.$$

The limit in (2.4.4) can therefore be written as

$$\begin{aligned} \lim_{\epsilon \downarrow 0} \left[\int_{|x|_\infty > 1} x 1_{|x|_{1,w} > \epsilon} N_t(dx) + \int_{|x|_\infty \in (\epsilon, 1]} x 1_{|x|_{1,w} > \epsilon} N_t(dx) \right] \\ = \int_{|x|_\infty > 1} x N_t(dx) + \lim_{\epsilon \downarrow 0} \int_{|x|_\infty \in (\epsilon, 1]} x N_t(dx), \end{aligned}$$

where the last equality holds because the first term has a limit in $|\cdot|_{1,w}$ by Lemma 2.16 and hence so does the second term. This gives (2.4.2).
(b) Now suppose $\eta \in \ell_\infty^+$ and the condition (2.4.3) holds. By Lemma 2.16, the term $\int_{|x|_\infty > 1} x N_t(dx)$ is almost surely in ℓ_∞^+, since it is equal to a sum of finitely many terms all taking values in ℓ_∞^+. It remains to show that

$$\widetilde{T}(t) := \int_{|x|_\infty \in (\epsilon, 1]} x N_t(dx) = \lim_{\epsilon \downarrow 0} \sum_{0 \leq u \leq t} \Delta T(u) 1_{|\Delta T(u)|_\infty \in (\epsilon, 1]}$$

takes values in ℓ_∞^+ as well. Note that for $\epsilon > 0$, since N_t is finite on $\{|x|_\infty \in (\epsilon, 1]\}$ a.s., the term $\sum_{0 \leq u \leq t} \Delta T(u) 1_{|\Delta T(u)|_\infty \in (\epsilon, 1]}$ is a sum of finitely many terms with values in ℓ_∞^+.

For $x \in \ell_{1,w}$, recall that the n-th coordinate x_n of x is given by $x_n = e_n^*(x) = \langle x, w_n^{-1} e_n \rangle$, where $w_n^{-1} e_n$ is a vector of norm w_n^{-1} in ℓ_∞. Since $\langle \,\cdot\,, w_n^{-1} e_n \rangle$ is continuous with respect to the norm $|\cdot|_{1,w}$ and the term $\widetilde{T}(t)$ is obtained as a $|\cdot|_{1,w}$-limit, we can write

$$\begin{aligned} |\widetilde{T}(t)|_\infty &= \sup_n \left\langle \lim_{\epsilon \downarrow 0} \sum_{0 \leq u \leq t} \Delta T(u) 1_{|\Delta T(u)|_\infty \in (\epsilon, 1]}, w_n^{-1} e_n \right\rangle \\ &= \sup_n \lim_{\epsilon \downarrow 0} \sum_{0 \leq u \leq t} \left\langle \Delta T(u), w_n^{-1} e_n \right\rangle 1_{|\Delta T(u)|_\infty \in (\epsilon, 1]} \\ &\leq \lim_{\epsilon \downarrow 0} \sum_{0 \leq u \leq t} |\Delta T(u)|_\infty 1_{|\Delta T(u)|_\infty \in (\epsilon, 1]}. \end{aligned}$$

Let $F_\epsilon : \ell_{1,w} \to \mathbb{R}$ be defined by $F(x) = |x|_\infty 1_{|x|_\infty \in (\epsilon, 1]}$, then F is obviously bounded on $\ell_{1,w}$ and belongs to $L^1(\ell_{1,w}, \mathcal{B}(\ell_{1,w}), \mathcal{T})$ by assumption (2.4.3). By Fatou's lemma,

$$\begin{aligned} \mathbb{E}|\widetilde{T}(t)|_\infty &\leq \liminf_{\epsilon \downarrow 0} \mathbb{E}\left[\sum_{0 \leq u \leq t} F_\epsilon(\Delta T(u)) \right] \\ &= \liminf_{\epsilon \downarrow 0} t \int_{|x|_\infty \in (\epsilon, 1]} |x|_\infty \mathcal{T}(dx) = t \int_{|x|_\infty \in (0, 1]} |x|_\infty \mathcal{T}(dx) < \infty, \end{aligned}$$

where the last equality follows by monotone convergence. Therefore $\widetilde{T}(t)$ is a bounded sequence almost surely and the proof is complete. □

2.5 Lévy measures on direct sums of Banach spaces

In order to define the weak subordination of X and T on $H \oplus \ell^+_{1,w}$, we need a characterisation of Lévy measure on the direct sum of spaces with different Rademacher types. This characterisation is given in Theorem 2.18.

Let $(E_1, |\cdot|_{E_1})$ and $(E_2, |\cdot|_{E_2})$ be separable Banach spaces of Rademacher type p_i and topological dual E_i^*, $i = 1, 2$. Then their direct sum $E_1 \oplus E_2$ is again a Banach space under any of the equivalent norms $|(x, y)|_p := (|x|^p_{E_1} + |y|^p_{E_2})^{1/p}$, $p \geq 1$. For simplicity, we will use $|(x, y)| := |(x, y)|_1 = |x|_{E_1} + |y|_{E_2}$. Furthermore $E_1 \oplus E_2$ is of type $p_1 \wedge p_2$. We endow $E_1 \oplus E_2$ with its Borel σ-algebra generated by any one of these equivalent norms.

For $x \in E_1$, $y \in E_2$, the maps $\pi_1 : E_1 \oplus E_2 \to E_1 \oplus \{0\}$ defined by $\pi_1(x, y) := (x, 0)$ and $\pi_2 : E_1 \oplus E_2 \to \{0\} \oplus E_2$ defined by $\pi_2(x, y) := (0, y)$ are bounded linear projections of norm 1. In particular, the function $(x, y) \mapsto |x|^{p_1}_{E_1} + |y|^{p_2}_{E_2}$ is Borel measurable.

2.5.1 Characterisation of Lévy measures

Now, let $\mathcal{Z}$ be a σ-finite measure on the Borel σ-algebra of a separable Banach space E. Recall K from 2.2.1. From Definition 2.3 we recall that $\mathcal{Z}$ is a Lévy measure if and only if $\int_E |K(x, u)|\mathcal{Z}(dx) < \infty$ for all $u \in E'$ and the function $u \mapsto \exp(\int_E K(x, u)\mathcal{Z}(dx))$ defines the characteristic function of a probability measure on E.

In practice, proving $\mathcal{Z}$ is a Lévy measure often requires uses of the Poisson exponent measure. We recall that for a finite measure μ on a separable Banach space E, the Poisson exponent measure $e(\mu)$ generated by μ (see [9]) is defined as

$$e(\mu) := e^{-\mu(E)} \sum_{k=0}^{\infty} \frac{\mu^{*k}}{k!}, \tag{2.5.1}$$

where μ^{*n} is the n-th convolution of μ and $\mu^{*0} := \delta_0$ is the Dirac measure at zero. It is known (see [114]) that $e(\mu)$ is an infinitely divisible Radon probability measure with

$$\widehat{e(\mu)}(u) = \exp\Big(\int_E (e^{i\langle x,u\rangle} - 1)\mu(dx)\Big).$$

The centering constant $x(\mu)$ of the measure μ is defined as the Bochner integral

$$x(\mu) := -\int_{|x|\leq 1} x\mu(dx).$$

The shifted Poisson probability measure $e_s(\mu)$ generated by μ is defined to be $e_s(\mu) :=$

$e(\mu) * \delta_{x(\mu)}$, whose Fourier transform is given by

$$\widehat{e_s(\mu)}(u) = \exp\Big(\int_E K(x,u)\mu(dx)\Big),$$

where recall $K(x,u) = e^{i\langle x,u\rangle} - 1 - i\langle x,u\rangle 1_B(x)$. Now suppose μ is only σ-finite instead of finite, in this case we will abuse some notation and still write $\widehat{e_s(\mu)}$ for the above expression whenever the integral in the exponent is well defined.

We are now ready to give a sufficient condition for a σ-finite measure $\mathcal{Z}$ to be a Lévy measure on $E_1 \oplus E_2$. Define the function

$$K(x,y,u,v) := e^{i\langle (x,y),(u,v)\rangle} - 1 - i\langle (x,y),(u,v)\rangle 1_{|(x,y)|\le 1}$$

for $(x,y) \in E_1 \oplus E_2$ and $(u,v) \in (E_1 \oplus E_2)^*$.

Theorem 2.18. *Let $\mathcal{Z}$ be a σ-finite measure on $E_1 \oplus E_2$ where E_1 and E_2 are of type p_1 and p_2 respectively. Then the integrability condition*

$$\int_{E_1\oplus E_2} 1 \wedge \Big(|x|_{E_1}^{p_1} + |y|_{E_2}^{p_2}\Big)\mathcal{Z}(dx,dy) < \infty \tag{2.5.2}$$

is sufficient for $\mathcal{Z}$ to be a Lévy measure on $E_1 \oplus E_2$.

Condition (2.5.2) has many equivalent forms due to the fact that all ℓ_p-type norms on $E_1 \oplus E_2$ are equivalent. We will precede the proofs of Theorem 2.18 with one such condition and a few technical lemmas.

Proposition 2.19. *Condition (2.5.2) is equivalent to*

$$\int_{E_1\oplus E_2} \Big[1_{|(x,y)|>1} + \Big(|x|_{E_1}^{p_1} + |y|_{E_2}^{p_2}\Big)1_{|(x,y)|\le 1}\Big]\mathcal{Z}(dx,dy) < \infty. \tag{2.5.3}$$

Proof. We will require the following three lemmas.

Lemma 2.20. *Let $x,y \in (0,\infty)$ and $p,q \ge 1$. Then the following holds.*

a) $\{x+y > 2\delta\} \subseteq \{x^p + y^q > \delta^p \wedge \delta^q\}$ *for all* $\delta > 0$.

b) $\{x^p + y^q > 1\} \subseteq \{x+y > 1\} \subseteq \{x^p + y^q > 2^{-p\vee q}\}$.

c) $\{x+y \le 1\} \subseteq \{x^p + y^q \le 1\} \subseteq \{x+y \le 2\}$.

Proof. (a) Since $p,q \ge 1$, the functions $x \mapsto x^p$ and $y \mapsto y^q$ are increasing. We first suppose $x \le y$; the case where $y \le x$ can be handled in the same way. Clearly the inequality $x+y > 2\delta$ implies $x\vee y > \delta$. Since $x \le y$, we have $(x\vee y)^p \wedge (x\vee y)^q = y^p \wedge y^q \le y^q \le x^p \vee y^q$. Therefore, using $x \vee y > \delta$ we obtain $x^p + y^q \ge x^p \vee y^q \ge \delta^p \wedge \delta^q$.

(b) Suppose $x^p + y^q > 1$. If both $x^p \le 1$ and $y^q \le 1$ then $x \ge x^p$ and $y \ge y^q$ since $p, q \ge 1$, so $x + y \ge x^p + y^q > 1$ and the first inclusion holds. If on the other hand $x^p > 1$ (resp. $y^q > 1$) then $x > 1$ (resp. $y > 1$) so $x + y > 1$ as well. The second inclusion follows from (a). (c) follows from (a) and (b). $\square$

Lemma 2.21. *Condition (2.5.3) implies $\mathcal{Z}(|(x, y)| > \delta) < \infty$ for all $\delta > 0$.*

Proof. The claim is trivial for $\delta > 1$. For $\delta \le 1$, let $\delta' := \left(\frac{\delta}{2}\right)^{p_1} \wedge \left(\frac{\delta}{2}\right)^{p_2} < 1$. By Lemma 2.20 (a) we have $1_{|x|+|y|>\delta} \le 1_{|x|^p+|y|^q>\delta'}$ and therefore

$$\mathcal{Z}(|(x, y)| \in (\delta, 1]) \le \frac{1}{\delta'} \int (|x|^{p_1} + |y|^{p_2}) 1_{|x|^{p_1}+|y|^{p_2}>\delta'} 1_{|(x,y)|\le 1} d\mathcal{Z},$$

which is finite by condition (2.5.3). $\square$

To prove Proposition 2.19, It is enough to bound the integrand in (2.5.3) with the integrand of (2.5.2) and vice versa.
(2.5.2) $\Rightarrow$ (2.5.3). Lemma 2.20 implies the inequalities $1_{|x|+|y|\le 1} \le 1_{|x|^{p_1}+|y|^{p_2}\le 1}$ and $1_{|x|+|y|>1} \le 1_{|x|^{p_1}+|y|^{p_2}>1} + 1_{|x|^{p_1}+|y|^{p_2}\in(\epsilon,1]}$ where $\epsilon := 2^{-p_1\vee p_2} < 1$. It remains to bound the integral of the second term by

$$\mathcal{Z}(|x|^{p_1} + |y|^{p_2} \in (\epsilon, 1]) \le \epsilon^{-1} \int (|x|^{p_1} + |y|^{p_2}) 1_{|x|^{p_1}+|y|^{p_2}\in(\epsilon,1]} d\mathcal{Z}.$$

(2.5.3) $\Rightarrow$ (2.5.2). Lemma 2.20 implies the inequalities $1_{|x|^{p_1}+|y|^{p_2}>1} \le 1_{|x|+|y|>1}$ and $1_{|x|^{p_1}+|y|^{p_2}\le 1} \le 1_{|x|+|y|\le 2}$, which gives the bound

$$\begin{aligned}(|x|^{p_1} + |y|^{p_2}) 1_{|x|^{p_1}+|x|^{p_2}\le 1} &\le (|x|^{p_1} + |y|^{p_2}) 1_{|x|+|x|\le 2} \\ &\le (2^{p_1} + 2^{p_2}) 1_{|x|+|y|\in(1,2]} + (|x|^{p_1} + |y|^{p_2}) 1_{|x|+|y|\le 1},\end{aligned}$$

since $|x| + |y| \le 2$ implies $|x|^{p_1} + |y|^{p_2} \le 2^{p_1} + 2^{p_2}$. $\square$

Lemma 2.22. *Suppose μ is a finite, symmetric Borel measure on $E_1 \oplus E_2$ with bounded support, i.e. there exists $\delta > 0$ such that $\mu(|(x, y)| > \delta) = 0$. Then for $i = 1, 2$, we have*

$$\int_{E_1\oplus E_2} |\pi_i(x, y)|^{p_i} \; e_s(\mu)(dx, dy) \le K_{p_i} \int_{E_1\oplus E_2} |\pi_i(x, y)|^{p_i} \mu(dx, dy),$$

where K_{p_i} are the type constants appearing in Theorem 2.2.

Proof. Since μ is a finite measure with bounded support, it is clear that $e_s(\mu)$ is a probability measure with finite moments of all orders. Let $|\mu| := |\mu(E_1 \oplus E_2)|$ and assume without loss of generality that $|\mu| \ne 0$. Recalling $\mu^{*0} := \delta_0$, we have

$$\int |\pi_i(x, y)|^{p_i} \; e_s(\mu)(dx, dy) = \sum_{n=0}^{\infty} e^{-|\mu|} \frac{|\mu|^n}{n!} \int |\pi_1(x, y)|^{p_i} (\mu/|\mu|)^{*n}(dx, dy)$$

$$= \sum_{n=1}^{\infty} e^{-|\mu|} \frac{|\mu|^n}{n!} \, \mathbb{E}\Big|\pi_1 \sum_{j=1}^{n} W_j\Big|^{p_i},$$

where W_j are independent, identically distributed random variables with distribution $\mu/|\mu|$. Furthermore each $\pi_i W_j$ is symmetric and takes values in a type p_i subspace of $E_1 \oplus E_2$. Therefore by Theorem 2.2 we have

$$\sum_{n=1}^{\infty} e^{-|\mu|} \frac{|\mu|^n}{n!} \, \mathbb{E}\Big|\pi_1 \sum_{j=1}^{n} W_j\Big|^{p_i} \leq K_{p_i} \sum_{n=1}^{\infty} e^{-|\mu|} \frac{|\mu|^n}{n!} \, n\mathbb{E}|\pi_1 W_1|^{p_i}$$
$$= K_{p_i} \left(\sum_{n=1}^{\infty} e^{-|\mu|} \frac{|\mu|^{n-1}}{(n-1)!} \right) \int |\pi_1(x,y)|^{p_i} \mu(dx,dy),$$

which completes the claim since the series in the last line sums up to 1. □

Proof of Theorem 2.18. Define the measure $\mathcal{Z}^-(E) := \mathcal{Z}(-E)$ so that $\mathcal{Z}^- + \mathcal{Z}$ is a symmetric, σ-finite measure on $E_1 \oplus E_2$. Since $\mathcal{Z}$ is a Lévy measure if and only if $\mathcal{Z} + \mathcal{Z}^-$ is ([114], p.70), replacing $\mathcal{Z}$ with $\mathcal{Z} + \mathcal{Z}^-$ if necessary we can assume $\mathcal{Z}$ is symmetric.

Let $\mathcal{Z}_1 := \mathcal{Z}|_{|x,y|>1}$ and $\mathcal{Z}_k := \mathcal{Z}|_{|(x,y)| \in (\frac{1}{k}, \frac{1}{k-1}]}$ for $k \geq 2$. By Proposition 2.19 it is clear that each $\mathcal{Z}_k$ is a finite measure, so by Proposition 5.3.1 of [114] $e_s(\mathcal{Z}_k)$ is the distribution of a random variable, say Y_k. Furthermore, since the supports of $(\mathcal{Z}_k)$ are disjoint Borel sets, by Proposition 5.3.2 we see that Y_k's are mutually independent and $e_s(\mathcal{Z}_l + \mathcal{Z}_m)$ is the distribution of $Y_l + Y_m$.

Note that $1 \wedge |(x,y)|^2 \lesssim 1 \wedge (|x|_{E_1}^2 + |y|_{E_2}^2) \leq 1 \wedge (|x|_{E_1}^{p_1} + |y|_{E_2}^{p_2})$, which is integrable with respect to $\mathcal{Z}$ by (2.5.2). Therefore the integral $\int K(x,y,u,v) d\mathcal{Z}$ is well-defined (see pg.71, Linde [114]). Then by the dominated convergence theorem, we have

$$\int_{E_1 \oplus E_2} K(x,y,u,v) \mathcal{Z}(dx,dy) = \lim_{n \to \infty} \sum_{k=1}^{n} \int_{E_1 \oplus E_2} K(x,y,u,v) \mathcal{Z}_k(dx,dy).$$

Taking exponentials of both sides gives us the expression

$$\widehat{e_s(\mathcal{Z})}(u,v) = \lim_{n \to \infty} \prod_{k=1}^{n} \widehat{e_s(\mathcal{Z}_k)}(u,v) = \lim_{n \to \infty} \mathbb{E}\left[\exp i \left\langle \sum_{k=1}^{n} Y_k, (u,v) \right\rangle \right].$$

From this it is clear that $\widehat{e_s(\mathcal{Z})}(u,v)$ is the c.f. of a probability measure, i.e. $\mathcal{Z}$ is a Lévy measure, provided that the series $\sum_{k=1}^{n} Y_k$ converges in distribution to a random variable. It is therefore sufficient to show that both $\sum_{k=1}^{n} \pi_1 Y_k$ and $\sum_{k=1}^{n} \pi_2 Y_k$ converge in distribution as $n \to \infty$.

For $2 \leq l \leq m$, the law of $\sum_{k=l}^{m} Y_k$ is the shifted Poisson probability measure $e_s(\sum_{k=l}^{m} \mathcal{Z}_k)$, where $\sum_{k=l}^{m} \mathcal{Z}_k = \mathcal{Z}|_{|x,y| \in (\frac{1}{m}, \frac{1}{l-1}]}$ is a measure satisfying the assumptions of Lemma 2.22. Therefore for $i = 1,2$ we have

$$\mathbb{E}\Big|\pi_i \sum_{k=l}^{m} Y_k\Big|^{p_i} = \int_{E_1\oplus E_2} |\pi_i(x,y)|^{p_i}\ e_s(\mathcal{Z}|_{|x,y|\in(\frac{1}{m},\frac{1}{l-1}]})(dx,dy)$$
$$\leq K_{p_i} \int_{|x,y|\in(\frac{1}{m},\frac{1}{l-1}]} |\pi_i(x,y)|^{p_i}\mathcal{Z}(dx,dy),$$

which converges to zero as $l, m \to \infty$ by (2.5.3) and the dominated convergence theorem. Hence $\pi_i \sum_{k=1}^{n} Y_k$ converges in probability and so does $\sum_{k=1}^{n} Y_k$. □

2.6 Weak subordination

We first give a formal definition of the weak subordination in Definition 2.23, which is a direct generalization of [49]. In Theorem 2.24 we establish the existence of the weak subordination as defined in Definition 2.23, which shows that the construction of [49] can indeed be extended to an infinite dimensional setting. The proof of Theorem 2.24 boils down to showing a certain measure defined on the direct sum of Banach spaces is a Lévy measure; this is the content of Theorem 2.26.

Suppose X is a Lévy process on a separable Hilbert space H with characteristic triplet $(\gamma, Q, \mathcal{X})$. Let $(T(t))_{t\geq 0}$ be a subordinator taking values in ℓ_∞^+ as defined in Section 2.4 , i.e. T is an subordinator on $\ell_{1,w}^+$ with drift $\eta \in \ell_\infty^+$ and Lévy measure $\mathcal{T}$ which is concentrated on ℓ_∞^+ and satisfies the integrability condition

$$\int_{\ell_\infty^+} 1 \wedge |\tau|_\infty \mathcal{T}(d\tau) < \infty. \tag{2.6.1}$$

Analogous to [49], we define the the weak subordination $(T, X \odot T)$ of X and T by specifying its characteristic triplet.

Definition 2.23. A Lévy process $Z(t) = (Z_1(t), Z_2(t))_{t\geq 0}$ taking values on the Banach space $\ell_\infty^+ \oplus H$ is called the weak subordination of X and T, which we write as $Z \stackrel{\mathcal{D}}{=} (T, X\odot T)$, if its characteristic triplet $(m, \Theta, \mathcal{Z})$ is given by

$$m = (m_1, m_2), \quad \Theta = 0 \oplus (\eta \diamond Q),$$
$$\mathcal{Z}(d\tau, dx) = \delta_0(d\tau)(\eta \diamond \mathcal{X})(dx) + \mathcal{T}(d\tau)\mathbb{P}(\mathbf{X}(\tau) \in dx), \tag{2.6.2}$$

where $m = (m_1, m_2)$ is given by the Bochner integrals

$$m_1 = \eta + \int_{\ell_{1,w}^+} \tau\mathbb{P}(|(\tau, \mathbf{X}(\tau))| \leq 1)\ \mathcal{T}(d\tau), \tag{2.6.3}$$

$$m_2 = \eta \diamond \gamma + C_{\eta\diamond\mathcal{X}} + \int_{\ell_{1,w}^+} \mathbb{E}[\mathbf{X}(\tau)1_{|(\tau,\mathbf{X}(\tau))|\leq 1}]\ \mathcal{T}(d\tau), \tag{2.6.4}$$

and $0 \oplus (\eta \diamond Q)$ is the direct sum of the zero operator on ℓ_∞ and $\eta \diamond Q$ on H, i.e. $(0\oplus(\eta\diamond Q))(\tau, x) = (0, (\eta\diamond Q)x)$ for any $(\tau, x) \in \ell_\infty^+\oplus H$. If in addition Z_1 is indistinguishable from T, the Lévy process Z is called the semi-strong subordination of X and T.

We remark that Definition 2.23 can seem quite involved upon first glance, and it is not at all clear that it captures the concept of weak subordination we have been describing so far. To unravel this definition, we recall the intuitive construction of the weak subordination we presented in Section 2.1.2 and explain how it matches up with the above definition.

Recall from Theorem 2.15 that T can be decomposed into $T(t) = t\eta + S(t)$, where $\eta \in \ell_\infty^+$ and S is a pure jump subordinator on ℓ_∞ with Lévy measure $\mathcal{T}$. Recall from Section 2.1.2 that to define the weak subordination, it suffices to consider the two cases $T(t) = t\eta$ and $T(t) = S(t)$ separately and define the process $X \odot T$ in each case. The weak subordination in the general case is then obtained by convolving the distributions of the weak subordinated process defined in each case.

When $T(t) = t\eta$, the weak subordination $X \odot T$ is defined to be the Lévy process whose marginal distributions are equal to those of $X \circ T$. In this case, from Theorem 2.10 we can conclude that the characteristic triplet of the process $X \odot T$ is given by $(\eta \diamond \gamma + C_{\eta\diamond\mathcal{X}}, \eta \diamond Q, \eta \diamond \mathcal{X})$. Moreover, since T is just a deterministic drift, it is clear that the triplet of the pair of processes $(T, X \odot T)$ is given by

$$\Big((\eta, \eta \diamond \gamma),\ 0 \oplus (\eta \diamond Q),\ \delta_0 \otimes (\eta \diamond \mathcal{X})\Big), \tag{2.6.5}$$

where $\delta_0 \otimes (\eta \diamond \mathcal{X})$ is the product measure on $\ell_\infty \oplus H$ formed by the Dirac measure δ_0 on ℓ_∞ and the Lévy measure $\eta \diamond \mathcal{X}$ on H.

On the other hand, when $T = S$, recall that $X \odot T$ is defined to be the (pure jump) Lévy process whose jump at time t, after conditioning on $\Delta T(t)$, is equal in distribution to the random variable $X(\Delta T(t))$. The Lévy measure of $(T, X \odot T)$ is thus given by

$$\begin{aligned}\mathbb{E}\Big[\sum_{t\le 1} 1_{\Delta T(t)\in d\tau} 1_{\Delta(X\odot T)(t)\in dx}\Big] &= \mathbb{E}\Big[\mathbb{E}\Big[\sum_{t\le 1} 1_{\Delta T(t)\in d\tau} 1_{X(\Delta T(t))\in dx} | \Delta T\Big]\Big] \\ &= \mathbb{E}\Big[\sum_{t\le 1} 1_{\Delta T(t)} \mathbb{P}(X(\Delta T(t)) \in dx) \in d\tau\Big] = \mathcal{T}(d\tau)\mathbb{P}(X(\tau) \in dx).\end{aligned}$$

Since $X \odot T$ and T are pure jump processes, the covariance operator in the triplet of $(T, X \odot T)$ is zero. To complete the triplet of $(T, X \odot T)$, it remains to compute the compensation term for the Lévy measure of $(T, X \odot T)$, which can be shown to be

$$\left(\int_{\ell_{1,w}^+} \tau \mathbb{P}(|(\tau, \mathbf{X}(\tau))| \le 1)\ \mathcal{T}(d\tau),\ \int_{\ell_{1,w}^+} \mathbb{E}[\mathbf{X}(\tau) 1_{|(\tau,\mathbf{X}(\tau))|\le 1}]\ \mathcal{T}(d\tau)\right) \in \ell_{1,w}^+ \oplus H.$$

We have therefore obtained the characteristic triplets of the weak subordination in the case $T(t) = t\eta$ and the case $T(t) = S(t)$. We recall that for the general case, the weak subordination is defined by setting its distribution to the convolution of distributions of $X \odot T$ in the two cases discussed above. This immediately implies that the characteristic triplet of $X \odot T$ in the general case is the sum of the characteristic triplets in each case. By summing up the triplets we described above we finally arrive at Definition 2.23.

We now present the main result of our work, which establishes the existence of the weak subordination as defined in Definition 2.23.

Theorem 2.24. *Let X be a Lévy process taking values on a separable Hilbert space and T be a subordinator taking values on the positive cone ℓ^+_∞. Then there exists a unique (in distribution) Lévy process Z taking values on $\ell^+_\infty \oplus H$ satisfying Definition 2.23.*

Since the weak subordination is defined via distribution, the existence of the weak subordination is guaranteed as long as $(m, \Theta, \mathcal{Z})$ is a valid characteristic triplet on $\ell^+_\infty \oplus H$. The semi-strong subordination is then always possible on a possibly augmented probability space that carries the process T, see Theorem 2.1 (ii) of [49].

From Proposition 2.9 it is clear that $\tau \diamond \gamma$ and $C_{d\diamond\tau}$ are elements of H and Θ is indeed a covariance operator on $\ell^+_{1,w} \oplus H$. The Bochner integrals in (2.6.3) are well-defined by (2.4.1) and the Bochner integrals in (2.6.4) are well-defined by (2.3.9) and (2.6.1). Therefore it remains to check that (2.6.2) defines a Lévy measure. We first give a preliminary result.

Proposition 2.25. *Expression (2.6.2) defines a σ-finite measure on $\ell^+_{1,w} \oplus H$. Furthermore the second term $\mathcal{T}(d\tau)\mathbb{P}(\mathbf{X}(\tau) \in dx)$ of (2.6.2) satisfies*

$$\int_{\ell^+_{1,w} \oplus H} f(\tau, x)\mathcal{T}(d\tau)\mathbb{P}(\mathbf{X}(\tau) \in dx) = \int_{\ell^+_{1,w}} \int_H f(\tau, x)\mathbb{P}(\mathbf{X}(\tau) \in dx)\mathcal{T}(d\tau)$$

for non-negative, measurable functions $f : \ell^+_{1,w} \oplus H \to \mathbb{R}$.

Proof. The proof can be adapted from the arguments on page 28 of [119]. We sketch the proof here for the reader's convenience. Let

$$\mathcal{Z}_0(d\tau, dx) := (\eta \diamond \mathcal{X})(dx)\delta_0(d\tau), \quad \mathcal{Z}_1(dx, d\tau) := \mathbb{P}(\mathbf{X}(\tau) \in dx)\mathcal{T}(d\tau).$$

Note that $\mathcal{Z}_0$ is simply the product measure on $\ell^+_{1,w} \oplus H$ and is therefore σ-finite since $\eta \diamond \mathcal{X}$ is σ-finite and δ_0 is finite.

By Corollary 2.12, we have $\mathbf{X}(\tau^k) \to \mathbf{X}(\tau)$ in probability whenever $\tau^k \to \tau$ in $\ell^+_{1,w}$ as $k \to \infty$. Let U be an arbitrary open set in H. By the Portmanteau theorem of weak convergence (see [35]) we have

$$\liminf_k \mathbb{P}(\mathbf{X}(\tau^k) \in U) \geq \mathbb{P}(\mathbf{X}(\tau) \in U)$$

which implies that $\tau \mapsto \mathbb{P}(\mathbf{X}(\tau) \in A)$ is lower semi-continuous and hence Borel measurable for all $\tau \in \ell^+_{1,w}$. By Lemma 1.40 in [94], $\mathbb{P}(\mathbf{X}(\tau) \in dx)$ is a Markov kernel from $\ell^+_{1,w}$ to H. Applying Theorem 6.11 of [56] gives the claimed result. $\square$

To prove Theorem 2.24, it remains to show the following.

Theorem 2.26. *Expression* (2.6.2) *defines a Lévy measure on* $\ell_{1,w}^{+} \oplus H$.

Proof. Since $\mathcal{Z}_0$ is a σ-finite product measure, by Fubini's theorem

$$\int_{\ell_{1,w}^{+} \oplus H} 1 \wedge (|x|^2 + |\tau|)\mathcal{Z}_0(d\tau, dx) = \int_H 1 \wedge |x|^2 (\eta \diamond \mathcal{X})(dx) < \infty$$

so $\mathcal{Z}_0$ is a Lévy measure by Theorem 2.18. It remains to show $\mathcal{Z}_1$ is a Lévy measure as well. Note that $|\tau|_{1,w} \leq |\tau|_\infty$ and $1_{|\tau|_{1,w}+|x|^2>1} \leq 1_{|\tau|_\infty+|x|^2>1}$. Then the set $\{|\tau|_\infty + |x|^2 > 1\}$ can be written as the disjoint union

$$\{|\tau|_\infty + |x|^2 > 1\} = \{|\tau|_\infty > 1, x \in H\} \cup \{|\tau|_\infty \leq 1, |x|^2 > 1 - |\tau|_\infty\}.$$

For the first set, we have

$$\mathcal{Z}_1(|\tau|_\infty > 1) = \int_{|\tau|_\infty > 1} \mathbb{P}(|\mathbf{X}(\tau)| \in H)\mathcal{T}(d\tau) = \mathcal{T}(|\tau|_\infty > 1) < \infty$$

since $\mathcal{T}$ satisfies (2.6.1). For the second set, by Lemma 2.11 we have

$$\begin{aligned}\mathcal{Z}_1(|\tau|_\infty \leq 1, |x|^2 > 1 - |\tau|_\infty) &= \int_{|\tau|_\infty \leq 1} \mathbb{P}\Big(|\mathbf{X}(\tau)|^2 > 1 - |\tau|_\infty\Big)\mathcal{T}(d\tau) \\ &\lesssim \int_{|\tau|_\infty \leq 1} |\tau|_\infty \mathcal{T}(d\tau).\end{aligned}$$

which is finite by (2.6.1). This shows $\int 1_{|\tau|_\infty + |x| > 1} d\mathcal{Z}_1 < \infty$. On the other hand, since $1_{|x|^2+|\tau|_\infty \leq 1} \leq 1_{|x|\leq 1}1_{|\tau|_\infty \leq 1}$, by Lemma 2.11 we have

$$\begin{aligned}\int (|x|^2+|\tau|_\infty)1_{|x|^2+|\tau|_\infty \leq 1} d\mathcal{Z}_1 &\leq \int_{|\tau|_\infty \leq 1} \int (|x|^2 + |\tau|_\infty) 1_{|x|\leq 1} d\mathcal{Z}_1 \\ &= \int_{|\tau|_\infty \leq 1} \mathbb{E}\Big[|\mathbf{X}(\tau)^2| 1_{|\mathbf{X}(\tau)|\leq 1}\Big]\mathcal{T}(d\tau) + \int_{|\tau|_\infty \leq 1} |\tau|_\infty \mathbb{P}(|\mathbf{X}(\tau)| \leq 1)\mathcal{T}(d\tau) \\ &\lesssim \int_{|\tau|_\infty \leq 1} (|\tau|_\infty + |\tau|_\infty^2)\mathcal{T}(d\tau)\end{aligned}$$

which is finite by (2.6.1) and the fact that $|\tau|_\infty^2 \leq |\tau|_\infty$ on the set $\{|\tau|_\infty \leq 1\}$. Therefore by Theorem 2.18 we conclude that $\mathcal{Z}$ is a Lévy measure. $\square$

So far we have defined the weak subordination as a pair of processes. To conclude our work we will state the marginal distributions of the weak subordination, which can be compared to the strong subordination $X \circ T$. We first state some elementary properties of the weak subordination $(T, X \odot T)$ in terms of the characteristics of T and X. We shall use the same notation $\langle \cdot, \cdot \rangle$ for the duality pairing between $\ell_{1,w}^{+}$ and ℓ_∞, the inner product of H, and the duality pairing of $\ell_{1,w}^{+} \oplus H$ and $\ell_\infty \oplus H \simeq (\ell_{1,w}^{+} \oplus H)^*$.

Firstly, from Definition 2.23, we immediately obtain the characteristic function of the weak subordination $Z \overset{\mathcal{D}}{=} (T, X \odot T)$ of X and T.

Proposition 2.27. *A Lévy process Z on $\ell_{1,w}^+ \oplus H$ is the weak subordination of X and T iff the characteristic exponent Ψ_Z of Z is given by*

$$\Psi_Z((\alpha, u)) = i\langle \eta, \alpha\rangle + \Psi_{X(\eta)}(u) + \int_{\ell_\infty^+} (e^{\Psi_{(\tau,\mathbf{X}(\tau))}(\alpha,u)} - 1)\mathcal{T}(d\tau),$$

for $\alpha \in \ell_\infty$ and $u \in H$, where $\Psi_{X(\eta)}$ is the characteristic exponent of $X(\eta)$.

Proof. Let $\alpha \in \ell_\infty$ and $u \in H$. From Definition 2.23 and the Lévy-Khintchine formula, we see that $Z \stackrel{\mathcal{D}}{=} (T, X \odot T)$ if and only if

$$\begin{aligned}\Psi_Z((u, v)) &= i\langle m_1, \alpha\rangle + i\langle m_2, u\rangle - \frac{1}{2}\langle u, (\eta \diamond Q)u\rangle + \int_H K(x, u)(\eta \diamond \mathcal{X})(dx) \\ &+ \int_{\ell_\infty^+ \oplus H} \Big[e^{i\langle(\tau,x),(\alpha,u)\rangle} - 1 - i\langle(\tau, x), (\alpha, u)\rangle 1_{|\tau|+|x|\le 1}\Big]\mathbb{P}(\mathbf{X}(\tau) \in dx)\mathcal{T}(d\tau),\end{aligned}$$

where the first two terms are given by

$$\begin{aligned}\langle m_1, \alpha\rangle &= \langle \eta, \alpha\rangle + \int_{\ell_\infty^+} \langle \tau, \alpha\rangle \mathbb{P}(|\tau| + |\mathbf{X}(\tau)| \le 1)\mathcal{T}(d\tau), \\ \langle m_2, u\rangle &= \langle \eta \diamond \gamma + C_{\eta\diamond\mathcal{X}}, u\rangle + \int_{\ell_\infty^+} \mathbb{E}[\langle \mathbf{X}(\tau), u\rangle 1_{|\tau|+|\mathbf{X}(\tau)|\le 1}]\mathcal{T}(d\tau),\end{aligned}$$

and the last integral in the expression for Ψ_Z simplifies to

$$\begin{aligned}&\int_{\ell_\infty^+ \oplus H} \Big[e^{i\langle(\tau,x),(\alpha,u)\rangle} - 1 - i\langle(\tau, x), (\alpha, u)\rangle 1_{|\tau|+|x|\le 1}\Big]\mathbb{P}(\mathbf{X}(\tau) \in dx)\mathcal{T}(d\tau) \\ &= \int_{\ell_\infty^+} \mathbb{E}[e^{i\langle(\tau,\mathbf{X}(\tau)),(\alpha,u)\rangle} - 1]\mathcal{T}(d\tau) - i\int_{\ell_\infty^+} \langle \tau, \alpha\rangle \mathbb{P}(|\tau| + |\mathbf{X}(\tau)| \le 1)\mathcal{T}(d\tau) \\ &\quad - i\int_{\ell_\infty^+} \mathbb{E}[\langle \mathbf{X}(\tau), u\rangle 1_{|\tau|+|\mathbf{X}(\tau)|\le 1}]\mathcal{T}(d\tau).\end{aligned}$$

The finiteness of all terms above follows directly from Lemma 2.11. The claim follows from collecting the above terms and Theorem 2.10. □

Finally, using Proposition 2.27 and setting either $u = 0$ or $\alpha = 0$, we can easily state the distributions of the marginal process Z_1 and Z_2:

Corollary 2.28. *Suppose $Z = (Z_1, Z_2) \stackrel{\mathcal{D}}{=} (T, X \odot T)$. Then $Z_1 \stackrel{\mathcal{D}}{=} T$ and Z_2 is a Lévy process on H with characteristics $(\zeta, \eta \diamond Q, \mathcal{Z})$,*

$$\begin{aligned}\zeta &= \eta \diamond \gamma + C_{\eta\diamond\mathcal{X}} + \int_{\ell_\infty^+} \mathbb{E}[\mathbf{X}(\tau)1_{|\mathbf{X}(\tau)|\le 1}]\mathcal{T}(d\tau), \\ \mathcal{Z}(dx) &= (\eta \diamond \mathcal{X})(dx) + \int_{\ell_\infty^+} \mathbb{P}(\mathbf{X}(\tau) \in dx)\mathcal{T}(d\tau).\end{aligned}$$

2.7 Conclusion and Future Works

In this chapter we have extended the construction of the weak subordination in [49] to an infinite dimensional setting. We have shown that for an arbitrary Lévy process X on a separable Hilbert space and a uniformly bounded sequence of subordinators T, the weak subordination $(T, X \odot T)$ as defined in (2.23) always exists as a Lévy process on $\ell_\infty^+ \oplus H$.

The next natural step for developing the theory of subordination is of course to construct concrete examples of it and analyse their properties. For instance, we recall that [107] defines an infinite dimensional α-stable noise using a cylindrical Wiener process subordinated by a one dimensional stable process. Using weak subordination and a multivariate stable subordinator, we can construct a generalized version of the α-stable noise. Towards this direction, we have some preliminary results suggesting that the weakly subordinated process can exhibit very different behaviours.

Another natural class of processes to consider is the WVαG processes we discussed in Section 1.2. An infinite dimensional version of it could be used as a high dimensional model for a large number of assets. Besides its uses in mathematical finance, this class of processes has some interesting theoretical properties as well, see [50].

Besides concentrating on the weak subordination itself, we may consider SPDEs driven by weakly subordinated Lévy processes. Towards this direction we have initiated some preliminary work with Prof. Ben Goldys. The motivation is that by using weakly subordination we can define a rather large class of Lévy driving noises. At the same time, we aim to have some explicit control over this general class of processes in terms of the triplets of X and T.

Chapter 3

Continuous Time Delayed Lévy-Driven GARCH Processes

3.1 Introduction

To continue our discussions in Section 1.3, we first give a more detailed overview of the CDGARCH process and the main results of [65, 66, 160] in Section 3.1.1. The setting of our work will be described in Section 3.1.2.

3.1.1 The CDGARCH Process

We recall that the discrete time GARCH(P,Q) process is defined by the pair of equations

$$Y_n := Y_{n-1} + \sqrt{X_n} Z_n, \tag{3.1.1}$$

$$X_n := \eta + \sum_{k=1}^{P} \beta_k X_{n-k} + \sum_{k=1}^{Q} \alpha_k X_{n-k} Z_{n-k}^2, \quad n \in \mathbb{N}, \tag{3.1.2}$$

subject to initial conditions $X_n = x_n$ for $n \leq 0$, where $(x_n)_{n\leq 0}$ is a sequence of positive constants. The sequence (Z_n) is assumed to be standardized white noise, i.e. uncorrelated random variables with zero mean and unit variance. Here η is a positive constant, and $(\beta_i)_{1\leq i\leq P}$ and $(\alpha_i)_{1\leq i\leq Q}$ are sequences of non-negative constants. The process (X_n) is referred to as the conditional variance process, since $X_n = \mathrm{Var}(Y_n|\mathcal{F}_{n-1})$, where $(\mathcal{F}_n)_n$ is the natural filtration of the process $(Y_n, X_n)_n$.

We now illustrate the resemblance between the discrete time GARCH process and the CDGARCH processes of [160]. Note that from (3.1.2) we have

$$X_n - X_{n-1} = \eta + (\beta_1 - 1)X_{n-1} + \sum_{k=2}^{P} \beta_k X_{n-k} + \sum_{k=1}^{Q} \alpha_k X_{n-k} Z_{n-k}^2.$$

Writing $\widetilde{\beta}_k = \beta_k - \mathbb{1}_{\{k=1\}}$, we can rewrite equation (3.1.2) into

$$X_n - X_0 = \sum_{i=1}^{n}(X_i - X_{i-1}) = n\eta + \sum_{i=1}^{n}\sum_{k=1}^{P}\widetilde{\beta}_k X_{i-k} + \sum_{i=1}^{n}\sum_{k=1}^{Q}\alpha_k X_{i-k} Z_{i-k}^2.$$

Interchanging the double sums, we can rewrite the GARCH equations into

$$Y_n = Y_0 + \sum_{i=1}^{n}(Y_i - Y_{i-1}), = Y_0 + \sum_{i=1}^{n}\sqrt{X_i}Z_i \tag{3.1.3}$$

$$X_n = X_0 + n\eta + \sum_{k=-P}^{-1}\widetilde{\beta}_{-k}\sum_{i=1+k}^{n+k} X_i + \sum_{k=-Q}^{-1}\alpha_{-k}\sum_{i=1+k}^{n+k} X_i Z_i^2. \tag{3.1.4}$$

Intuitively, by replacing the sums in (3.1.4) by appropriate integrals, we obtain the CDGARCH equation (1.3.1). This intuition is made rigorous in [65, 66, 160], which shows that after embedding into continuous time, as the time between observations tends to zero, the solution to (3.1.4) indeed converges in a certain sense to the solution of

$$Y_t = Y_0 + \int_0^t \sqrt{X_{s-}}dL_s, \tag{3.1.5}$$

$$X_t = \theta_t + \int_{-p}^{0}\int_{u}^{t+u} X_s ds\mu(du) + \int_{-q}^{0}\int_{u+}^{t+u} X_{s-}d[L,L]_s\nu(du), \tag{3.1.6}$$

where L, θ are semimartingales and $[L, L]$ is the quadratic variation of L. Here μ and ν are signed Borel measures on $[-p, 0]$ and $[-q, 0]$ respectively. These measures arise as the limits in a certain sense of the coefficients $(\widetilde{\beta}_k)_{1\le k\le P}, (\alpha_k)_{1\le k\le Q}$ and capture the serial dependence of the conditional variance process. Making precise statements on the convergence of the coefficients $(\widetilde{\beta}_k)_k, (\alpha_k)_k$ to the measures μ, ν requires a lot of additional notations; we instead refer the reader to [160] for the details.

Interestingly, the sufficient conditions imposed on the coefficients $(\alpha_k)_k$ and the measures ν to obtain the convergence to (3.1.6) depend on the choice of driving noises Z and L. In particular, when L is assumed to be a continuous process, $(\alpha_k)_k$ is only required to converge to ν in a very weak sense, and the limit ν can be any signed Borel measure with finite variation. That is, when the driving noise L is a Brownian motion, there is very little restriction on the serial dependence structure specified by ν.

On the other hand, when the continuity assumption on L is dropped, $(\alpha_k)_k$ is required to converge in a much stronger sense and ν can only be a Dirac measure at zero. We remark that this requirement is not explicitly stated in [160] but follows from equation (7.8) and Assumption 7.2.1 therein. To reiterate, when L is taken to be a Lévy process, to establish the convergence of (3.1.4) to (3.1.6), the author [160] requires the measure ν to be a point mass at zero. This is equivalent to requiring $q = 0$ in which case the setting reduces to the simpler case of a CDGARCH$(p, 0)$ process.

Since the goal of our work is to study a Lévy driven CDGARCH(p, q) process as a generalization of the GARCH and COGARCH processes, we will need to overcome this issue in some way. We remark that while it might be possible to continue the work of

[160] and investigate whether the sufficient conditions posed in [160] are in fact necessary as well, we instead take a more direct approach. Observe that (3.1.6) still makes sense as a stand-alone stochastic integral equation, even if the sufficient conditions in [160] are not satisfied. In this case (3.1.6) might not be the limit of a sequence of approximating GARCH processes in the sense of [160], but clearly still resembles the structure of the GARCH process as can be seen from (3.1.2).

We therefore take equation (3.1.6) as a starting point of our analysis. Essentially, we are treating the CDGARCH(p,q) process defined by (3.1.6) as a stand-alone process of interest rather than a limit of discrete time GARCH processes. In doing so however, we can no longer rely on the results of [160] to obtain the existence and uniqueness of a solution to (3.1.6). Instead we have to establish the existence and uniqueness of a solution directly from (3.1.6) without the use of an approximating sequence of GARCH processes. We now give a formal description of the setting of our work.

3.1.2 Our Setting

Let $p,q>0$ and put $r := p \vee q$. Let $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t\geq -r}, \mathbb{P})$ be a filtered probability space that satisfies the usual assumptions (see [145], page 3). Let Ψ be a $\mathcal{F}_0$-measurable non-negative random variable and $(\Phi_u)_{u\in[-(p\vee q),0]}$ be a non-negative adapted process.

In this chapter, we will study the CDGARCH(p,q) equation

$$\begin{aligned}
Y_t &= Y_0 + \int_0^t \sqrt{X_{s-}}dL_s,\\
X_t &= \theta_t + \int_{-p}^0 \int_u^{t+u} X_s ds\mu(du) + \int_{-q}^0 \int_{u+}^{t+u} X_{s-}d[L,L]_s\nu(du), \quad t>0,
\end{aligned}$$

with initial conditions $Y_0 = \Psi$ and $X_u = \Phi_u$ for $u \in [-(p\vee q),0]$. We assume L is a Lévy process adapted to $\mathbb{F}$ and $[L,L]$ is the quadratic variation process of L.

Following the settings of [66, 65, 160], we recall that μ and ν are signed Borel measures supported on $[-p,0]$ and $[-q,0]$ respectively. Although in theory they could be chosen almost arbitrarily, in our work we assume that they have specific forms that resemble the GARCH process. The measures μ and ν are assumed to have point masses at zero and are absolutely continuous with respect to the Lebesgue measure on $[-r,0)$. We remark that these choices are in fact natural extensions of the constraints of the original GARCH process. Specifically, we assume that there exist positive constants c_μ, c_ν and nonnegative, continuous functions f_μ, f_ν supported on $[-p,0]$ and $[-q,0]$ respectively, such that for any Borel set $E \in \mathcal{B}([-r,0))$, we have

$$\mu(E) := \int_{E\cap[-p,0]} f_\mu(u)du - c_\mu\delta_0(E), \tag{3.1.7}$$

$$\nu(E) := \int_{E\cap[-q,0]} f_\nu(u)du + c_\nu\delta_0(E). \tag{3.1.8}$$

The stationarity of GARCH processes depends on the size of the coefficients, in fact, usual weak stationarity constraints effectively require $\widetilde{\beta}_1 < 0$, $\widetilde{\beta}_k \in (0,1)$ for $k \neq 1$ and $\alpha_k \in (0,1)$ for all k. Our assumptions on μ and ν, including the signs of the constants, are analogous to these constraints. It will be shown that analogous to the GARCH process, the stationarity of the CDGARCH process depends on the relative sizes of c_μ, c_ν, f_μ and f_ν. Furthermore, it is known that the GARCH process exhibits mean reverting behaviour. This feature is retained in our setup; it will be shown that the constant $-c_\mu$ captures the effect of mean reversion.

Finally, to specify the process $(\theta_t)_t$ in (3.1.6), we first introduce some notations. Let $\mathscr{D}_{[a,b]} := \mathscr{D}([a,b])$ (resp. $\mathbb{D}_{[a,b]}$) be the space of càdlàg functions (resp. processes) on $[a,b] \subseteq \mathbb{R}$ and write $\mathscr{D} := \mathscr{D}_{[-r,0]}$ and $\mathbb{D} := \mathbb{D}_{[-r,0]}$. Given an initial process $\Phi_\cdot \in \mathbb{D}$, we extend it to $\mathbb{D}_{[-r,\infty)}$ by setting $\Phi_t = \Phi_0$, for all $t > 0$. Fix a positive constant η. Throughout the paper we will assume θ takes the form

$$\theta_t := \Phi_t + \eta t \mathbb{1}_{[0,\infty)}(t), \quad t \in [-r,\infty). \tag{3.1.9}$$

We note that this choice is completely analogues to the $n\eta$ term in (3.1.4).

We finally remark that our setting is general enough to include the earlier models we discussed in Section 1.3. When $p = q = 0$, $c_\mu, c_\nu > 0$, and $\theta_t = X_0 + \eta t \mathbb{1}_{[0,\infty)}(t)$, equation (3.1.6) reduces to a stochastic differential equation

$$dX_t = \eta dt - c_\mu X_t dt + c_\nu X_{t-} d[L,L]_t, \quad t > 0, \tag{3.1.10}$$

which (after a reparameterization) is the SDE specifying the COGARCH process (see [100] Proposition 3.2). On the other hand, taking L to be a Brownian motion, it is possible to define a similar pair of SFDEs that generalizes Nelson's diffusion and Lorenz's limit.

The rest of the chapter will be organized as follows. Section 3.2 collects some preliminary material on Lévy processes and stochastic integrals. Section 3.3 establishes the existence, uniqueness and positivity of a solution to (3.1.6). Since (3.1.6) is rather difficult to work with, we derive a more convenient representation of the solution to (3.1.6). Using this we can study sample paths of the process, and characterize its jumps and mean reverting behaviour. Section 3.4 studies the second order behaviour of the CDGARCH process. We give conditions for the process to be stationary and derive an equation for the asymptotic mean and covariance function. The behaviour of the CDGARCH process is shown to be similar to the discrete time GARCH process. Finally, the proofs in this chapter are collected in Section 3.5.

3.2 Preliminaries

We first collect some preliminary results. We follow Jacod and Shiryaev [89], Protter [145] for semimartingale theory, Applebaum [6] for Lévy processes, and Diekmann, van Gils, Lunel, and Walther [62] for deterministic delay differential equations.

3.2.1 Driving Lévy Process

Let $r := p \vee q > 0$ and suppose we have a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t \geq -r}, \mathbb{P})$ that satisfies the "usual conditions" (see Definition 12, [89]). Given a stochastic process Z, we write $\sigma(Z) := (\sigma\{Z_u, u \leq t\})_{t \geq -r}$ for the natural filtration of Z.

Let $(M_t)_{t \geq -r}$ be a càdlàg, adapted martingale with respect to $\mathbb{F}$. We follow Protter [145] and call M a square integrable martingale if $\mathbb{E}[M_t^2] < \infty$ for every $t \geq -r$. For a process Z with finite second moments, i.e. $\mathbb{E}[Z_t^2] < \infty$ for all t, write $[Z] := [Z, Z]$ (resp. $\langle Z \rangle := \langle Z, Z \rangle$) for the quadratic variation (resp. predictable quadratic variation) process of Z. Let $L^2(Z)$ be the set of all predictable processes H such that the integral process $H^2 \cdot \langle Z \rangle$ is integrable, i.e. $\mathbb{E}[\int_{-r}^T H_s^2 d\langle Z \rangle] < \infty$ for each fixed T. The following lemma follows from Theorems I.4.31 - I.4.40 of Jacod and Shiryaev [89].

Lemma 3.1. *Let Z be a semimartingale and suppose H is càdlàg and predictable. Then the integral process $H \cdot Z$ is a càdlàg, adapted process. If furthermore Z is a square integrable martingale and $H \in L^2(Z)$, then $H \cdot Z$ is a square integrable martingale.*

We assume that the space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ supports a càdlàg, $\mathbb{F}$-adapted Lévy process $(L_t)_{t \geq -r}$, such that $L_{-r} = 0$ a.s., L is stochastically continuous and for all $-r \leq s < t < \infty$, $L_t - L_s$ is independent of $\mathcal{F}_s$ and has the same distribution as L_{t-s-r}. Put $\mathbb{R}_0 := \mathbb{R} \setminus \{0\}$ and write $\mathcal{B}(\mathbb{R}_0)$ for the Borel sigma-algebra on $\mathbb{R}_0$. When $U \in \mathcal{B}(\mathbb{R}_0)$ with $0 \notin \overline{U}$, write

$$N(t, U) := \sum_{-r \leq s \leq t} \mathbb{1}_U(\Delta L_s), \quad t > 0$$

for the Poisson random measure on $\mathcal{B}(0, \infty) \times \mathcal{B}(\mathbb{R}_0)$ associated with $(L_t)_{t \geq -r}$ and write $\Pi_L(U) := \mathbb{E}[N(-r+1, U)]$ for the corresponding Lévy measure on $\mathcal{B}(\mathbb{R}_0)$. Write $\widetilde{N}(dt, dz) := N(dt, dz) - \Pi_L(dz)dt$ for the compensated Poisson random measure.

Recall that a Lévy measure Π_L always satisfies $\int_{\mathbb{R}_0} (1 \wedge z^2) \Pi_L(dz) < \infty$. Throughout the chapter, we will also assume that Π_L has finite second moment and L is centered, so that $(L_t)_{t \geq -r}$ is a square integrable martingale with respect to $\mathbb{F}$, i.e., $\mathbb{E}[L_t] = 0$ and $\mathbb{E}[L_t^2] < \infty$ for all $t \geq -r$. The characteristic function of L_t is given by the Lévy-Khintchine formula

$$\mathbb{E}\Big[e^{iuL_t}\Big] = \exp\Big\{(t+r)\Big(-\frac{1}{2}\sigma_L^2 u^2 + \int_{\mathbb{R}_0} (e^{iuz} - 1 - iuz)\Pi_L(dz)\Big)\Big\}, \quad u \in \mathbb{R},$$

where $\sigma_L > 0$. Furthermore, the Lévy-Itô decomposition of L gives

$$L_t = \sigma_L B_t + \int_{-r}^{t} \int_{\mathbb{R}_0} z\widetilde{N}(dz, ds), t \geq -r, \tag{3.2.1}$$

where $(B_t)_{t \geq -r}$ is a standard Brownian motion with respect to $\mathbb{F}$, having $B_{-r} = 0$, a.s. The quadratic variation process $S := [L, L]$ of L is the subordinator

$$S_t = \sigma_L^2(t + r) + \int_{-r}^{t} \int_{\mathbb{R}_0} z^2 N(dz, dt), \quad t \geq -r. \tag{3.2.2}$$

Put $\kappa_2 := \mathbb{E}[S_{-r+1}] = \sigma_L^2 + \int_{\mathbb{R}_0} z^2 \Pi_L(dz) < \infty$, so that the process $(\widetilde{S}_t)_{t \geq -r}$ defined by

$$\widetilde{S}_t := S_t - \kappa_2(t + r) = \int_{-r}^{t} \int_{\mathbb{R}_0} z^2 \widetilde{N}(dz, dt), \quad t \geq -r, \tag{3.2.3}$$

is a martingale with respect to $\mathbb{F}$ (see [6], Theorem 2.5.2). If furthermore L has finite fourth moments, then $\kappa_4 := \mathbb{E}[\widetilde{S}_{-r+1}^2] < \infty$ and $\widetilde{S}$ is a square integrable martingale, with predictable quadratic variation process $d\langle S \rangle_t = \kappa_4 dt$.

3.2.2 Delay Differential Equations

Consider the deterministic functional differential equation

$$\frac{d}{dt} x(t) = \int_{[-r,0]} x(t + u) \mu(du), \quad t \geq 0, \tag{3.2.4}$$

with initial condition $x|_{[-r,0]} = \varphi$ for some $\varphi \in \mathscr{D}$. Here μ is a signed Borel measure with finite total variation on $[-r, 0]$. For each initial condition $\varphi \in \mathscr{D}$, there exists a unique solution $t \mapsto x(t, \varphi)$ on $[-r, \infty)$, i.e. $x(u, \varphi) = \varphi(u)$ for all $u \in [-r, 0]$, $t \mapsto x(t, \varphi)$ is continuously differentiable on $(0, \infty)$, and (3.2.4) holds on $(0, \infty)$. The asymptotic stability of this solution as $t \to \infty$ is governed by the roots of the so-called *characteristic function* $\Delta : \mathbb{C} \to \mathbb{C}$ of μ, defined as

$$\Delta(z) := z - \hat{\mu}(z) = z - \int_{[-r,0]} e^{zu} \mu(du). \tag{3.2.5}$$

Let $x(\cdot, \varphi)$ be a solution to (3.2.4) and fix any $\lambda \in \mathbb{R}$ such that $\Delta(z) \neq 0$ on the line $\mathrm{Re}\, z = \lambda$. Then [62] gives the following asymptotic expansion of $t \mapsto x(t, \varphi)$:

$$x(t, \varphi) = \sum_{j=1}^{n} p_j(t) e^{z_j t} + o(e^{\lambda t}), \quad t \to \infty, \tag{3.2.6}$$

where $z_1, \ldots, z_n$ are finitely many zeros of $\Delta(z)$ with real part exceeding λ, and $p_j(t)$ is a $\mathbb{C}$-valued polynomial in t of degree less than the multiplicity of z_j as a zero of $\Delta(z)$. In particular, it's clear from (3.2.6) that if $\Delta(z)$ is root free in the right half-plane $\{z | \mathrm{Re}\, z \geq 0\}$, then the zero solution is asymptotically stable, that is, all solutions $x(\cdot, \varphi)$

of the functional differential equation (3.2.4) converge to the zero solution exponentially fast as $t \to \infty$.

3.3 The Solution Process

We first establish the existence and uniqueness of a solution to (3.1.6) in Section 3.3.1. We show that this solution satisfies a stochastic functional differential equation which is easier to work with than the original formulation (3.1.6). Using this new formulation we study the jumps of the solution in Section 3.3.2 and establish the positivity of the solution. All proofs and supporting lemmas are deferred to Section 3.5.

3.3.1 Existence and Uniqueness

We first specify the space on which we are solving equation (3.1.6) and define the notion of a strong solution. Given a stochastic process $(Z_t) \in \mathbb{D}_{[-r,\infty)}$, define $Z_t^* := \sup_{0\leq s\leq t} |Z_s|$ and $Z^* := \sup_{s\geq 0} |Z_s|$. Let $(|\cdot|_t)_{t\geq -r}$ be a family of semi-norms given by

$$|Z|_t := |Z_t^*|_{L^2(\Omega,\mathbb{P})} = \left(\mathbb{E}\left[\sup_{s\in[-r,t]} |Z_s|^2\right]\right)^{1/2}. \tag{3.3.1}$$

We denote by $\mathcal{S}^2$ the class of càdlàg processes with finite $|\cdot|_t$ for every $t \geq -r$.

Definition 3.2. A stochastic process $X = (X_t^{\Phi})_{t\geq -r}$ adapted to $\mathbb{F}$ is called a strong solution to equation (3.1.6) with $\mathbb{D}$-valued initial condition Φ if X belongs to $\mathcal{S}^2$, satisfies $X|_{[-r,0]} = \Phi$, and the equation (3.1.6) holds for all $t \in (0,\infty)$. We refer to this solution X as the CDGARCH(p,q) variance process.

The following set of conditions ensures (3.1.6) has a unique strong solution in the sense of Definition 3.2.

Assumption 3.1.

a) The initial process $\Phi \in \mathbb{D}$ is adapted to $\sigma(L)$, with $|\Phi|_0 < \infty$.

b) The process S as defined in (3.2.2) is square integrable, i.e. $\mathbb{E}[L_1^4] < \infty$.

Here we remark that (b) of Assumption 3.1 does indeed seem very strong for the purpose of obtaining a solution to the equation. However, it is necessary since we are interested in the second order properties of the solution. The following theorem establishes the existence and uniqueness of a solution to the CDGARCH equations. To the extent of our best knowledge, the form of the CDGARCH equation is not covered by any existing results in the literature, hence we include a proof for the following results.

Theorem 3.3. *Suppose S and Φ satisfy Assumption 3.1. Then* *a) There exists a unique strong solution X to (3.1.6) with initial condition Φ.*

b) For all $\alpha \in [0,2]$, the function $t \mapsto \mathbb{E}[|X_t|^\alpha]$ is finite valued and càdlàg.

We note that the equation (3.1.6) is rather difficult to work with. To obtain a more convenient expression for the solution to (3.1.6), we first introduce some notations. Recall the functions f_μ and f_ν from (3.1.7). Let $F_\mu, F_\nu : \mathbb{R}^2 \to \mathbb{R}$ be functions defined by

$$F_\mu(t,s) := \int_{[-p\vee(s-t),s\wedge 0]} f_\mu(u)du, \quad F_\nu(t,s) := \int_{[-q\vee(s-t),s\wedge 0]} f_\nu(u)du. \tag{3.3.2}$$

Then F_μ and F_ν are known as Volterra type kernels on $\mathbb{R}^2$, i.e. $F(t,s) = 0$ for all $s \geq t$. Finally, we define a stochastic process $(\Xi(X)_t)_{t\geq 0}$ by

$$\Xi(X)_t := \int_{[-p,t]} F_\mu(t,s)X_s ds + \int_{(-q,t]} F_\nu(t,s)X_{s-}dS_s, \quad t \geq 0. \tag{3.3.3}$$

The process $\Xi(X)$ is known in the literature as a convoluted Lévy process; we will discuss this connection in more detail in Remark 3.6. We first present some elementary properties of the functions F_μ, F_ν and the process $\Xi(X)$.

Proposition 3.4. *Suppose the functions f_μ and f_ν are non-negative and continuous on $[-p,0]$ and $[-q,0]$ respectively. Then*

a) The kernels F_μ and F_ν are non-negative Lipschitz continuous functions on $\mathbb{R}^2$.

b) The process $(\Xi(X)_t)_{t\geq 0}$ has locally Lipschitz continuous sample paths. Furthermore, it is differentiable at Lebesgue almost every $t \geq 0$ almost surely, with derivative

$$\xi(X)_t := \frac{d}{dt}\Xi(X)_t = \int_{-p}^{0} f_\mu(u)X_{t+u}du + \int_{-q+}^{0} f_\nu(u)X_{t+u-}dS_{t+u}. \tag{3.3.4}$$

Using F_μ, F_ν and $\Xi(X)$, we can express X as a stochastic functional differential equation driven by the quadratic variation process S.

Theorem 3.5. *Let X be the unique strong solution to the CDGARCH(p,q) variance equation (3.1.6), with parameters specified in (3.1.9), (3.1.7) and driving noise S defined in (3.2.2). Then the process X satisfies the stochastic (Volterra) integral equation*

$$X_t = X_0 + \int_0^t \Big(\eta - c_\mu X_s\Big)ds + c_\nu \int_{0+}^t X_{s-}dS_s + \Xi(X)_t, \quad t \geq 0, \tag{3.3.5}$$

which can be rewritten into a differential form

$$dX_t = \Big(\eta - c_\mu X_t + \xi(X)_t\Big)dt + c_\nu X_{t-}dS_t, \quad t \geq 0, \tag{3.3.6}$$

with initial condition $X_u = \Phi_u$ on $[-r,0]$. In particular, X is a semimartingale and has paths of finite variation on compacts sets.

We note that equation (3.3.6) is of a form commonly seen in the literature while the original CDGARCH equation (3.1.6) is more difficult to deal with. Furthermore, from (3.3.6) it is clear that the CDGARCH(p, q) process is a direct generalization of the COGARCH process. Indeed, without the term $\xi(X)_t$, the equation (3.3.6) is exactly the SDE for the COGARCH process (3.1.10). We can therefore interpret the CDGARCH(p, q) process as a COGARCH process with an extra delay-type drift coefficient $\xi(X)_t$.

Since the coefficient $\xi(X)_t$ depends on the paths of both X and S, the CDGARCH process $(X_t)_t$ clearly is not a Markovian process. A common technique in the literature to deal with delayed equations like (3.3.6) is to "lift" the solution X up into a functional space so that it becomes Markovian, for instance, see [147]. We do not pursue this direction further in our work. A few remarks are in place to discuss (3.3.6), the process $\Xi(X)$ and its connections to the literature.

Remark 3.6. a) The stochastic process $t \mapsto \Xi(X)_t$ in (3.3.3) is an example of a *convoluted Lévy process*, studied in [32]. In fact, with a different choice of kernel, the process $\Xi(X)$ could be a fractional Lévy process considered in [30, 31] and [126]. A recent work [88] also considered a similar process with convolution type kernels and Brownian driving noise, with applications to modeling asset volatility.

b) The process $\xi(X)$ has been studied in [22] (and in multiple related works by the same group of authors) over the past few years to model stochastic volatility and turbulent flows. In particular $\xi(X)$ is referred to as a *volatility modulated Lévy driven Volterra (VMLV)*, or more specifically, a *Lévy semi-stationary (LSS) process* in [23]. Furthermore, these processes are special cases of a much more general class of objects called Ambit fields. We refer to [22] and [143] for surveys of relevant results.

Finally, we observe from (3.2.2) that the Brownian component of L appears in the quadratic variation process S as a positive drift $\sigma_L^2(t + r)$. From (3.3.4) and (3.3.6), it is clear that we could absorb this drift into the constant c_μ and the function f_μ. That is, by replacing c_μ with $c_\mu - \sigma_L^2 c_\nu$ and f_μ with $f_\mu + \sigma_L^2 f_\nu$, we can assume without loss of generality that $\sigma_L^2 = 0$ and S is a pure jump Lévy process.

3.3.2 Sample Path of the Solution

We now focus on the path properties of the CDGARCH(p, q) process X. The jumps of X are shown to exhibit similar behaviours to those of the COGARCH process, however the behaviour of X in between jumps is more complex and interesting. Furthermore, we will show that X stays positive and is bounded away from zero, which is to be expected since X is the conditional variance process of the process Y.

We will first focus on the case where the driving noise L is a compound Poisson process. In general a pure jump Lévy process L could have infinite activity, i.e. L could have infinite number of jumps in any compact time interval. In the literature of Lévy processes it is standard to truncate the small jumps of L and approximate L using a sequence of compound Poisson processes which have finite activity, see [6, 152]. In Section 3.3.3 we will show that this approximation can be carried over to the solution of (3.3.6) driven by L as well. More specifically, we define a sequence $(L^n)_n$ of compound Poisson processes approximating L and show that the solution to (3.3.6) driven by L^n converges in a fairly strong sense to the solution (3.3.6) with L being the driving noise.

We start with the path properties of X.

Proposition 3.7. *Let $(L_t)_{t\geq 0}$ be a compound Poisson process, i.e. $\sigma_L = 0$ and the Lévy measure Π_L is a finite measure. Let $-r < T_0 < T_1 < \dots$ be the times of jumps of L and $(\Delta L_t)_{t\geq -r}$ be the sizes of those jumps. Then*

a) The jumps of X are driven by the jumps of $S = [L, L]$ and

$$\Delta X_t = c_\nu X_{t-}\Delta S_t = c_\nu X_{t-}(\Delta L_t)^2, \quad t \geq -r.$$

b) Suppose $f_\nu(-q) = 0$, then on each $[T_n, T_{n+1})$, the process $\xi(X)$ is continuous and X is continuously differentiable, with derivative given by

$$\frac{d}{dt}X_t = \eta - c_\mu X_t + \xi(X)_t, \quad t \in (T_j, T_{j+1}).$$

Furthermore, when $p > 0$ and $q = 0$, between two consecutive jump times, the process X satisfies the deterministic differential equation

$$\frac{d}{dt}X_t = \eta - c_\mu X_t + \int_{-p}^{0} f_\mu(u)X_{t+u}du, \quad t \in (T_j, T_{j+1}). \tag{3.3.7}$$

In the case $p = q = 0$, i.e. when X is a COGARCH process, X decays exponentially between its jump times, and we have a closed form solution

$$X_t = \frac{\eta}{c_\mu} + \left(X_{T_j+} - \frac{\eta}{c_\mu}\right)e^{-c_\mu(t-T_j)}, \quad t \in (T_j, T_{j+1}).$$

From (a) of the Proposition it is clear that the jump structure of X is the same between the COGARCH process and the CDGARCH(p, q) process. However, the behavior of X in between jumps is very different. We illustrate these differences by simulating sample paths of the CDGARCH(p, q) processes of different orders via a simple Euler scheme.

For clarity we set L to be a compound Poisson process with unit intensity and jumps equal to ± 1 with equal probability. The top figure in Figure 3.1 shows a simulated path of L. The processes below are the COGARCH (or CDGARCH$(0, 0)$), the CDGARCH$(p, 0)$ and the CDGARCH(p, q) variance processes respectively, driven by the same realization of L. The horizontal lines are the theoretical stationary means of the variance processes, computed in Section 3.4. The delay functions f_μ and f_ν are chosen to be exponential, and comparable parameters are chosen between all three processes.

For the COGARCH process, i.e. when $p = q = 0$, the process X is deterministic between jumps and decays exponentially. In the CDGARCH$(p, 0)$ case, the process X follows a deterministic differential equation (given in Proposition 3.7) between the jumps of the driving noise L, but the decay towards the baseline level is slower than the exponential function, indicating a longer memory effect.

In the case $p, q > 0$, the process X is no longer deterministic between jumps; instead it is a continuous process of finite variation that depends on $\{X_u, u \in [T_j - p, T_j]\}$ as well as $\{\Delta S(u), u \in [T_j - q, T_j]\}$. For the particular realization shown below, X is in fact increasing immediately after a jump, then starts decaying towards the baseline level. Depending on choices and sizes of f_ν, it is possible to have a range of different behaviors between jumps.

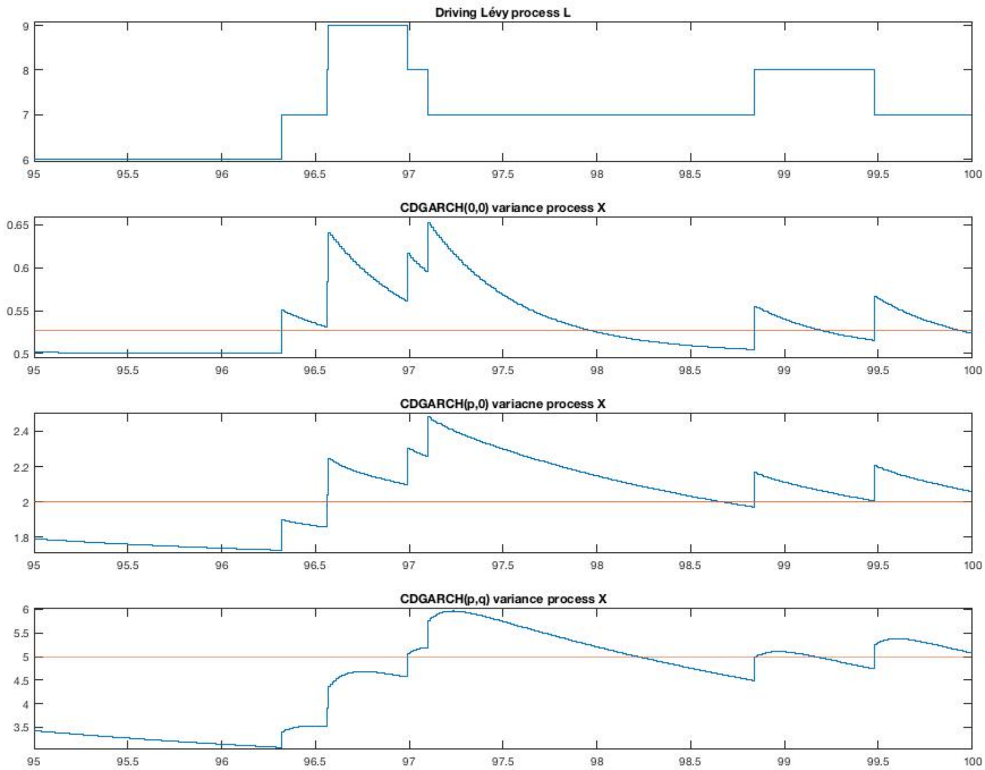

Figure 3.1: simulated paths of CDGARCH processes with different orders

In comparison to the COGARCH process, the CDGARCH(p, q) process decays significantly slower in between jumps. This is a behaviour shared by higher order GARCH processes as well in comparison to the GARCH$(1, 1)$. In this sense, the CDGARCH(p, q) process does a better job at capturing high order delays than the COGARCH process. From Proposition 3.7 it is clear that for the COGARCH process, the rate of decay is controlled by the constant c_μ, which can be interpreted as the speed of mean reversion. In the CDGARCH$(p, 0)$ process, beisdes of the decay rate c_μ, there is an additional upward drift whose size depends on f_μ as well as past values of X. As will be shown, the constant c_μ needs to be large enough relative to f_μ for the process X to behave nicely.

The simulations in Figure 3.2 illustrate the behaviour of X around each jump. We can also simulate the CDGARCH process over a longer period of time using settings reminiscent of real financial data. Figure 3.2 shows a simulated sample path of the driving noise L, the return process dY, the CDGARCH process Y and its volatility process $\sqrt{X}$. Here L is set to a compound Poisson process with high intensity. We note that the CDGARCH(p, q) process exhibits many of the features of the GARCH process, including volatility clustering and the persistence of volatility.

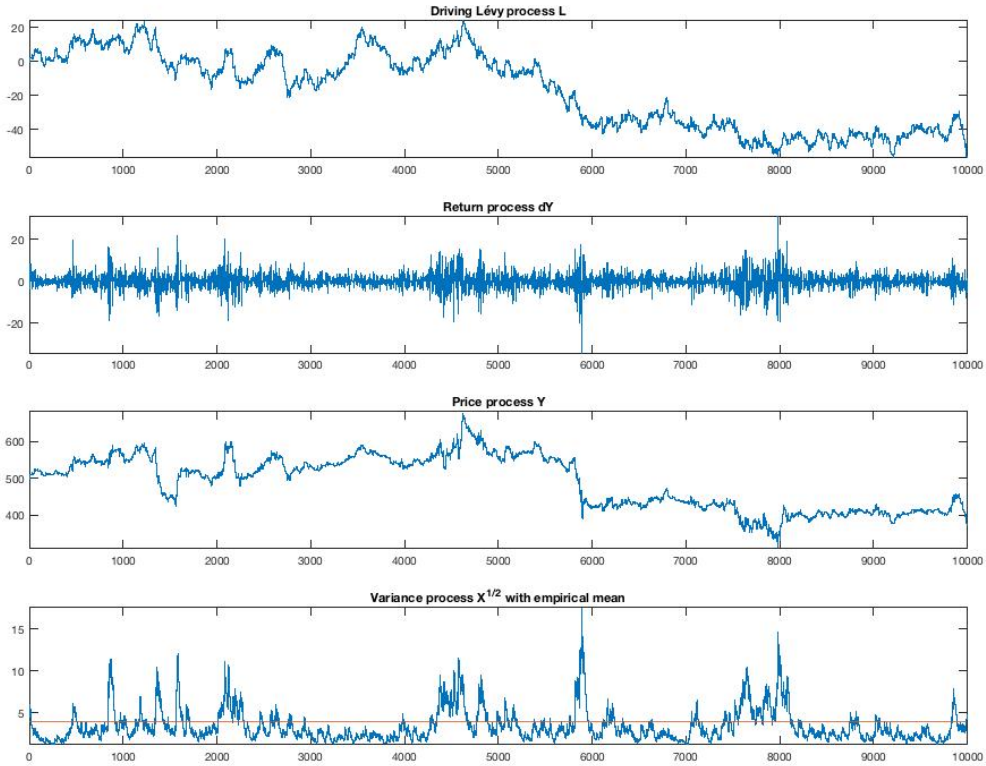

Figure 3.2: Simulated paths of the processes L, dY, Y and X

Finally we give conditions for X to remain positive.

Proposition 3.8. *Let S be a compound Poisson process satisfying Assumption 3.1(b) and X be the unique solution to (3.3.6) driven by S. Suppose $\eta > 0$, $c_\mu > |f_\mu|_{L^1}$ and*

$$X_u \geq x^- := \frac{\eta}{c_\mu - |f_\mu|_{L^1}} > 0, \quad \forall u \in [-r, 0]. \tag{3.3.8}$$

Then $X_t \geq x^-$ for all $t > 0$, i.e. X is positive and bounded away from zero

3.3.3 An Approximation Result

Our analysis in the preceding section is carried out under the assumption that L is a compound Poisson process. We now relax this assumption by showing that the general case can be well approximated by this special case.

Without loss of generality we suppose L is a pure jump process in which case

$$S_t = \sum_{-r<s\leq t} (\Delta L_s)^2. \tag{3.3.9}$$

For each $n \in \mathbb{N}$, define the approximating process S^n by

$$S_t^n := \sum_{-r<s\leq t} (\Delta L_s)^2 \mathbb{1}_{\{|\Delta L_s| \geq \frac{1}{n}\}}. \tag{3.3.10}$$

Then $(S^n)_n$ is a sequence of compound Poisson processes satisfying $S_t^n \leq S_t$ for all $n \in \mathbb{N}$ and $t \geq -r$. For each n, we will consider equation (3.3.6) driven by S^n:

$$dX_t^n = b^n(X^n)_t dt + c_\nu X_{t-}^n dS_t^n, \tag{3.3.11}$$

where the drift coefficient $b^n : \mathbb{R}_+ \times \mathbb{R} \times \Omega$ is defined as

$$b^n(H)_t := \eta - c_\mu H_t + \int_{t-p}^t f_\mu(u-t) H_u du + \int_{t-q+}^t f_\nu(u-t) H_{u-} dS_u^n.$$

Applying Theorem 3.3, we see that equation (3.3.11) has a unique solution X^n in $\mathcal{S}^2$ for each initial value Φ satisfying Assumption 3.1a. Similar to the definition of b^n, we will write b for the drift coefficient of (3.3.6) driven by S. The main result of the current section is to show that X^n converges to X in the following sense. We recall from [145] that a sequence of processes $(H^n)_n$ converges to H uniformly on compacts in probability (ucp) if for each $t > 0$, the sequence $\sup_{s\leq t} |H_s^n - H_s|$ converges to 0 in probability as $n \to \infty$.

We will need the following approximation results on the drift term of the equation. Let $(U_t)_{t\geq 0}$ be the (finite and increasing) process given by

$$U_t := c_\mu + |f_\mu|_{L^1} + (|f_\nu|_{L^1} + f_\nu(0))S_t^*. \tag{3.3.12}$$

Proposition 3.9. *Let X and $(X^n)_{n\in\mathbb{N}}$ be the unique solutions to (3.3.6) and (3.3.11).*

a) For each $t \geq 0$, S^n converges to S in $|\cdot|_t$ and hence in ucp.

b) The drift coefficient b is functional Lipschitz (page 256, [145]) with Lipschitz process $(U_t)_t$ defined in (3.3.12). That is, for every Y and Z in $\mathcal{S}^2$, we have

$$|b(Y)_t - b(Z)_t| \leq U_t \sup_{s\leq t} |Y_s - Z_s|, \quad \forall t \geq 0.$$

Furthermore, for all $n \in \mathbb{N}$, b^n is functional Lipschitz with the same U.

c) *For each $t \geq 0$, the process $b^n(X)$ converges to $b(X)$ in $|\cdot|_t$ and hence in ucp.*

We are now in a position to state the approximation result, which allows us to extend Proposition 3.8 to a general driving noise S.

Theorem 3.10. *Let S and $(S^n)_n$ be given by (3.3.9) and (3.3.10). Let X and $(X^n)_n$ be the solutions to (3.3.6) and (3.3.11). Then as $n \to \infty$, X^n converges to X in ucp.*

Corollary 3.11. *Proposition 3.8 holds for any S of the form (3.3.9) satisfying Assumption 3.1b. In particular, suppose $\eta > 0$ and $c_\mu > |f_\mu|_{L^1}$, then for each $t > 0$, X_t is positive and bounded away from zero by x^- defined in (3.3.8).*

3.4 Moments and Stationarity

We switch our attention to the second order structure of the process X and compare to the results of [100]. Observe from (3.3.6) that c_μ can be interpreted as the speed of mean reversion and acts as a negative drift in the SDE (3.3.6). On the other hand, the constant c_ν and the functions f_μ and f_ν all contribute to the positive drift in (3.3.6). Intuitively, the value of c_μ has to be large enough to balance out the effects of c_ν, f_μ, f_ν and keep the solution X from exploding. The following result shows that when c_μ is large enough, the solution X is in fact uniformly bounded in L^1 or L^2.

Recall η from (3.1.6). Let $C_1^+, C_1^-, C_2^+, C_2^-$ be given by

$$
\begin{aligned}
C_1^{\pm} &= c_\mu - \kappa_2 c_\nu \pm \Big(|f_\mu|_{L^1} + \kappa_2 |f_\nu|_{L^1} \Big), \\
C_2^{\pm} &= c_\mu - \kappa_2 c_\nu - \frac{1}{2}\kappa_4 c_\nu^2 \pm \Big(|f_\mu|_{L^1} + \kappa_4 |f_\nu|_{L^2} \Big).
\end{aligned}
$$

Proposition 3.12. *Suppose Assumptions 3.1 hold and X is a positive solution to (3.3.6).*

a) *Suppose $\mathbb{E}[X_0] < \infty$ and $C_1^- > 0$, or equivalently,*

$$c_\mu > \kappa_2 c_\nu + |f_\mu|_{L^1} + \kappa_2 |f_\nu|_{L^1}. \tag{3.4.1}$$

Then X is uniformly bounded in L^1 with $\sup_t \mathbb{E}[X_t] \leq 2\eta/C_1^- + \mathbb{E}[X_0]C_1^+/C_1^-$.

b) *Suppose $\mathbb{E}[X_0^2] < \infty$ and $C_2^- > 0$, or equivalently,*

$$c_\mu > \kappa_2 c_\nu + \frac{1}{2}\kappa_4 c_\nu^2 + |f_\mu|_{L^1} + \kappa_4 |f_\nu|_{L^2}. \tag{3.4.2}$$

Then X is uniformly bounded in L^2 with $\sup_t \mathbb{E}[X_t^2] \leq \big(\eta/C_2^-\big)^2 + \mathbb{E}[X_0^2]C_2^+/C_2^-$.

3.4.1 The Moment Processes

Let m be the mean function of X, i.e., $m(t) := \mathbb{E}[X_t]$, $t \in [-r, \infty)$. Write $\varphi(\cdot)$ for the mean function of the initial segment Φ. For $t > 0$, define the *segment process* $m_{(t)} : [-r, 0] \to \mathbb{R}$ of the process m as $m_{(t)}(u) := m(t+u), u \in [-r, 0]$. For notational simplicity, we will from here onwards write $c_0 := c_\mu - \kappa_2 c_\nu$ and $f := f_\mu + \kappa_2 f_\nu$. We first show that the mean function of the process X satisfies a deterministic delayed differential equation.

Proposition 3.13. *Suppose Assumption 3.1 is satisfied. Then*

a) The mean function m is finite-valued, continuously differentiable on $(0, \infty)$, and satisfies the (deterministic) functional differential equation

$$\frac{d}{dt} m(t) = \eta - c_0 m(t) + \int_{-r}^{0} m(t+u) f(u) du, \quad t > 0, \tag{3.4.3}$$

with the initial condition $m(u) = \varphi(u)$ for $u \in [-r, 0]$.

b) The mean function m also satisfies the renewal equation

$$m(t) = \int_0^t \zeta(t-u) m(u) du + h(t), \quad t > 0,$$

with initial condition $m(0) = \varphi(0)$. The convolution kernel ζ is given by

$$\zeta(t) = -c_0 \mathbb{1}_{(0,\infty)}(t) + \int_0^{t \wedge r} f(-u) du, \quad t \in [0, \infty), \tag{3.4.4}$$

and the forcing function $h : [0, \infty) \to \mathbb{R}$ is given by

$$h(t) = m_{(0)}(0) + \frac{\eta}{\zeta(r)} \int_0^t \zeta(u) du + \int_0^r (\zeta(t+u) - \zeta(u)) \left(m_{(0)}(-u) + \frac{\eta}{\zeta(r)} \right) du.$$

Recall that X is said to be mean stationary if $m(t) = M$ for all $t > 0$ for some $M > 0$ and X is said to be asymptotically mean stationary if $m(t) \to M$ as $t \to \infty$. Using Proposition 3.13, we can show that condition 3.4.1 is in fact necessary and sufficient for the mean stationarity or asymptotic mean stationarity of X.

Theorem 3.14. *Suppose Proposition 3.13 holds so $m(t)$ satisfies the functional differential equation (3.4.3) with some positive initial condition $\varphi \in \mathscr{D}$. Then*

a) The mean function m converges to a (positive) limit M exponentially fast as $t \to \infty$, if and only if $c_0 > |f|_1$. If it exists, the limit M is equal to

$$M = \frac{\eta}{c_0 - |f|_1}. \tag{3.4.5}$$

b) *The process X admits a stationary (positive) mean, i.e. $m(t) = M$ for all $t \in [0, \infty)$, if and only if $c_0 > |f|_{L^1}$ and $\varphi \equiv M$ on $[-r, 0]$, where M is given by (3.4.5).*

Recall that $c_\mu, c_\nu, f_\mu, f_\nu$ are in our case analogues of the coefficients of the GARCH process. Keeping this in mind it is easy to see that the necessary and sufficient condition (3.4.1) for mean stationarity and the formula 3.4.5 for the asymptotic mean are exact analogues of the discrete time GARCH process, see [116].

The second moment of the process X involves the term $\mathbb{E}[X_t \int_{t-q+}^t X_s f_\nu(s) dS_s]$, which in general cannot be computed easily. However, we can formulate some asymptotic results.

Theorem 3.15. *Suppose condition (3.4.1) is satisfied so that $\mathbb{E}[X_u] \to M$ as $u \to \infty$. For every $t > 0$ and $\mathcal{F}_t$-measurable random variable Z with $\mathbb{E}[Z^2 X_u^2] < \infty$, we have $\mathbb{E}[Z X_u] \to M\mathbb{E}[Z]$ exponentially fast as $u \to \infty$.*

The asymptotic behavior of the covariance function $\mathrm{Cov}(X_t, X_{t+u})$ of the process X is an immediate corollary to Theorem 3.15 by taking $Z = X_t$.

Corollary 3.16. *Suppose X is asymptotically mean stationary and has finite fourth moments. Then for every $t > 0$, the covariance function $\mathrm{Cov}(X_t, X_{t+u})$ tends to zero exponentially fast as $u \to \infty$.*

We finally look at the properties of the price and return processes under the CDGARCH model. Recall the price process $Y_t = Y_0 + \int_{0+}^t \sqrt{X_{s-}} dL_s, t \geq 0$, and define the return process $(\widetilde{Y}_t)_{t>1}$ by $\widetilde{Y}_t := Y_t - Y_{t-1} = \int_{t-1+}^t \sqrt{X_{s-}} dL_s$.

Corollary 3.17. *Let $(X_t)_t$ be the solution to (3.3.6) and $Y, \widetilde{Y}$ be defined as above. Suppose X is mean stationary, with mean M defined in (3.4.5).*

a) *The return process $\widetilde{Y}$ is covariance stationary, with zero mean and auto-covariance function given by $\mathrm{Cov}(\widetilde{Y}_t, \widetilde{Y}_{t+u}) = \mathbb{E}\Big[\widetilde{Y}_t \widetilde{Y}_{t+u}\Big] = \kappa_2 M (1-u)_+$.*

b) *Suppose $\widetilde{Y}$ has finite fourth moments. Then for any $t > 1$, the squared return process $(\widetilde{Y}_t^2)_{t>1}$ satisfies $\mathrm{Cov}(\widetilde{Y}_t^2, \widetilde{Y}_{t+u}^2) \to 0$ exponentially fast as $u \to \infty$.*

We note that a stochastic process H with the property that $\mathrm{Cov}(H_t, H_{t+u}) \to 0$ exponentially fast is said to have short memory, see [126, 130, 151] and the references therein. From Corollary 3.16 and (3.17) we see that that the CDGARCH process indeed has short memory, just like the discrete time GARCH process.

3.5 Proofs

3.5.1 Existence and uniqueness of the solution

We first introduce some notations. Given a signed measure μ on a measure space (E, Σ), we denote its corresponding total variation measure $|\mu|$ by $|\mu|(E) = \sup_\pi \sum_{A \in \pi} |\mu(A)|$, for all $E \in \Sigma$, where the supremum is taken over all Σ-measurable partitions π of E. We also denote the total variation norm of μ as $|\mu| := |\mu|(S)$.

Let μ and ν be signed Borel measures on $[-r, 0]$ with finite total variations and S be a càdlàg, adapted process with paths of finite variation. Recall θ from (3.1.9) and $|\cdot|_t$ from (3.3.1). We can write the variance equation (3.1.6) as $X_t = \theta_t + \mathcal{R}(X)_t$, where $\mathcal{R}$ is a linear map on $\mathscr{D}$ given by

$$\mathcal{R}(Z)_t = \int_{-r}^{0} \int_{u}^{t+u} Z_s ds \mu(du) + \int_{-r}^{0} \int_{u+}^{t+u} Z_{s-} dS_s \nu(du), \quad t > 0, \tag{3.5.1}$$

and $\mathcal{R}(Z)_u = 0$ for all $u \leq 0$. Using Lemma 3.1, it is easy to see that $\mathcal{R}(Z)$ is càdlàg and adapted, whenever Z is càdlàg and adapted. We first obtain some norm estimates on $\mathcal{R}$:

Lemma 3.18. *Let $(H_t)_{t \geq -r}$ be a process in $\mathcal{S}^2$, as defined in (3.3.1). Then, under Assumption 3.1, for all $T \geq -r$,*

$$|\mathcal{R}(H)|_T^2 \leq K_T \int_{-r}^{T} \mathbb{E}\Big[H_s^2\Big] ds \leq K_T' |H|_T^2,$$

where $K_T = 2(|\mu|^2 + 2\kappa_2 |\nu|^2)T + 16\kappa_4 |\nu|^2 < \infty$ and $K_T' = K_T(T + r) < \infty$.

Proof. For notational convenience, we will define the semi-norm

$$|Z|_{[0,t]} := \left(\mathbb{E}\left[\sup_{s \in [0,t]} |Z_s|^2\right]\right)^{1/2}.$$

Since $\mathcal{R}(H)_u = 0$ for $u \leq 0$, by the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, we have

$$|\mathcal{R}(H)|_T^2 \leq 2\left|\int_{-r}^{0} \int_{u}^{u+\cdot} H_s ds \mu(du)\right|_{[0,T]}^2 + 2\left|\int_{-r}^{0} \int_{u+}^{u+\cdot} H_{s-} dS_s \nu(du)\right|_{[0,T]}^2 =: \mathbf{I} + \mathbf{II}.$$

Since $H = 0$ on $[-r, 0]$, an application of the Cauchy-Schwarz inequality yields the bound

$$\mathbf{I} \leq 2|\mu|^2 \mathbb{E}\left[\sup_{t \in [0,T]} \sup_{u \in [-r,0]} \left(t \int_{u}^{t+u} |H_s|^2 ds\right)\right] \leq 2T|\mu|^2 \mathbb{E}\left[\int_0^T |H_s|^2 ds\right].$$

For $\mathbf{II}$, recall $\widetilde{S}_t := S_t - (t + r)\kappa_2$ from (3.2.3). Using the same reasoning as above,

$$\mathbf{II} \leq 4\kappa_2^2 \left|\int_{-r}^{0} \int_{u}^{u+\cdot} H_s ds \nu(du)\right|_{[0,T]}^2 + 4\left|\int_{-r}^{0} \int_{u+}^{u+\cdot} H_{s-} d\widetilde{S}_s \nu(du)\right|_{[0,T]}^2 =: \mathbf{III} + \mathbf{IV}.$$

By similar workings as in $\mathbf{I}$, we have $\mathbf{III} \leq 4\kappa_2^2 T |\nu|^2 \mathbb{E}\Big[\int_0^T |H_s|^2 ds\Big]$. For $\mathbf{IV}$, recall $\widetilde{S}$ is a square integrable martingale and $d\langle \widetilde{S} \rangle_t = \kappa_4 dt$. Since $|H|_T < \infty$, H is clearly in $L^2(\widetilde{S})$, and the process $H \cdot \widetilde{S}$ is a square integrable martingale by Lemma 3.1. By Jensen's inequality, Doob's inequality and the Ito isometry, we have

$$\mathbf{IV} \leq 4|\nu| \int_{-r}^{0} \mathbb{E} \sup_{t\in[0,T]} \left| \int_{u+}^{t+u} H_{s-} d\widetilde{S}_s \right|^2 |\nu|(du) \leq 16|\nu| \int_{-r}^{0} \mathbb{E} \left| \int_{u+}^{T+u} H_{s-} d\widetilde{S}_s \right|^2 |\nu|(du)$$
$$\leq 16|\nu|^2 \sup_{u\in[-r,0]} \mathbb{E}\left[\int_{u}^{T+u} H_s^2 d\langle \widetilde{S}, \widetilde{S}\rangle_s \right] \leq 16\kappa_4 |\nu|^2 \mathbb{E}\left[\int_{-r}^{T} H_s^2 ds \right]$$

The lemma follows immediately by collecting all terms. □

Proof of Theorem 3.3. (a) Let S and Φ satisfy Assumption 3.1 and θ be defined in (3.1.9). For existence, we use a Picard iteration to produce a sequence of $\mathcal{S}^2$-processes that converges to a limit. Set the initial term $X^{(0)} := \theta \in \mathcal{S}^2$, and define recursively for each $n \geq 1$ the process $X^{(n)} := \theta + \mathcal{R}X^{(n-1)}$. We see that the differences between each term are given by $X^{(1)} - X^{(0)} = \mathcal{R}(\theta)$ and $X^{(n)} - X^{(n-1)} = \mathcal{R}(X^{(n-1)} - X^{(n-2)})$, for $n \geq 2$.

Write $D_{n,T} := |X^{(n+1)} - X^{(n)}|_T$, so that $D_{0,T} = |\mathcal{R}(\theta)|_T$ and for $n \geq 1$, $D_{n,T} = |\mathcal{R}\big(X^{(n)} - X^{(n-1)}\big)|_T$. The first term D_0 is finite by an application of Lemma 3.18 to $H = \theta$. Since $X^{(n+1)} = \theta + \mathcal{R}(X^{(n)})$, by Lemma 3.1 and the second bound in Lemma 3.18, $X^{(n+1)}$ is in $\mathcal{S}^2$ whenever $X^{(n)}$ is in $\mathcal{S}^2$. Therefore by induction, for each $n \geq 1$, the difference $X^{(n)} - X^{(n-1)}$ is in $\mathcal{S}^2$ and we can apply Lemma 3.18 to each $D_{n,T}$.

Since $t \mapsto D_{n,t}$ is non-decreasing and non-negative on $[-r, T]$, applying Lemma 3.18 to each $D_{n,T}$ and expanding the recursion yields a Gronwall type inequality

$$D_{n,T}^2 \leq K_T \int_0^T D_{n-1,t_2}^2 dt_2 \leq K_T^n \int_0^T \int_0^{t_2} \cdots \int_0^{t_n} D_{0,t_{n+1}}^2 dt_{n+1} \cdots dt_3 dt_2 \leq \frac{K_T^n T^n}{n!} D_{0,T}^2.$$

The sequence $(D_{n,T})_n$ is therefore Cauchy for each $T > 0$. Since $\mathcal{D}[-r, \infty)$ is complete in $|\cdot|_T$, taking $n \to \infty$, the sequence of processes

$$X^{(n)} = X^{(0)} + \sum_{k=1}^{n} \Big(X^{(k)} - X^{(k-1)} \Big)$$

converges in $|\cdot|_T$ to a limit X, which is also in $\mathcal{S}^2$.

It remains to show that this limit X is indeed a solution to (3.1.6), i.e. satisfies $X = \theta + \mathcal{R}(X)$. First, observe that since $(D_{n,T})_n$ is summable for every $T > 0$,

$$|X - X^{(n)}|_T = \left| \sum_{k=n}^{\infty} X^{(k+1)} - X^{(k)} \right|_T \leq \sum_{k=n}^{\infty} D^{(k)}(T) \to 0, \tag{3.5.2}$$

as $n \to \infty$. Using $X^{(n)} = \theta + \mathcal{R}(X^{(n-1)})$, by Lemma 3.18, we also have

$$|\theta + \mathcal{R}X - X^{(n)}|_T^2 = |\mathcal{R}(X - X^{(n-1)})|_T^2 \leq K_T' |X - X^{(n-1)}|_T^2 \to 0.$$

Then by the triangle inequality, for any $n \geq 1$,

$$|\theta + \mathcal{R}X - X|_T \leq |\theta + \mathcal{R}X - X^{(n)}|_T + |X - X^{(n)}|_T,$$

which implies that $\sup_{0\le s\le T}|\theta_s + \mathcal{R}X_s - X_s| = 0$ a.s. and hence (3.1.6) is satisfied.

To establish the uniqueness of the solution, suppose X and X' are two strong solutions to (3.1.6), i.e., we have $X = \theta + \mathcal{R}(X)$ and $X' = \theta + \mathcal{R}(X')$. Let $D_T := |X - X'|_T = |\mathcal{R}(X - X')|_T$. By Lemma 3.18, for any $0 \le T < \infty$, we have $D_T^2 \le K_T \int_{-r}^T D_t^2 dt$, where K_T is defined in Lemma 3.18. By Gronwall's inequality (Thm V.68, [145]), $|X - X'|_T^2 = 0$ for all $t \ge 0$ which implies that $\sup_{-r\le s\le T}|X_s - X'_s| = 0$, almost surely, and the two solutions are indistinguishable.

((b).) Since $|x|^\alpha \le 1 + |x|^2$ for any $0 \le \alpha \le 2$ and $x \in \mathbb{R}$, we have,

$$\mathbb{E}[(X_T^*)^\alpha] \le 1 + \mathbb{E}\Big[(X_T^*)^2\Big] < \infty.$$

Hence the function $t \mapsto \mathbb{E}[|X_t|^\alpha]$ is finite-valued. Since X is càdlàg, by the dominated convergence theorem with $|X_T^*|^\alpha$ as dominating functions, $t \mapsto \mathbb{E}[|X_t|^\alpha]$ is a càdlàg function on $[0, \infty)$ for $0 \le \alpha \le 2$. □

3.5.2 Properties of the solution

Proof of Proposition 3.4. (a) We first rewrite $F_\nu(t,s) = \int_{-q}^0 \mathbb{1}_{[s-t,s]}(u) f_\nu(u) du$. For any (t_2, s_2) and $(t_1, s_1) \in \mathbb{R}^2$, by the triangle inequality, F_ν is Lipschitz on $\mathbb{R}^2$ since

$$\begin{aligned}
|F_\nu(t_2,s_2) - F_\nu(t_1,s_1)| &\le |F_\nu(t_2,s_2) - F_\nu(t_1,s_2)| + |F_\nu(t_1,s_2) - F_\nu(t_1,s_1)| \\
&\le \int_{-q}^0 \Big|\mathbb{1}_{[s_2-t_2,s_2]}(u) - \mathbb{1}_{[s_2-t_1,s_2]}(u)\Big| f_\nu(u) du + \int_{-q}^0 \Big|\mathbb{1}_{[s_2-t_1,s_2]}(u) - \mathbb{1}_{[s_1-t_1,s_1]}(u)\Big| f_\nu(u) du \\
&\le |f_\nu| \Big(|t_2 - t_1| + 2|s_2 - s_1|\Big) \le 2|f_\nu| \Big|(t_2,s_2) - (t_1,s_1)\Big|,
\end{aligned}$$

where $|\cdot|$ is the sup-norm and $|\cdot|$ is the Euclidean distance on $\mathbb{R}^n$. Similarly for F_μ.

(b) Since F_μ and F_ν are identically zero whenever $s \ge t$ or $s \le -r$, we will omit the region of integration and write $\Xi(X)_t = \int F_\mu(t,s) X_{s-} ds + \int F_\nu(t,s) X_{s-} dS_s$.

Since for almost every $\omega \in \Omega$, $t \mapsto S_t(\omega)$ is a non-decreasing càdlàg function, we will fix such an ω and treat the stochastic integral above as a Lebesgue-Stieljes integral with respect to the function $t \mapsto S_t(\omega)$. Since F_μ and F_ν vanishes for $s \notin (-r, t)$, by Proposition 3.4(a), for any $t_2, t_1 \in \mathbb{R}_+$,

$$\begin{aligned}
|\Xi(X)_{t_2} - \Xi(X)_{t_1}| &\le \int |F_\mu(t_2,s) - F_\mu(t_1,s)||X_{s-}| ds + \int |F_\nu(t_2,s) - F_\nu(t_1,s)||X_{s-}| dS_s \\
&\le 2|f_\mu| \Big(\int_{-p}^{t_2 \vee t_1} |X_s| ds\Big) |t_2 - t_1| + 2|f_\nu| \Big(\int_{-q+}^{t_2\vee t_1} |X_{s-}| dS_s\Big) |t_2 - t_1|.
\end{aligned}$$

It follows that $t \mapsto \Xi(X)_t$ is locally Lipschitz continuous almost surely, since with probability one X is locally bounded and S has finite variation on compacts.

We first compute $dF_\nu(t,s)/dt$ - the case of F_μ is identical and omitted. In the expression

$F_\nu(t,s) = \int_{-q}^{0} \mathbb{1}_{[s-t,s]}(u) f_\nu(u)du$, the integrand clearly does not depend on t whenever $t \notin (s \vee 0, s+q)$, hence $t \mapsto F_\nu(t,s)$ is constant on these regions and $dF_\nu(t,s)/dt = 0$. On the interval $t \in (s \vee 0, s+q)$, we have $F_\nu(t,s) = \int_{s-t}^{s\wedge 0} f_\nu(u)du$ so by the Fundamental Theorem of Calculus, $t \mapsto F_\nu(t,s)$ is continuously differentiable and $dF_\nu(t,s)/dt = f_\nu(s-t)$ on the interval $t \in (s \vee 0, s+q)$. We can therefore write $dF_\nu(t,s)/dt = f_\nu(s-t)\mathbb{1}_{[s\vee 0,s+q]}(t)$ for almost every $t \geq 0$. Clearly, $t \mapsto F_\nu(t,s)$ is not differentiable at $t = s \vee 0$ or $t = s+q$, unless $f_\nu(0)$ and $f_\nu(-q)$ are equal to zero.

We now compute the derivative of the second integral in (3.3.3):

$$I_t := \int F_\nu(t,s) X_{s-} dS_s, t \geq 0. \tag{3.5.3}$$

The case of the first integral is similar and omitted. Again, we fix an $\omega \in \Omega$ such that $t \mapsto S_t$ is a non-decreasing càdlàg function and treat the dS integral as a Stieljes integral. For every $t \in \mathbb{R}_+$, the map $s \mapsto F_\nu(t,s)X_{s-}$ is in $L^1_{\text{loc}}(\mathbb{R}, dS)$ since X and S are locally bounded and $s \mapsto F_\nu(t,s)$ is supported on a compact set. For every $s \in \mathbb{R}$, the map $t \mapsto F_\nu(t,s)X_{s-}$ is continuously differentiable in $(s \vee 0, s+q)$ by the previous argument. For every t, the derivative $s \mapsto \frac{d}{dt}F_\nu(t,s)X_{s-}$ is locally bounded and hence also in $L^1_{\text{loc}}(\mathbb{R}, dS)$. Then by the differentiation lemma ([98, Theorem 6.28]), $t \mapsto I_t$ is differentiable almost everywhere with derivative

$$\frac{d}{dt} I_t = \int f_\nu(u-t)\mathbb{1}_{[u\vee 0,u+q]}(t) X_{u-} dS_u = \int_{(t-q,t]} f_\nu(u-t) X_{u-} dS_u, \quad t > 0. \tag{3.5.4}$$

The expression (3.3.4) then follows with a simple change of variable. □

Proof of Theorem 3.5. Recalling $\nu(du) = c_\nu \delta_0(du) + f_\nu(u)du$, we have

$$\int_{-q}^{0} \int_{u+}^{t+u} X_{s-} dS_s \nu(du) = c_\nu \int_{0+}^{t} X_{s-} dS_s + \int_{-q}^{0-} \int_{u+}^{t+u} X_{s-} dS_s f_\nu(u) du := \mathbf{I} + \mathbf{II}.$$

Since $X \in \mathcal{S}^2$ and is hence locally bounded and progressively measurable, by Fubini's theorem, exchanging the order of integration of $\mathbf{II}$ gives

$$\mathbf{II} = \int_{(-q,t]} \left(\int \mathbb{1}_{[-q,0)}(u) \mathbb{1}_{(u,t+u]}(s) X_{s-} f_\nu(u) du \right) dS_s = \int_{(-q,t]} F_\nu(t,s) X_{s-} dS_s,$$

for $t \geq 0$, where the kernel F_ν is given by 3.3.2. The computations for the $d\mu$ integral in (3.1.6) are exactly the same and the integral equation (3.3.5) follows immediately.

For the functional differential equation, first observe that $t \mapsto I_t$ in (3.5.3) is Lipschitz and hence absolutely continuous. Hence $I_t - I_0 = \int_{(0,t]} \frac{d}{dt} I_s ds$, where $\frac{d}{dt} I_t$ is given by (3.5.4), with $I_0 = 0$ since $F_\nu(0,s) = 0$ for any s. The integral involving F_μ can be differentiated in exactly the same way. The functional differential equation follows immediately.

Finally, since S is of finite variation and X is càdlàg, X is a semimartingale with finite variations by Theorem I.4.31 of Jacod and Shiryaev [89]. $\square$

Proof of Proposition 3.7. (a) follows immediately from (3.3.6).
(b) On (T_j, T_{j+1}), $S_t = S(T_j)$, so by (a) of Proposition 3.7, X is continuous on (T_j, T_{j+1}). With the normalization $f(-q) = 0$, $\Delta\xi(X)_t = f_\nu(0)X_{t-}\Delta S_t$, so $\xi(X)$ is continuous on (T_j, T_{j+1}) as well. The rest of the proposition follows immediately. $\square$

Proof of Proposition 3.8. Suppose $X_t \geq x^-$ for all $t \in [-r, T]$ for some $T \geq 0$ and let $T' := \inf\{t > T, \Delta S > 0\}$. Since S is a compound Poisson process, we have $T' > T$ almost surely so that the interval $[T, T')$ is non-empty. Then by Proposition 3.7(b), X is continuously differentiable in $[T, T')$ with derivative given by $\dot{X}_t = \eta - c_\mu X_t + \xi(X)_t$. Note that $\Delta X_t \geq 0$ whenever $X_{t-} \geq 0$ by Proposition 3.7. By iterating this argument, it suffices to show that $X_t \geq x^-$ for all $t \in [T, T')$.

Let $T'' := \inf\{t > T, X_t < x^-\}$ and suppose for a contradiction that $T'' \leq T'$ with positive probability. Note that X is continuous at T'' a.s. by definition of T''. Then necessarily we have $X_{T''} = x^-$ and $\dot{X}_{T''} < 0$. But

$$\frac{d}{dt}X_t\Big|_{t=T''} \geq \eta - c_\mu X_{T''} + \int_{-p}^{0} X_{T''+u} f_\mu(u)du \geq 0$$

almost surely, which contradicts our assumption. $\square$

3.5.3 Approximation by processes of finite activity

Proof of Proposition 3.9. (a) From the construction of S^n in (3.3.10), we have

$$S_u - S_u^n = \int_{-r}^{u} \int_{0<|z|<\frac{1}{n}} z^2 N(dz, ds),$$

which in non-decreasing in u. Fixing $t > 0$, we have

$$\mathbb{E}\Big[\sup_{u\leq t} |S_u - S_u^n|^2\Big] = \mathbb{E}\left[\left(\int_{-r}^{t} \int_{0<|z|<\frac{1}{n}} z^2 N(dz, ds)\right)^2\right] = \int_{-r}^{t} \int_{0<|z|<\frac{1}{n}} z^4 \Pi_L(dz).$$

Since $n \geq 1$, the integrand is dominated by z^2 which is $d\Pi_L$ integrable. By the dominated convergence theorem, $\mathbb{E}\big[\sup_{u\leq t} |S_u - S_u^n|^2\big] \to 0$ as $n \to \infty$. That is, S^n approximates S in each $|\cdot|_t$, $t > 0$. This clearly implies convergence in the ucp topology.

(b) Let Y and Z be càdlàg processes in $\mathcal{S}^2$, then

$$\begin{aligned} |b(Y)_t - b(X)_t| &\leq c_\mu |Y_t - Z_t| + \int_{t-p}^{t} f_\mu(u-t)|Y_u - Z_u|du + \int_{t-q+}^{t} f_\nu(u-t)|Y_u - Z_u|dS_u \\ &\leq \sup_{s\leq t} |Y_s - Z_s| \Big(c_\mu + |f_\mu|_{L^1} + \int_{t-q+}^{t} f_\nu(u-t)dS_u\Big). \end{aligned}$$

Since f_ν is continuous and normalized to $f(-q) = 0$, integrating by parts gives

$$\int_{t-q+}^{t} f_\nu(u-t)dS_u = S_t f_\nu(0) - \int_{q+}^{0} S_{t+u} df_\nu(u) \leq S_t^*(f_\nu(0) + |f_\nu|_{L^1}),$$

which implies that b is functional Lipschitz. For each b^n, it suffices to carry through the same computation and observe that by construction, $S_t^n \leq S_t$ for each $n \geq 1$ and $t > 0$.

(c) From the definitions of b^n, for each $t \geq 0$,

$$b(X)_t - b^n(X)_t = \int_{t-q+}^{t} f_\nu(u-t)X_{u-}dS_u - \int_{t-q+}^{t} f_\nu(u-t)X_{u-}dS_u^n.$$

By the construction of S^n, we have

$$|b(X)_t - b^n(X)_t| \leq \sup_{u\in(t-q,t]} |f_\nu(u-t)X_{u-}|\Big(S_t - S_t^n - (S_{t-q+} - S_{t-q+}^n)\Big),$$

which converges to zero in $|\cdot|_t$ for each t and hence in ucp by Proposition 3.9(a). □

Proof of Theorem 3.10. The claim directly follows from Proposition 3.9 and Theorem V.15 of [145]. More accurately, we invoke a trivial extension of Theorem V.15 of [145] to the case with multiple driving semimartingales (see comments on page 257 of [145]). □

Proof of Corollary 3.11. For a given S of the form (3.3.9) satisfying Assumption 3.1(b), let $(S^n)_n$ be as defined in (3.3.10). By Theorem 3.3, we can set X and $(X^n)_n$ to be unique solutions to (3.3.6) and (3.3.11) driven by S and $(S^n)_n$ respectively.

By Theorem 3.10, X^n converges to X in ucp, which trivially implies that for each $t > 0$, $X_t^n \to X_t$ in probability and hence in distribution. Furthermore, since each S^n is a compound Poisson process by construction, by Proposition 3.8, for each $n \geq 1$ and $t > 0$, we have $X_t^n \geq x^-$ with probability one, where $x^- > 0$ is defined in (3.3.8). Finally, since $(-\infty, x^-)$ is open in $\mathbb{R}$, by the Portmanteau theorem of weak convergence (Theorem 2.1 Billingsley [35]), we have for each $t > 0$,

$$\mathbb{P}(X_t < x^-) \leq \liminf_n \mathbb{P}(X_t^n < x^-) = 0,$$

which completes the proof. □

3.5.4 Moment bounds

We precede the proof of Proposition 3.12 with the following two lemmas.

Lemma 3.19 (Lemma 8.1 - 8.2, Ito and Nisio [87]). *Suppose $x, y : [0,\infty) \to \mathbb{R}_+$ are continuous functions, $\alpha > 0$ and $\lambda_1 > \lambda_2 > 0$. For every $0 \leq t < \infty$,*

(a) if $x_t \leq x_0 - \lambda_1 \int_0^t x_u du + \int_0^t y_u du$, then $x_t \leq x_0 + \int_0^t e^{-\lambda_1(t-u)} y_u du$;

(b) if $x_t \le \alpha + \lambda_2 \int_0^t e^{-\lambda_1(t-s)} x_u du$, then $x_t \le \alpha\lambda_1/(\lambda_1 - \lambda_2)$. □

Lemma 3.20. *Suppose Assumptions 3.1 hold, let $(X_t)_{t\ge 0}$ be the unique strong solution to (3.3.6) with initial condition Φ and let $(\xi(X)_t)_{t\ge0}$ be as defined in (3.3.4). For $n \in \{1,2\}$, we have the estimate*

$$\mathbb{E}[|X_t^{n-1}\xi(X)_t|] \le C_k \sup_{u\in[t-r,t]} \mathbb{E}[|X_u|^n], \quad t>0,$$

where $C_1 = |f_\mu|_{L^1} + \kappa_2 |f_\nu|_{L^1}$ and $C_2 = |f_\mu|_{L^1} + \kappa_4|f_\nu|_{L^2}$.

Proof. For the case of $n=2$, by Fubini's theorem and the Cauchy-Schwartz inequality,

$$\begin{aligned}
\mathbb{E}[|X_t\xi(X)_t|] &\le \mathbb{E}\Big[|X_t|\int f_\mu(u-t)|X_u|du\Big] + \mathbb{E}\Big[|X_t|\int f_\nu(u-t)|X_{u-}|dS_u\Big]\\
&\le \int f_\mu(u-t)\mathbb{E}[|X_t||X_u|]du + \mathbb{E}[X_t^2]^{\frac12}\mathbb{E}\Big[\Big(\int f_\nu(u-t)|X_{u-}|dS_u\Big)^2\Big]^{\frac12}\\
&\le \int f_\mu(u-t)\mathbb{E}[X_t^2]^{\frac12}\mathbb{E}[X_u^2]^{\frac12}du + \mathbb{E}[X_t^2]^{\frac12}\kappa_4\Big(\int f_\nu(u-t)^2\mathbb{E}[X_u^2]du\Big)^{\frac12}\\
&\le \Big(|f_\mu|_{L^1} + \kappa_4|f_\nu|_{L^2}\Big)\sup_{u\in[t-r,t]}\mathbb{E}[X_u^2].
\end{aligned}$$

The case of $n=1$ easily follows from the same computations. □

Proof of Proposition 3.12. (a) and (b) The following proof holds for both $n=1$ and $n=2$, with different corresponding constants. Let X be a positive solution to (3.3.6) with $\eta>0$. For $n=2$, it follows from Ito's Lemma ([89, Theorem I.4.57]) that (the $n=1$ case is trivial),

$$X_t^n = X_0^n + n\int_0^t X_s^{n-1}\Big(\eta - c_\mu X_s + \xi(X)_s\Big)ds + \sum_{0<s\le t}\Big\{X_s^n - X_{s-}^n\Big\} \tag{3.5.5}$$

where $\Delta X_t = c_\nu X_{t-}\Delta S_t$ and

$$\begin{aligned}
\sum_{0<s\le t}\Big\{X_s^2 - X_{s-}^2\Big\} &= \sum_{0<s\le t}\Big\{(X_{s-} + c_\nu X_{s-}\Delta S_s)^2 - X_{s-}^2\Big\}\\
&= \sum_{0<s\le t}\Big\{X_{s-}^2\Big(2c_\nu\Delta S_s + c_\nu^2(\Delta S_s)^2\Big)\Big\}.
\end{aligned}$$

Let $E_n(t) := \mathbb{E}[X_t^n]$ and put $K_1 := \kappa_2 c_\nu$ and $K_2 := \kappa_2 c_\nu + \frac12\kappa_4 c_\nu^2$. From (3.5.5), we have

$$E_n(t) = E_n(0) + n\mathbb{E}\Big[\int_0^t X_s^{n-1}\Big(\eta + (-c_\mu + K_n)X_s\Big)ds\Big] + n\mathbb{E}\Big[\int_0^t X_s^{n-1}\xi(X)_s ds\Big]. \tag{3.5.6}$$

Let C_1, C_2 be given as in Lemma 3.20 and suppose

$$\lambda_n := c_\mu - K_n - C_n > 0. \tag{3.5.7}$$

An exercise in calculus gives $\sup_{x\geq 0} 2x\Big(\eta - \frac{1}{2}\lambda_2 x\Big) = \eta^2/\lambda_2 =: a_2$ and $\sup_{x\geq 0}\Big(\eta - \frac{1}{2}\lambda_1 x\Big) = \eta =: a_1$ Then for all $X_s \geq 0$,

$$nX_s^{n-1}\Big(\eta + \frac{1}{2}\Big(-c_\mu + K_n + C_n\Big)X_s\Big) < a_n,$$

which we rearrange to get a bound for the integrand in the first integral in (3.5.6):

$$nX_s^{n-1}\Big(\eta + \Big(-c_\mu + K_n\Big)X_s\Big) < a_n - \frac{n}{2}\Big(c_\mu - K_n + C_n\Big)X_s^n. \tag{3.5.8}$$

For the second integral of (3.5.6), Lemma 3.20 gives the bound

$$n\mathbb{E}\Big[X_s^{n-1}\xi(X)_s\Big] \leq nC_n \sup_{u\leq s}\mathbb{E}[X_u^n].$$

Combining this with (3.5.6) and (3.5.8) and writing $\bar{E}_n(s) = \sup_{u\leq s} E(u)$, we have

$$E_n(t) \leq E_n(0) - \lambda_n' \int_0^t E_n(s)ds + \int_0^t \Big(a_n + \lambda_n''\bar{E}_n(s)\Big)ds,$$

where the constants λ_n' and λ_n'' are given by $\lambda_n' := \frac{1}{2}n(c_\mu - K_n + C_n)$ and $\lambda_n'' := nC_n$. Our assumed condition (3.5.7) gives $\lambda_n' - \lambda_n'' = \frac{1}{2}n\lambda_n > 0$. By Lemma 3.19 ((a)),

$$E_n(t) \leq E_n(0) + \int_0^t e^{-\lambda_n'(t-s)}\Big(a_n + \lambda_n''\bar{E}_n(s)\Big)ds. \tag{3.5.9}$$

Since $\bar{E}$ is non-decreasing and $\lambda_n' > 0$, an integration by parts shows

$$\begin{aligned}
&\int_0^t e^{-\lambda_n'(t-s)}\Big(a_n + \lambda_n''\bar{E}_n(s)\Big)ds \\
&\quad = \frac{(1-e^{-\lambda_n' t})(a_n + \lambda_n''\bar{E}_n(0))}{\lambda_n'} + \frac{\lambda_n''}{\lambda_n'}\int_0^t \Big(1 - e^{-\lambda_n'(t-s)}\Big)d\bar{E}_n(s) \\
&\quad = \frac{(1-e^{-\lambda_n' t})(a_n + \lambda_n''\bar{E}_n(0))}{\lambda_n'} + \frac{\lambda_n''}{\lambda_n'}(\bar{E}_n(t) - \bar{E}_n(0)) - e^{-\lambda_n' t}\int_0^t e^{\lambda_n' s}d\bar{E}_n(s).
\end{aligned}$$

The last expression is a non-decreasing function of t. Hence from (3.5.9) we have

$$\begin{aligned}
\bar{E}_n(t) := \sup_{u\in[0,t]} E(u) &\leq E_n(0) + \sup_{u\in[0,t]}\int_0^u e^{-\lambda_n'(u-s)}\Big(a_n + \lambda_n''\bar{E}_n(s)\Big)ds \\
&\leq \mathbb{E}[X_0^n] + \frac{a_n}{\lambda_n'} + \lambda_n''\int_0^t e^{-\lambda_n'(t-s)}\bar{E}_n(s)ds.
\end{aligned}$$

By Lemma 3.19 (b), since $\lambda_n' > \lambda_n'' > 0$ and $\lambda_n' - \lambda_n'' = \frac{1}{2}n\lambda_n$ for all $t \geq 0$, we have

$$\bar{E}_n(t) \leq \left(\mathbb{E}[X_0^n] + \frac{a_n}{\lambda_n'}\right)\frac{\lambda_n'}{\lambda_n' - \lambda_n''} = \frac{2a_n/n}{c_\mu - K_n - C_n} + \mathbb{E}[X_0^n]\frac{c_\mu - K_n + C_n}{c_\mu - K_n - C_n} < \infty.$$

The theorem follows immediately. □

3.5.5 Moment processes

Proof of Proposition 3.13. (a) Since $X \in \mathcal{S}^2$, the stochastic integral process $X \cdot \widetilde{S}$ is a true martingale. Taking expectation of the equation (3.3.6), we get

$$\begin{aligned} m(t) &= m(0) + \int_0^t \Big(\eta - c_\mu m(s) + \mathbb{E}[\xi(X)_s]\Big) ds + \kappa_2 c_\nu \int_0^t m(s) ds \\ &= m(0) + \int_0^t \Big(\eta - c_0 m(s) + \int_{s-p}^s f_\mu(u-s) m(u) du + \kappa_2 \int_{s-q}^s f_\nu(u-s) m(u) du\Big) ds. \end{aligned}$$

Recalling the definitions of f and c_0 before Proposition 3.13, we have the integral equation

$$m(t) = m(0) + \int_0^t \Big(\eta - c_0 m(s) + \int_{-r}^0 f(u) m(s+u) du\Big) ds.$$

From Theorem 3.3, we know that the function $t \mapsto m(t)$ is càdlàg and hence locally bounded. Since f is integrable, for any $t_1 \leq t_2$, by the dominated convergence theorem,

$$\begin{aligned} \left|\int_{-r}^0 f(u)\Big(m(t_2+u) - m(t_1+u)\Big) du\right| &\leq \int_{t_1-r}^{t_2} \Big|f(u-t_2) - f(u-t_1)\Big| |m(u)| du \\ &\leq \Big(\sup_{u \in [t_1-r, t_2]} |m(u)|\Big) \int \Big|f(u-t_2) - f(u-t_1)\Big| du \to 0, \quad \text{as } |t_2 - t_1| \to 0, \end{aligned}$$

so the function $t \mapsto \int_{-r}^0 f(u) m(t+u) du$ is continuous. Furthermore,

$$|m(t_2) - m(t_1)| \leq |t_2 - t_1| \Big(\eta + c_0 \sup_{u \in [t_1, t_2]} |m(u)| + |f|_{L^1} \sup_{u \in [t_1-r, t_2]} |m(u)|\Big),$$

so the function $t \mapsto m(t)$ is continuous as well. Therefore $t \mapsto m(t)$ is continuously differentiable and the differential equation follows.

(b) The proof is adapted from Section 6.1 of Hale and Lunel [79]. Put $M := \eta/(c_0 - |f|_{L^1})$ and $\widetilde{m} := m - M$ and $\widetilde{\varphi} := \varphi - M$, then clearly $\widetilde{m}$ is the solution to the delay equation

$$\widetilde{m}'(t) = -c_0 \widetilde{m}(t) + \int_{-r}^0 \widetilde{m}(t+u) f(u) du, \quad t \in [0, \infty), \tag{3.5.10}$$

with initial condition $\widetilde{m} = \widetilde{\varphi}$ on $[-r, 0]$. With ζ defined in (3.4.4), we have

$$\widetilde{m}'(t) = \int_0^r \widetilde{m}(t-u) d\zeta(u).$$

For $0 \leq t \leq r$, we can separate the initial condition in (3.5.10) to obtain

$$\widetilde{m}'(t) = \int_0^t \widetilde{m}(t-u) f(u) du + \int_t^r \widetilde{\varphi}(t-u) f(u) du. \tag{3.5.11}$$

Now, since ζ is by construction constant for $t \geq r$, (3.5.11) holds for $t > r$ also. Integrating by parts, we obtain a renewal equation for $\widetilde{m}'$:

$$\widetilde{m}'(t) = \int_0^t \widetilde{m}'(t-u)\zeta(u)du + g(t), \quad t \in [0, \infty), \tag{3.5.12}$$

with initial condition $\widetilde{m}(0) = \widetilde{\varphi}(0)$, where $g(t) := \zeta(t)\widetilde{m}(0) + \int_t^r \widetilde{\varphi}(t-u)f(u)du$. Integrating (3.5.12) and changing the order of integration, we obtain

$$\begin{aligned}
\widetilde{m}(t) - \widetilde{m}(0) &= \int_0^t \int_0^s \zeta(u)\widetilde{m}'(s-u)du\, ds + \int_0^t g(s)ds \\
&= \int_0^t \zeta(u) \int_u^t \widetilde{m}'(s-u)ds\, du + \int_0^t g(s)ds \\
&= \int_0^t \zeta(u)\widetilde{m}(t-u)du - \int_0^t \zeta(u)\widetilde{m}(0)du + \int_0^t g(s)ds.
\end{aligned}$$

Changing variables $u \mapsto t-u$, we arrive at a renewal equation for $\widetilde{m}$:

$$\widetilde{m}(t) = \int_0^t \zeta(t-u)\widetilde{m}(u)du + \widetilde{h}(t), \quad t \in [0, \infty), \tag{3.5.13}$$

with initial condition $\widetilde{m}(0) = \widetilde{\varphi}(0)$. The forcing function, $\widetilde{h}$, given by

$$\begin{aligned}
\widetilde{h}(t) :&= \widetilde{\varphi}(0) - \int_0^t \zeta(u)\widetilde{\varphi}(0)du + \int_0^t g(s)ds \\
&= \widetilde{\varphi}(0) + \int_{-r}^0 (\zeta(t+u) - \zeta(u))\widetilde{\varphi}(-u)du, \qquad (3.5.14)
\end{aligned}$$

is Lipschitz continuous on $[0, r]$ and constant for $t \geq r$ [62, p.18]. Since $\zeta(-r) = -M$, substituting $\widetilde{m} = m + \eta/\zeta(r)$ and $\widetilde{\varphi} = \varphi + \eta/\zeta(r)$ back into (3.5.13) and (3.5.14) completes the computations. □

Proof of Theorem 3.14. (a) Let M, $\widetilde{m}$ and $\widetilde{\varphi}$ be as defined in the proof of Proposition 3.13 (b). The *characteristic function* Δ of (3.5.10) defined in (3.2.5) is given by

$$\Delta(z) = z + c_0 - \int_{-r}^0 e^{zu} f(u)du.$$

It's clear from (3.2.6) that if $\Delta(z)$ is root free in the right half-plane $\{z|\text{Re } z > 0\}$, then all solutions $\widetilde{m}$ of the functional differential equation (3.5.10) converge to zero exponentially fast as $t \to \infty$.

For sufficiency, it is enough to show that $c_0 > |f|_{L^1}$ implies $\Delta(z) \neq 0$ for any z with Re $z \geq 0$. Let $z = \alpha + i\beta$ where $\alpha \geq 0$. Then the real part of Δ can be written as

$$\text{Re } \Delta(z) = \alpha + c_0 - \int_{-r}^0 e^{\alpha u} \cos(\beta u) f(u)du$$

Since $e^{\alpha u}$ and $\cos(\beta u)$ are no greater than 1 on $[-r, 0]$, we have Re $\Delta(z) \geq \alpha + c_0 -$

$\int_{-r}^{0} f(u)du > 0$, whenever $c_0 > |f|_{L^1}$, so Δ is root free on $\{z|\text{Re } z \geq 0\}$. For necessity, the expansion (3.2.6) implies that 0 is the only possible limit of $\widetilde{m}(t)$, which gives the uniqueness of m as a limiting mean. Since we require this limit to be positive, necessarily we require $c_0 > |f|_{L^1}$.

(b) Suppose that $\varphi \equiv M$ where M is defined in (3.4.5), and assume $c_0 > |f|_{L^1}$ so that $M > 0$. Then $\tilde{\varphi}$ is identically zero on $[-r, 0]$, and the function h defined in (3.5.14) is identically zero on $[-r, \infty)$. From (3.5.13), the centered mean process $\widetilde{m}(\cdot)$ satisfies satisfies the homogeneous renewal equation

$$\widetilde{m}(t) = \int_0^t \zeta(u)\widetilde{m}(t-u)du.$$

Applying the representation in Theorem 2.12 of Diekmann et al. [62] shows that the only solution to this renewal equation is $\widetilde{m}(t) = 0$ for all $t \in [0, \infty)$. This gives $m(t) = M$ for all $t \in [-r, \infty)$. Conversely, suppose that for all $t \in [0, \infty)$, $m(t) = M$ for some positive M. Then (3.4.3) gives $0 = \eta + M(-c_0 + |f|_{L^1})$, which implies that M is uniquely given by (3.4.5) and $c_0 > |f|_{L^1}$. Recall that the delay equation (3.4.3) has a unique solution once the initial condition φ is fixed. Therefore the solution $m \equiv M$ for all $t \geq 0$ then corresponds uniquely to the initial condition $\varphi \equiv M$ on $[-r, 0]$, and the proof is complete. $\square$

Proof of Theorem 3.15. Let Z be an $\mathcal{F}_t$ measurable random variable with $\mathbb{E}[Z^2X_t^2] < \infty$ for any $t > 0$. Since X has finite variation, for any $t > 0$ and $u > r$, we have

$$X_{t+u} = ZX_{t+r} + \int_{t+r+}^{t+u} ZdX_s = ZX_{t+r} + \int_{t+r+}^{t+u} Z\Big\{\Big(\eta - c_0X_s + \xi(X)_s\Big)ds + c_\nu X_{s-}d\widetilde{S}_s\Big\}.$$

Taking expectations and using Fubini's theorem gives

$$\begin{aligned}\mathbb{E}[ZX_{t+u}] = \mathbb{E}[ZX_{t+r}] + \eta(u-r)\mathbb{E}[Z] - c_0\int_{t+r}^{t+u} \mathbb{E}[ZX_s]ds + \int_{t+r}^{t+u}\int_{s-r}^{s} \mathbb{E}[ZX_u]f(u-s)duds\\ + \int_{t+r}^{t+u} \mathbb{E}\Big[Z\int_{s-q+}^{s} X_{u-}f_\nu(u-s)d\widetilde{S}_u\Big]ds + c_\nu\mathbb{E}\Big[\int_{t+r+}^{t+u} ZX_{s-}d\widetilde{S}_s\Big].\end{aligned}$$

Since Z is $\mathcal{F}_t$ measurable and hence $\mathcal{F}_{s-q+}$ measurable for any $s \geq t+r$, the two stochastic integrals in the last expression have zero expectation. Therefore

$$\mathbb{E}[ZX_{t+u}] = \mathbb{E}[ZX_{t+r}] + \eta(u-r)\mathbb{E}[Z] - c_0\int_{t+r}^{t+u} \mathbb{E}[ZX_s]ds + \int_{t+r}^{t+u}\int_{-r}^{0} \mathbb{E}[ZX_{s+u}]f(u)duds,$$

from which we obtain a functional differential equation,

$$\frac{d}{du}\mathbb{E}[ZX_{t+u}] = \eta\mathbb{E}[Z] - c_0\mathbb{E}[ZX_{t+u}] + \int_{-r}^{0} \mathbb{E}[ZX_{u+u}]f(u)duds.$$

We note that this is a functional differential equation of a similar form as we dealt with in Theorem 3.14 (a). Since we assumed $c_0 > |f|_{L^1}$, we can compute the characteristic function Δ of this equation like we did in Theorem 3.14 (a), which allows us to establish the convergence to a limiting mean. Note that the limiting mean is given by

$$\mathbb{E}[ZX_{t+u}] \to \frac{\eta \mathbb{E}[Z]}{c_0 - |f|_{L^1}} = M\mathbb{E}[Z]$$

and the convergence is exponential. $\square$

Proof of Corollary 3.17. (a) Since $\widetilde{Y}_t = \int_{t-1+}^{t} \sqrt{X_{s-}} dL_s$ and L has zero mean, we have $\mathbb{E}[\widetilde{Y}_t] = 0$ and $\operatorname{Cov}(\widetilde{Y}_t, \widetilde{Y}_{t+u}) = \kappa_2 \int \mathbb{1}_{[t-1,t]}(s)\mathbb{1}_{[t+u-1,t+u]}(s)\mathbb{E}[X_s]ds = \kappa_2 M(1-u)_+$.

(b) Write $\kappa_1 := \int_{\mathbb{R}_0} z\Pi_L(dz)$. Since $dY_t = -\kappa_1\sqrt{X_t}dt + \int_{\mathbb{R}_0} \sqrt{X_{t-}} z\widetilde{N}(dz, dt)$, by Ito's lemma, it holds that $dY_t^2 = -2Y_t\kappa_1\sqrt{X_t}dt + Y_t^2 - Y_{t-}^2$, where

$$Y_t^2 - Y_{t-}^2 = \int_{\mathbb{R}_0} \left((Y_{t-} + \sqrt{X_{t-}} z)^2 - Y_{t-}^2 \right) N(dz, dt) = \int_{\mathbb{R}_0} \left(2Y_{t-}\sqrt{X_{t-}} z + X_{t-}z^2 \right) N(dz, dt).$$

Then, since $\widetilde{Y}_{t+u}^2 = (Y_{t+u} - Y_{t+u-1})^2 = Y_{t+u}^2 - Y_{t+u-1}^2 - 2Y_{t+u-1}(Y_{t+u} - Y_{t+u-1})$, we have

$$\widetilde{Y}_{t+u}^2 = \int_{t+u-1+}^{t+u} 2Y_{s-}\sqrt{X_{s-}} dL_s + \int_{t+u-1+}^{t+u} \int_{\mathbb{R}_0} X_{t-}z^2 N(dz, dt) - 2Y_{t+u-1} \int_{t+u-1+}^{t+u} \sqrt{X_{s-}} dL_s.$$

Now suppose $u > 1$ so that $\widetilde{Y}_t$ is $\mathcal{F}_{t+u-1}$ measurable. Taking expectations, we obtain

$$\mathbb{E}[\widetilde{Y}_t^2 \widetilde{Y}_{t+u}^2] = \kappa_2 \int_{t+u-1}^{t+u} \mathbb{E}[\widetilde{Y}_t^2 X_s] ds.$$

By Theorem 3.15 and Corollary 3.17 (a), $\mathbb{E}[\widetilde{Y}_t^2 X_u] \to \kappa_2 M^2$ exponentially fast as $u \to \infty$, i.e. there exist constants C and $T, \lambda > 0$ such that $\left|\mathbb{E}[\widetilde{Y}_t^2 X_u] - \kappa_2 M^2\right| \le Ce^{-\lambda u}$, for all $u > T$. Therefore

$$\begin{aligned}
\left|\mathbb{E}[\widetilde{Y}_t^2 \widetilde{Y}_{t+u}^2] - \kappa_2^2 M^2\right| &\le \kappa_2 \int_{t+u-1}^{t+u} \left|\mathbb{E}[\widetilde{Y}_t^2 X_s] - \kappa_2 M^2\right| ds \\
&\le \kappa_2 \int_{t+u-1}^{t+u} Ce^{-\lambda s} ds = \frac{\kappa_2 C e^{1-\lambda t}}{\lambda} e^{-\lambda u},
\end{aligned}$$

for all $u > T$, which finishes the proof. $\square$

Chapter 4

CLT for Spiked Eigenvalues of Sample Auto-covariance Matrices

4.1 Introduction

This chapter focuses on the asymptotic theory of the high dimensional factor model we introduced in Section 1.4 of Chapter 1. We work under a high-dimensional setting where both the dimension p and the sample size T tend to infinity simultaneously. In this high dimensional setting, most asymptotic results from the finite p setting no longer apply, which motivated [10, 11, 12, 104, 105, 133] and more recent works [110, 111] to develop an appropriate asymptotic theory. To set the context for our current work, we begin with an overview of these results in Section 4.1.1. We observe that some of these works are naturally related to the theory of random matrices, especially the study of large spiked covariance matrices. We therefore take a detour and introduce some elements of random matrix theory in Section 4.1.2 before returning to the discussion of factor models in Section 4.1.3. Finally we give an overview of our current work in Section 4.1.4.

4.1.1 High Dimensional Factor Models

The seminal work [12] provides a consistent estimator for the number of factors K of a static factor model by minimizing a penalized loss function. Under mostly the same setting [10] analyses the factor model via PCA and establishes asymptotic normality of estimated common components. A quasi-maximum likelihood method is proposed in [11] to estimate the factors and the authors establish the asymptotic normality of the estimators. We observe that the methods in the three aforementioned works do not explicitly model the auto-covariance structure of the factor model. The PCA method in [10] is essentially based on the covariance matrix instead of auto-covariance matrices and the other two approaches do not model the serial correlations directly neither.

An alternative perspective stems from the idea that when strong serial correlation is exhibited by the data, it is more natural to analyse the model using the sample auto-covariance matrices. This approach is explored in [133] and more recently in [104, 105]. Define the matrix $M := \sum_{\tau=1}^{\tau_0} \Sigma_\tau \Sigma_\tau^\top$, where Σ_τ is the lag-τ (population) auto-covariance matrix of the time series and τ_0 is some chosen constant. It is not difficult to see that the eigenvalues and eigenvectors of M capture a lot of information on the structure of the factor model. Indeed, observe that a vector x belongs to the kernel of M if and only if $\Sigma_\tau x = 0$ for all $\tau \leq \tau_0$, and it can be shown that this happens if and only if x is in the orthogonal complement of the factor loading matrix. This implies that the factor loading space is in fact spanned by the eigenvectors of M corresponding to non-zero eigenvalues. Therefore by estimating the eigenvectors of the matrix M, we are effectively estimating the factor loading space of the model. Moreover, the matrix M contains exactly K non-zero eigenvalues which represent the strength of the K factors in the model. This implies that the ratio μ_{i+1}/μ_i, where μ_i is the i-th largest eigenvalue of M, $i = 1, \ldots, p$ is equal to zero for all $i > K$. Therefore by estimating the number of non-zero ratios we obtain an estimate for the number of factors.

In [104, 105] the authors consider a factor model where the strength of the factors, or equivalently the eigenvalues $\{\mu_i\}$ of M tend to infinity as $T \to \infty$. More specifically, it is assumed that each μ_i diverges at a rate of $p^{1-\delta}$ for $i \leq K$ where $\delta \in [0, 1]$ is a constant. Under this assumption [104] proposes a ratio-based estimator $\widehat{K} := \arg\min_{i \leq p} \lambda_{i+1}/\lambda_i$, where λ_i is the i-th largest eigenvalue of some estimate $\widehat{M}$ of M. The authors did not obtain explicit asymptotic results for $\widehat{K}$ but instead developed the asymptotic theory for the estimators $\widehat{\lambda}_i$. Using the asymptotic properties of $\widehat{\lambda}_i$ as well as empirical results the authors argued that $\widehat{K}$ is a good estimator for the number of factors K. A similar setting is considered in [105] in which the authors provide estimators of the factor loading space using the eigenvectors $\widehat{M}$ and establish the asymptotic theory.

As can be seen from the above discussion, accurately estimating the eigenvalues $\{\mu_i\}$ is of paramount importance in the analysis of factor models. This naturally brings the discussion to the corresponding asymptotic theory of empirical eigenvalues $\{\lambda_i\}$, which is the main focus of our current work. As observed in [104], when p diverges at the same time as T, empirical eigenvalues are no longer consistent estimators of true eigenvalues. In the case of [104], the authors obtain a rate of $|\lambda_i - \mu_i| = O_p(p^{2-\delta}T^{-1/2})$ for $i \leq K$, which is directly related to not only T but the dimension p as well. A similar type of result is obtained in [105] for estimates of the factor loadings matrix. We remark that this type of results, especially the observation that the rate depends explicit on the dimension p, is a known phenomenon in high dimensional statistics and its closely related field, large dimensional random matrix theory. We will take on this perspective and examine these results in the context of random matrix theory.

To see the connection to high dimensional statistics and random matrix theory, we observe that the spectrum of the matrix M discussed above consists of K large (in fact, diverging) eigenvalues while the rest of the eigenvalues are small and similar in size. This is an example of the so-called spiked covariance model which has been an area of extensive research since the pioneering work of [91]. The theory of spiked covariance matrices is by now very well developed and plays an essential role in the asymptotic theory of principal component analysis in a high dimensional setting. We refer to the monograph [14] for a detailed treatise of the topic and [90] for a survey of recent developments and related topics. On the other hand, the topic of spiked auto-covariance matrix has only recently started to attract attention and the literature is still relatively sparse. We will therefore first discuss some simple spiked covariance models to outline the landscape of the theory before returning to the topic of factor models.

4.1.2 Spiked Sample Covariance Matrices

To set the scene, suppose $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is a $p \times n$ matrix of random variables with p interpreted as the dimension of the model and n the sample size. Assuming the columns $(\mathbf{x}_i)$ are independent and identically distributed (i.i.d.), let $\Sigma := \mathbb{E}[\mathbf{x}_1\mathbf{x}_1^\top]$ be the (population) covariance matrix and $\Sigma = U\Lambda U^\top = U\mathrm{diag}(\mu_1, \ldots, \mu_p)U^\top$ be its spectral decomposition where the eigenvalues $\{\mu_i\}$ are arranged in non-increasing order. A spiked covariance matrix informally refers to the assumption that $\mu_1, \ldots, \mu_K$ are 'much bigger' than the rest of the eigenvalues for some number K, which can be unknown or even diverging as well. The rest of the eigenvalues are assumed to be small and of 'similar' size. The case where $K = 0$ and all eigenvalues are of similar size is referred to as the null case.

We will focus on a particular asymptotic regime where p and n diverge simultaneously and $p/n \to c$ for constant $c > 0$. The cases where $p/n \to 0$ or $p/n \to \infty$ are interesting as well, some recent developments include [53, 54, 97, 67, 36]. We also refer to [90] for a survey of some relevant results. Let $\{\lambda_i\}_{1\le i\le n}$ be the eigenvalues of the sample covariance matrix $\widehat{\Sigma} := n^{-1}XX^\top$. It is easy to show that one cannot in general expect λ_i to be consistent estimators of μ_i, even in very simple cases. For an immediate counter-example, consider the simple case where $p > n$, $K = 0$ and population covariance matrix is given by $\Sigma = I_p$. Then clearly $\widehat{\Sigma}$ is a matrix of rank at most n and therefore contains at least $p - n$ zero eigenvalues which do not converge to the true value one 1. Similarly, in general, the largest eigenvalue $\widehat{\lambda}_i$ of $\widehat{\Sigma}$ does not converge to its theoretical counterpart either.

In the null case where $K = 0$, the asymptotic properties of the spectrum of $\widehat{\Sigma}$ are best captured by its limiting spectral distribution (LSD). Put $F := p^{-1}\sum_{i=1}^{p}\delta_{\lambda_i}$ where δ_a is the Dirac measure at a, i.e. F is the empirical measure of eigenvalues of $\widehat{\Sigma}$. Then F is a random probability measure on $\mathcal{B}(\mathbb{R})$ and is known as the empirical spectral distribution (ESD) of $\widehat{\Sigma}$. Under certain regularity conditions, the measure F converges almost surely in the weak topology of measures to a deterministic probability measure known as the Marčenko-Pastur distribution. As a consequence, it can be shown that the k-th largest eigenvalue λ_k converges a.s. to the right end-point of the support of the Marčenko-Pastur distribution for any fixed k. For more details on the LSD of covariance matrices and the Marčenko-Pastur law we refer to [14] and [90].

In the case where $K > 0$, i.e. where spiked eigenvalues are present, even more interesting phenomena occur. The number $\gamma := 1 + \sqrt{c}$ functions as a critical threshold in the sense that spikes below this threshold cannot be distinguished from the non-spiked eigenvalues asymptotically. More specifically, supposing $\mu_i < \gamma$, under certain regularity conditions it can be shown that λ_i converges a.s. to the right end point of the support of the Marčenko-Pastur distribution at a convergence rate of $n^{-2/3}$. On the other hand, if $\mu_i > \gamma$, then λ_i has an a.s. limit equal to $\mu_i + c\mu_i/(\mu_i - 1)$ at a convergence rate of $n^{-1/2}$. In particular, this limit is outside the support of the Marčenko-Pastur distribution, allowing the spike to be distinguished from the non-spiked eigenvalues.

Taking a closer inspection at the difference in convergence rates discussed above, it is natural to postulate that besides tending to different limits, empirical eigenvalues could exhibit different types of limiting distributions depending on the size of their population counterparts in relation to the threshold. Indeed, for eigenvalue μ_i below the threshold, it can be shown that after centering by the right end-point of the Marčenko-Pastur law and scaling by $n^{2/3}$, the empirical eigenvalue λ_i tends to the Tracy-Widom distribution. On the other hand, for μ_i above the threshold, the estimator λ_i is instead asymptotically Gaussian with the usual scaling of $n^{1/2}$. This dichotomy of asymptotic behaviours of estimated eigenvalues is known as the Baik/Ben Arous/Péché (BBP) phase transition, named after the authors of the pioneering work [17].

Since [17], the phase transition phenomenon and related results were investigated in various different settings; see [16, 36, 51, 92, 93] and the references therein. Both the case above and below the transition threshold have attracted significant research from probabilists, mathematicians, statisticians and mathematical physicists. Recent works in the latter direction such as [15, 19, 52, 135, 165] manage to establish the CLT for spiked eigenvalues of the covariance matrix under quite general settings. These theoretical results are of great importance in the study of the asymptotic behaviours of the PCA. Our current work fits under this area - we aim to establish the asymptotic normality for spiked eigenvalues of the auto-covariance matrix.

The phase transition phenomenon immediate illustrates the striking differences between high dimensional asymptotic theory and the traditional, fixed dimensional theory. In the high dimensional setting, empirical eigenvalues are not consistent estimators of true eigenvalues, and the bias as well as the asymptotic distribution of λ_i depends on whether the true eigenvalue μ_i is above or below the phase transition threshold. In particular, only eigenvalues above the threshold can be detected asymptotically, while the rest are indistinguishable from each other. This has immediate implications for the estimation of the number of factors, since only factors with strength above the phase transition threshold can be detected asymptotically.

4.1.3 Spiked Sample Auto-Covariance Matrices

To conclude the above introduction on spiked covariance matrices, we remark that the asymptotic theory for the spectral structure of spiked covariance matrices is very well developed. In contrast, the theory around spiked auto-covariance matrices (or more precisely of the matrix $\widehat{\Sigma}_\tau\widehat{\Sigma}_\tau^\top$ discussed above) is not yet well understood. Nevertheless, due to its connection and importance in the analysis of high dimensional factor models and time series, this topic has been gaining attention. Recent work including [26, 38] and [110, 111, 161] are a first step toward a better and more systematic understanding of the asymptotic properties of spiked auto-covariance matrices.

Mirroring the theory on spiked covariance matrices, the first step in the analysis is to study the empirical distribution of sample eigenvalues. Towards this, the limiting spectral distribution for $\widehat{\Sigma}_\tau\widehat{\Sigma}_\tau^\top$ in the null case ($K = 0$ number of factors) was studied recently [110] and [161] using the Stieltjes transform and moment methods respectively. Results of a similar type were obtained in [38] based on the theory of non-commutative probability. The phase transition phenomena and the limits of eigenvalues of $\widehat{\Sigma}_\tau\widehat{\Sigma}_\tau^\top$ in a spiked model were first established in the recent work [111]. The setting of [111] is based on the factor model proposed by [104] but assumes that all spiked eigenvalues are finite instead of diverging. In [111], the authors give a precise description of the asymptotic property of the ratio λ_i/λ_{i+1}, which is used in [104] to estimate the number of factors K. As a consequence of the developed theory, the authors in [111] are able to propose a strongly consistent estimator for K, which is a sizable improvement upon the original methodology and results of [104].

Based on the discussions above, the asymptotic properties of $\widehat{\Sigma}_\tau\widehat{\Sigma}_\tau^\top$ studied in the literature so far exhibit a certain resemblance to those of the spiked covariance matrix. The actual form of the LSD and phase transition threshold obviously differ from those of the covariance matrix, but there is a clear parallel in the type of behaviours observed in large dimensional random matrices. It is therefore interesting, from both a theoretical and a practical perspective, to identify what other important features are common between the covariance and the auto-covariance matrix, as well as what features are unique to the auto-covariance matrix. On the other hand, due to the presence of temporal correlation in the data and the more complex structure of the matrix $\widehat{\Sigma}_\tau\widehat{\Sigma}_\tau^\top$, these results are much harder to establish in the case of the auto-covariance matrix and new techniques need to be invented.

In our work we focus on one such feature - the asymptotic normality of spiked eigenvalues whose population counterparts are above the transition threshold. The asymptotic distributions of the eigenvalues of covariance matrices proved to be of great importance in the asymptotic theory of the PCA, as shown by [52, 165]. Analogously, the theory of asymptotic distributions for eigenvalues of spiked auto-covariance matrices could provide a powerful tool in studying the asymptotics of high dimensional factor models. To the best extent of our knowledge however, identifying the limiting distribution of leading eigenvalues in the case of the auto-covariance matrix remains an open topic. Based on what is known for the covariance matrix and the work of [111], it is reasonable to suspect that the leading eigenvalues are asymptotically Gaussian with the usual scaling of $T^{1/2}$. Indeed, in our current work we will establish this result under quite general conditions.

4.1.4 Overview of our work

We now give an overview of our settings and the contributions of our work. Our setting is based on the factor model studied in [104, 111] and we will too be working in a high-dimensional setting where p and T diverge at the same time such that $p/T \to c \in (0, \infty)$. The main object of our study is the symmetrized lag-τ sample auto-covariance matrix $\widehat{\Sigma}_\tau \widehat{\Sigma}_\tau^\top$ and its eigenvalues λ_i, in particular, we will establish the asymptotic normality of λ_i under appropriate conditions.

Similar to [104] and [105] we assume that the factor strength $\{\mu_i\}_{i \le K}$ diverges as $T \to \infty$. We remark that most results on spiked sample covariance matrices assume that the spikes are bounded as $p, T \to \infty$. On the other hand, in the context of factor modelling, it is perhaps more natural to consider situations where the factor strengths are divergent, as seen in the recent literature [105, 104, 11]. The authors in [104, 105] assumes that each μ_i diverges at a specified rate of $p^{1-\delta}$ where $\delta \in [0, 1]$ and argue that this choice is in fact quite natural. We relax this assumption and allow μ_i to diverge at any arbitrary rate instead of as a specified function of p, T. As a consequence our results are applicable to a much wider range of cases where the factors are not as strong. Additionally, we also allow the number of factors K to be possibly diverging as $T \to \infty$. This type of assumption has been made in the literature for covariance matrices in [52, 136] but has not been incorporated into the factor model setting. Lastly, we consider both the standard case where the lag τ in the auto-covariance matrix $\widehat{\Sigma}_\tau \widehat{\Sigma}_\tau^\top$ is a fixed constant, as well as the new asymptotic regime where τ is diverging as well. It will be shown that in these two cases, the scalings for the central limit theorems are not of the same order. Consequently, if one is interested in the eigenvalues of $\widehat{\Sigma}_\tau \widehat{\Sigma}_\tau^\top$ for a moderately large τ, the central limit theorem in the regime where $k \to \infty$ might provide a more accurate result.

A major source of difficulty in our setting is that we do not impose any restriction on the rate of divergence of the factor strength μ_i. We argue that strong assumptions on the speed of μ_i such as ones used in [104, 105] essentially reduce the analysis of a high-dimensional factor model to the study of just the factors, which is a low dimensional problem (see the remarks below Theorem 4.2). While this aligns with the goals of dimension reduction in [104, 105], it obfuscates some interesting features otherwise seen in high-dimensional models. Without such assumptions, the idiosyncratic noise is no longer negligible and we obtain a clearer picture of how the high-dimensional noise accumulates and affects the location of eigenvalues. More specifically, even though λ_i is close to μ_i asymptotically, it will be shown that the speed of convergence rate of $\lambda_i - \mu_i$ (after appropriate scaling) is in general slower than $T^{-1/2}$, i.e. we will not be able to obtain a CLT using μ_i as the centering term. What happens here is that the bias of λ_i decays too slowly for the purpose of obtaining a CLT and a more accurate centering is need. In our work this centering term will be defined implicitly as the solution to an equation. The phenomenon described above is more common in the random matrix literature where there is less emphasis on reducing high-dimensional models to low dimensional ones, see for instance [52].

Lastly, instead of working with the auto-covariance matrix $\widehat{\Sigma}_\tau$, we deal only with the symmetrized version $\widehat{\Sigma}_\tau \widehat{\Sigma}_\tau^\top$ in our analysis. The matrix $\widehat{\Sigma}_\tau \widehat{\Sigma}_\tau^\top$ does not factor into a matrix with independent entries like the covariance matrix $\widehat{\Sigma}$ does. Consequently the central ideas of works like [15] and [52] are not applicable in our case and we need a new approach to establish asymptotic normality. The approach we develop here could potentially be applied to other types of products of covariance type matrices.

The rest of the chapter is organized as follows. Section 4.2 introduces the setting and assumptions of our work, sets up the relevant notations and presents some preliminary results. The results of our work are given in Section 4.3. In Section 4.3.1 we investigate the asymptotic location of empirical eigenvalues and construct an accurate centering for these eigenvalues. The central limit theorem for the empirical eigenvalues, which is the main result of our work, is given in Section 4.3.2. The proof of the CLT is quite involved and is thus divided into a series of intermediate results collected in Section 4.4 and technical lemmas collected in Section 4.5. We give a summary of the strategy of the proof in Section 4.3.2 and explain how the intermediate results are used to obtain the CLT.

4.2 The Setting

As discussed in the introduction, in this work we study a high-dimensional time series arising from a factor model considered in [104, 105, 111]. Suppose $(\mathbf{y}_t)_{t=1,\ldots,T} \subseteq \mathbb{R}^{K+p}$ is a $K+p$ dimensional stationary time series consisting of K factors, observed over a time period of length T. Here the choice of writing $K+p$ for the dimension of the time series is purely for notational convenience in our exposition and proofs. Then we may write

$$\mathbf{y}_t = L\mathbf{f}_t + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T, \tag{4.2.1}$$

where the $K \times T$ matrix $(\mathbf{f}_t)_{t=1,\ldots,T}$ contains K independent factors, each assumed to be a stationary time series. The matrix L is the $(p+K) \times K$ factor loading matrix and $\boldsymbol{\epsilon}_t$ is a

$K+p$ dimensional idiosyncratic noise time series to be specified below.

It is well-known that the factor model (4.2.1) is not identifiable without additional constraints on L and $\mathbf{f}_t$. There are multiple ways to impose such constraints, see Table 1 of [11] for a discussion and comparisons between different constraints found in the literature. The constraint we chose to work with, mainly for notational convenience, is to assume $L^\top L$ is equal to a diagonal matrix and all factors are standardized, i.e. $\mathbb{E}[f_{it}] = 0$ and $\mathbb{E}[f_{it}^2] = 1$ for all $i = 1, \ldots, K$ and $t = 1, \ldots, T$.

We work in a high-dimensional setting where p and T diverge simultaneously and the ratio p/T tends to a constant $c > 0$ as $T \to \infty$. We allow the number of factors K to diverge as $T \to \infty$, but impose conditions on the speed of its divergence so that the number of factors remains small in comparison to the dimension of the entire observation (see Assumption 4.2 and Assumption 4.3).

Each factor $(f_{it})_t$ is assumed to be a stationary time series of the form

$$f_{it} = \sum_{l=0}^{\infty} \varphi_{il} z_{i,t-l}, \quad i = 1, \ldots, K, \quad t = 1, \ldots, T, \tag{4.2.2}$$

where the random variables (z_{it}) are i.i.d. with zero mean, unit variance and finite $(4+\epsilon)$-th moment for some small $\epsilon > 0$. Under this setup, the constraint $\mathrm{Var}(f_{it}) = 1$ mentioned above directly translates to the constraint $\|\boldsymbol{\varphi}_i\|_{\ell_2} = 1$ where $\boldsymbol{\varphi}_i := (\varphi_{il})_l$ is the vector of coefficients for the i-th factor and $\|\cdot\|_{\ell_2}$ is the ℓ_2 norm on sequence spaces. Write $\gamma_i(\tau) := \mathbb{E}[f_{i,1} f_{i,\tau+1}]$ for the population lag-τ auto-covariance of the i-th factor time series $\mathbf{f}_i$. Then clearly $\gamma_i(\tau)$ can be written as

$$\gamma_i(\tau) := \mathbb{E}[f_{i,1} f_{i,\tau+1}] = \sum_{l=0}^{\infty} \varphi_{i,l} \varphi_{i,l+\tau}. \tag{4.2.3}$$

In general, the loading matrix L is important in the analysis of the factor model as it appears in the (population) covariance and auto-covariance matrices of $\mathbf{y}_t$. However, the recent work [111] makes an important observation that under additional Gaussian assumptions on the error time series $\boldsymbol{\epsilon}_t$, the factor model can be reduced to a canonical form where $L = \begin{pmatrix} I_K & \mathbf{0}_{K\times p} \end{pmatrix}^\top$. The authors of [111] are able to obtain explicit results on the phase transition of leading eigenvalues under this assumption. As previously mentioned, for notational convenience we employ a slightly different normalization for the matrix L. Nevertheless, we argue that under Gaussian assumptions on the error $\boldsymbol{\epsilon}_t$, the factor model can be reduced to a canonical form where L takes the form

$$L = \begin{pmatrix} \mathrm{diag}(\sigma_1, \ldots, \sigma_K) \\ \mathbf{0}_{p\times K} \end{pmatrix},$$

where $(\sigma_1, \ldots, \sigma_K)$ is a sequence of positive real numbers. For the completeness of our exposition, we give a detailed explanation of this simplification.

Recalling our constraint on $L^\top L$ being a diagonal matrix, without loss of generality we can assume $L^\top L := \mathrm{diag}(\sigma_1^2, \ldots, \sigma_K^2)$ where $(\sigma_1, \ldots, \sigma_K)$ is a sequence of positive numbers. Clearly the $(p+K) \times K$ matrix $\overline{L} := L\ \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_K^{-1})$ satisfies $\overline{L}^\top \overline{L} = I_K$, thus there exists a $(p+K) \times p$ matrix $\underline{L}$ with orthogonal columns such that $\widetilde{L} := (\overline{L}, \underline{L})$ is an orthogonal matrix. Recall from (4.2.1) that $\mathbf{y}_t = L\mathbf{f}_t + \boldsymbol{\epsilon}_t$. Define

$$\mathbf{z}_t := \widetilde{L}^\top \mathbf{y}_t = \begin{pmatrix} \overline{L}^\top \\ \underline{L}^\top \end{pmatrix} L\mathbf{f}_t + \widetilde{L}^\top \boldsymbol{\epsilon}_t = \begin{pmatrix} \overline{L}^\top \\ \underline{L}^\top \end{pmatrix} \overline{L}\ \mathrm{diag}(\sigma_1, \ldots, \sigma_K)\ \mathbf{f}_t + \widetilde{L}^\top \boldsymbol{\epsilon}_t.$$

By definition we clearly have $\overline{L}^\top \overline{L} = I_K$ and $\underline{L}^\top \overline{L} = \mathbf{0}_p$, therefore

$$\mathbf{z}_t = \widetilde{L}^\top \mathbf{y}_t = \begin{pmatrix} \mathrm{diag}(\sigma_1, \ldots, \sigma_K) \\ \mathbf{0}_{p \times K} \end{pmatrix} \mathbf{f}_t + \widetilde{L}^\top \boldsymbol{\epsilon}_t. \tag{4.2.4}$$

Note that $\mathbf{z}_t$ is simply the original data $\mathbf{y}_t$ subjected to an orthogonal transformation, in particular, the sample auto-covariance matrix of $(\mathbf{z}_t)$ contains the same information as that of $(\mathbf{y}_t)$. More precisely, define the sample auto-covariance matrices

$$\Sigma_{\mathbf{y}} := \frac{1}{T} \sum_{t=1}^{T-\tau} \mathbf{y}_{t+\tau} \mathbf{y}_t^\top, \quad \Sigma_{\mathbf{z}} := \frac{1}{T} \sum_{t=1}^{T-\tau} \mathbf{z}_{t+\tau} \mathbf{z}_t^\top = \widetilde{L}^\top \Sigma_{\mathbf{y}} \widetilde{L}.$$

It is easy to see that the spectrum of $\Sigma_{\mathbf{y}} \Sigma_{\mathbf{y}}^\top$ coincides with that of $\Sigma_{\mathbf{z}} \Sigma_{\mathbf{z}}^\top$. Indeed, we have

$$\Sigma_{\mathbf{z}} \Sigma_{\mathbf{z}}^\top = \widetilde{L}^\top \Sigma_{\mathbf{y}} \widetilde{L}\ \widetilde{L}^\top \Sigma_{\mathbf{y}}^\top \widetilde{L} = \widetilde{L}^\top \Sigma_{\mathbf{y}} \Sigma_{\mathbf{y}}^\top \widetilde{L},$$

where $\widetilde{L}$ is orthogonal so a conjugation by $\widetilde{L}$ does not affect the spectrum $\Sigma_{\mathbf{y}} \Sigma_{\mathbf{y}}^\top$.

Recall that the main goal of our work is to establish the asymptotic distribution of the leading eigenvalues of $\Sigma_{\mathbf{y}} \Sigma_{\mathbf{y}}^\top$. By the above arguments, it suffices to consider $\Sigma_{\mathbf{z}} \Sigma_{\mathbf{z}}^\top$ instead of $\Sigma_{\mathbf{y}} \Sigma_{\mathbf{y}}^\top$, that is, we may without any loss of generality assume that

$$\mathbf{y}_t = \begin{pmatrix} \mathrm{diag}(\sigma_1, \ldots, \sigma_K) \\ \mathbf{0}_{p \times K} \end{pmatrix} \mathbf{f}_t + \widetilde{L}^\top \boldsymbol{\epsilon}_t.$$

Finally, when $\boldsymbol{\epsilon}_t$ is assumed to be standard Gaussian and hence unitarily invariant, the transformed error $\widetilde{L}^\top \boldsymbol{\epsilon}_t$ is equal in distribution to $\boldsymbol{\epsilon}_t$. Under this assumption, we have

$$\mathbf{y}_t \stackrel{\text{dist.}}{=} \begin{pmatrix} \mathrm{diag}(\sigma_1, \ldots, \sigma_K) \\ \mathbf{0}_{p \times K} \end{pmatrix} \mathbf{f}_t + \boldsymbol{\epsilon}_t \tag{4.2.5}$$

and we may take this as the canonical form of the factor model 4.2.1. Motivated by these observations, we will work under the assumption that $\epsilon_{it} \stackrel{\text{iid}}{\sim} N(0, 1)$ for all $i = 1, \ldots, K$,

$t = 1, \ldots, T$ and start with the canonical form (4.2.5).

4.2.1 Assumptions

Observe that under the canonical representation (4.2.5), the sequence $(\sigma_1, \ldots, \sigma_K)$ are in fact the standard deviations of the factors. We adopt similar assumptions as in [104] and assume that every $\sigma_i \to \infty$ as $p \to \infty$, i.e. the strength of factors in the model is much stronger than that of the noise. In [104] it is assumed that all σ_i diverge at the specific rate $p^{1-\delta}$ where $\delta \in [0, 1]$ is fixed. We do not impose such strong assumptions and instead allow σ_i to diverge at any rate, no matter how slow. The only restriction we impose is that all factors are asymptotically equal in strength, i.e. there exists a constant $C > 0$ such that $\sigma_i/\sigma_j \le C$ for any $i, j = 1, \ldots, K$ and $T > 0$. As a result, our result is applicable to a much wider range of situations where the factors are not as strong as required by [104].

Recall $\gamma_i(\tau) := \mathbb{E}[f_{i,1} f_{i,\tau+1}]$ from (4.2.3). Under the canonical form (4.2.5), the (population) lag-τ auto-covariance function for each time series $(y_{it})_t$ can be written as

$$\mu_{i,\tau} := \mathbb{E}[y_{i,t} y_{i,t+\tau}]^2 = \sigma_i^4 \gamma_i(\tau)^2, \quad i = 1, \ldots, K, \quad \tau \ge 0. \tag{4.2.6}$$

For increased generality, we will consider two different types of asymptotic regimes on $\mu_{i,\tau}$ as $T \to \infty$. In the first case we assume τ is a fixed integer for all T; in the second case we allow τ to vary with T and assume $\tau \to \infty$ as $T \to \infty$. In the case where τ is fixed, we will assume without any loss of generality that the sequence $(\mu_{i,\tau})_i$ is arranged in decreasing order. Furthermore, we assume that $\{\mu_{i,\tau}\}$ is well separated, i.e. there exists $\epsilon > 0$ such that $\mu_{i,\tau}/\mu_{i+1,\tau} > 1 + \epsilon$ for all i and T. This assumption is standard (see e.g. [52]) and ensures that the empirical eigenvalues are separated asymptotically.

In the case where τ is allowed to vary with T, it is too restrictive to assume that such an ordering on $\mu_{i,\tau}$ exists for all $\tau \ge 0$. For example, suppose that the first coordinate $(y_{1t})_t$ has a large variance σ_1^2 but a very rapidly decaying auto-covariance function $\gamma_1(\cdot)$, while $(y_{2t})_t$ has a smaller variance but a slow decaying auto-covariance function. Then we can easily have $\mu_{1,1} > \mu_{2,1}$ as well as $\mu_{1,\tau} < \mu_{2,\tau}$ for a larger τ so the assumption $\mu_{1,\tau} > \mu_{2,\tau}$ for all τ is unrealistic. Instead, we will assume that the sequence $(\mu_{i,\tau})_i$ is well separated only asymptotically, i.e. we assume there exists T_0 large enough and $\epsilon > 0$ such that

$$\mu_{i,\tau}/\mu_{i+1,\tau} > 1 + \epsilon, \quad \forall T > T_0, \quad i = 1, \ldots, K.$$

We will assume that each $\gamma_i(\tau)$ decays at the same speed asymptotically, i.e. $\gamma_i(\tau)/\gamma_j(\tau) < C_1$ for $i, j = 1, \ldots, K$ and some constant C_1. This implies that the $\mu_{i,\tau}$'s are of the same order as well and a comparison between them is more reasonable.

For clarity and the convenience of the reader we summarize our settings into the following sets of conditions which will be referred to in later parts of the paper.

Assumption 4.1. *a) $p, T \to \infty$ and $p/T \to c > 0$.*

b) $\sigma_i \to \infty$ and there exists $C > 0$ such that $\sigma_i/\sigma_j < C$ for all $i, j = 1, \ldots, K$.

c) $(z_{it})_{1 \le i \le K, 1-L \le t \le T+1}$ is independent, identically distributed with $\mathbb{E}[z_{it}] = 0$, $\mathbb{E}[z_{it}^2] = 1$ and uniformly bounded $4 + \epsilon$ moment for some $\epsilon > 0$.

d) $(\epsilon_{it})_{1 \le i \le p+K, 1 \le t \le T+1}$ is i.i.d. standard Gaussian.

e) $\sup_i \|\boldsymbol{\varphi}_i\|_{\ell_1} < \infty$.

We note that (a) and (b) of Assumption 4.1 capture our asymptotic regime where p diverges at the same rate as T and the strength of all factors diverge at comparable rates. Moment conditions such as (c) of Assumption 4.1 are standard in the literature; see for instance [18, 52, 110, 162, 163]. The normality assumption in (d) is solely for the purpose of reducing the model to a canonical form, as discussed in the previous section. Finally, condition (e) is very standard in the time series literature, see [43]; we list it here for ease of reference. For instance, condition (e) is satisfied by any causal auto-regressive moving average process written in the form (4.2.2).

The following two sets of assumptions encapsulate the two asymptotic schemes discussed above. Most of our main results hold under either set of assumptions.

Assumption 4.2. *a) τ is a fixed, non-negative integer*

b) $K = o(T^{1/16})$ and $K = o(\sigma_1^2)$ as $T \to \infty$.

c) the sequence $(\mu_{1,\tau}, \ldots, \mu_{K,\tau})$ is arranged in decreasing order and there exists $\epsilon > 0$ such that $\mu_{i,\tau}/\mu_{i+1,\tau} > 1 + \epsilon$ for all $i = 1, \ldots, K-1$.

Assumption 4.3. *a) $\tau \in \mathbb{N}$ and $\tau \to \infty$ as $T \to \infty$.*

b) $K = o(T^{1/16}\gamma_1(\tau)^{1/2})$ and $K = o(\sigma_1^2\gamma_1(\tau)^3)$ as $T \to \infty$.

c) there exists $C_1 > 0$ such that $\mu_{i,\tau}/\mu_{j,\tau} \le C_1$ for all $i, j = 1, \ldots, K$ and $\tau \ge 0$.

d) there exists T_0 large enough and some $\epsilon > 0$ such that $\mu_{i,\tau}/\mu_{i+1,\tau} > 1 + \epsilon$ for all $i = 1, \ldots, K-1$ and $T > T_0$.

Assumption 4.2 describes the asymptotic regime where τ is a fixed integer and Assumption 4.3 allows τ to diverge along with T. We note that condition (b) of both of the above set of assumptions is trivially satisfied when the number of factors K is assumed to be finite. Under (b) of Assumption 4.1, condition (c) of Assumption 4.3 ensures that the strengths of factors are comparable when $\tau \to \infty$. Finally, (c) of Assumption 4.2 and (d) of Assumption 4.3 are standard and ensure that the empirical eigenvalues are separated from each other asymptotically, see for instance see e.g. [52].

4.2.2 Notations and Preliminaries

In our exposition and proofs we will often encounter various resolvent matrices, which capture the spectral information of the random matrices we are studying. Since we are constantly dealing with many different matrices, assigning to each a different letter will easily exhaust the alphabet. Instead, we adopt some non-standard notations for matrices and sub-matrices. Write (a_{ij}) for a matrix where the (i,j)-th entry is equal to a_{ij}. For such a matrix (a_{ij}), we will write

$$\mathbf{a}_{[i:j],[k:l]} := \begin{pmatrix} a_{ik} & \dots & a_{il} \\ \vdots & \ddots & \vdots \\ a_{jk} & \dots & a_{jl} \end{pmatrix}$$

for a specified sub-matrix. Similarly we will write $\mathbf{a}_{i,[j:k]}$ and $\mathbf{a}_{[i:j],k}$ for the column vectors $(a_{ij}, \dots, a_{ik})^\top$ and $(a_{ik}, \dots, a_{jk})^\top$ respectively.

First we introduce notations for some of the more important random matrices in our study. We denote $x_{it} = \sigma_i f_{it} + \epsilon_{it}$, $i = 1, \dots, K$, $t = 1, \dots, T$ and write

$$\begin{aligned} X_0 &:= \frac{1}{\sqrt{T}} \mathbf{x}_{[1:K],[1:T-\tau]}, \quad X_\tau := \frac{1}{\sqrt{T}} \mathbf{x}_{[1:K],[\tau+1:T]}, \\ E_0 &:= \frac{1}{\sqrt{T}} \boldsymbol{\epsilon}_{[K+1:K+p],[1:T-\tau]}, \quad E_\tau := \frac{1}{\sqrt{T}} \boldsymbol{\epsilon}_{[K+1:K+p],[\tau+1:T]}, \end{aligned} \tag{4.2.7}$$

for matrices containing the factors and noises in our model. We will also write

$$Y_0 := \frac{1}{\sqrt{T}} \mathbf{y}_{[1:p+K],[1:T-\tau]}, \quad Y_\tau := \frac{1}{\sqrt{T}} \mathbf{y}_{[1:p+K],[\tau+1:T]}, \tag{4.2.8}$$

i.e. we have $Y_0 = (X_0^\top, E_0^\top)^\top$ and $Y_\tau = (X_\tau^\top, E_\tau^\top)^\top$. For an integer $\tau \geq 0$, the lag-τ sample auto-covariance matrix of $\mathbf{y}_t$ can then be written as

$$\widehat{\Sigma}_\tau := \frac{1}{T} \sum_{t=1}^{T-\tau} \mathbf{y}_{t+\tau} \mathbf{y}_t^\top = \begin{pmatrix} X_\tau \\ E_\tau \end{pmatrix} \begin{pmatrix} X_0 \\ E_0 \end{pmatrix}^\top = \begin{pmatrix} X_\tau X_0^\top & X_\tau E_0^\top \\ E_\tau X_0^\top & E_\tau E_0^\top \end{pmatrix}.$$

Next we introduce resolvent matrices which are central to the study of spectral properties of random matrices. Most of our results rely on certain bilinear forms formed using the resolvents. For $a \in \mathbb{R}$ outside of the spectrum of the matrix $E_\tau^\top E_\tau E_0^\top E_0$ write

$$R(a) := (I_{T-\tau} - a^{-1} E_\tau^\top E_\tau E_0^\top E_0)^{-1} = a(a - E_\tau^\top E_\tau E_0^\top E_0)^{-1} \tag{4.2.9}$$

for the (scaled) resolvent of $E_\tau^\top E_\tau E_0^\top E_0$ at a. The resolvent $R(a)$ satisfies

$$R(a) = I_{T-\tau} + a^{-1} R(a) E_\tau^\top E_\tau E_0^\top E_0, \tag{4.2.10}$$

which follows from rearranging $R(a)(I_{T-\tau} - a^{-1}E_\tau^\top E_\tau E^\top E) = I_{T-\tau}$. Using the identity

$$A(\lambda I - BA)^{-1} = (\lambda I - AB)^{-1}A \tag{4.2.11}$$

we may also obtain the following identities

$$R(a)E_\tau^\top E_\tau = E_\tau^\top E_\tau R(a)^\top, \quad E_0^\top E_0 R(a) = R(a)^\top E_0^\top E_0. \tag{4.2.12}$$

In our analysis we will constantly be dealing with certain quadratic forms involving matrices X_0, X_τ, E_0, E_τ and the resolvent $R(a)$. To simplify notations we will write

$$A(a) := \frac{1}{\sqrt{a}} X_0 R(a) X_\tau^\top, \quad B(a) := \frac{1}{a} X_\tau E_0^\top E_0 R(a) X_\tau^\top, \tag{4.2.13}$$

$$\overline{Q}(a) := I_K - a^{-1} X_\tau E_0^\top E_0 R(a) X_\tau^\top, \quad Q(a) := I_K - a^{-1} X_0 R(a) E_\tau^\top E_\tau X_0^\top. \tag{4.2.14}$$

For any a outside the spectrum of the matrix $X_0 R(a) E_\tau^\top E_\tau X_0^\top$, the matrix $Q(a)$ defined above is invertible and similar to (4.2.10), we have

$$Q(a)^{-1} = I_K + \frac{1}{a} Q(a)^{-1} X_0 R(a) E_\tau^\top E_\tau X_0^\top. \tag{4.2.15}$$

For two sequences of positive numbers (a_n) and (b_n), we write $a_n \lesssim b_n$ if there exists a constant $c > 0$ such that $a_n \le c b_n$. We write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold simultaneously. A sequence of events (F_n) is said to hold with high probability if there exist constants $c, C > 0$ such that $\mathbb{P}(F_n^c) \le C n^{-c}$. The operator and Hilbert-Schmidt norms of a matrix M are denoted by $\|M\|$ and $\|M\|_F$ respectively, and we write $\|(a_n)\|_{\ell^p}$ for the ℓ_p norm of a sequence (a_n). We will write $(\mathbf{e}_i)_{i=1}^n$ for the standard orthonormal basis of Euclidean space $\mathbb{R}^n$, often without specifying the dimension n.

We will use the usual o_p and O_p notations for convergence in probability and stochastic compactness. For $p \ge 1$, we will write o_{L^p} and O_{L^p} for convergence to zero and boundedness in L^p, i.e. for a sequence of random variables $(X_n)_n$ and real numbers (a_n), we write $X_n = O_{L^p}(a_n)$ if $\mathbb{E}|X_n/a_n|^p = O(1)$ and $X_n = o_{L^p}(a_n)$ if $\mathbb{E}|X_n/a_n|^p = o(1)$. For matrices (A_n) we will write $A_n = O_{p,\|\cdot\|}(a_n)$ if $\|A_n\| = O_p(a_n)$.

Throughout the paper we will make use of certain events of high probability. Define

$$\mathcal{B}_0 := \left\{ \|E_0^\top E_0\| + \|E_\tau^\top E_\tau\| \le 4\left(1 + \frac{p}{T}\right) \right\},$$
$$\mathcal{B}_1 := \left\{ \|X_0^\top X_0\| + \|X_\tau^\top X_\tau\| \le 2 \sum_{i=1}^K \sigma_i^2 \right\} \tag{4.2.16}$$

and $\mathcal{B}_2 := \mathcal{B}_1 \cap \mathcal{B}_2$. We first state a preliminary result showing that these events happen with high probability as $T \to \infty$. The proof will be given in Section 4.5.

Lemma 4.1. *Under Assumption 4.1 and either Assumption 4.2 or 4.3, we have* *a)* $\mathcal{B}_0$ *holds with probability* $\mathbb{P}(\mathcal{B}_0) = 1 - o(T^{-l})$ *for any* $l \in \mathbb{N}$ *as* $T \to \infty$*.*

b) For $k = 1, 2$*,* $\mathcal{B}_k$ *holds with probability* $\mathbb{P}(\mathcal{B}_k) = 1 - O(KT^{-1})$ *as* $T \to \infty$*.*

As an immediate consequence of this lemma and (b) of Assumption 4.1, we have

$$\|E_0^\top E_0\| + \|E_\tau^\top E_\tau\| = O_p(1), \quad \|X_0^\top X_0\| + \|X_\tau^\top X_\tau\| = O_p(K\sigma_1^2). \tag{4.2.17}$$

Furthermore, we observe that under the event $\mathcal{B}_0$, for any sequence $(a_T)_T$ such that $a_T \to \infty$, the matrix $I_{T-\tau} - a_T^{-1} E_\tau^\top E_\tau E_0^\top E_0$ is eventually invertible. Moreover, we note that under $\mathcal{B}_0$ we have $\|a_T^{-1} E_\tau^\top E_\tau E_0^\top E_0 1_{\mathcal{B}_0}\| \leq 4a_T^{-1}(1 + p/T) = O(a_T^{-1})$, which is a deterministic upper-bound. By the reverse triangle inequality we immediately have $\|I_{T-\tau} - a_T^{-1} E_\tau^\top E_\tau E_0^\top E_0 1_{\mathcal{B}_0}\| \geq 1 - O(a_T^{-1})$ and therefore

$$\|R(a_T) 1_{\mathcal{B}_0}\| = 1 + o(1), \quad \|R(a_T)\| = 1 + o_p(1). \tag{4.2.18}$$

Similarly, under the event $\mathcal{B}_2$ the matrix $Q(a_T)$ is eventually invertible as $a_T \to \infty$ and

$$\|Q(a_T)^{-1} 1_{\mathcal{B}_2}\| = 1 + o(1), \quad \|Q(a_T)^{-1}\| = 1 + o_p(1). \tag{4.2.19}$$

Finally, let $\mathcal{F}_p$ be the σ-algebra generated by the noise time series $(\boldsymbol{\epsilon}_t)$, i.e.

$$\mathcal{F}_p := \sigma\Big(\{\epsilon_{it}, i = K+1, \ldots, K+p, t = 1, \ldots, T\}\Big). \tag{4.2.20}$$

We will often take expectations conditional on the noise series, in which case we shall write

$$\underline{\mathbb{E}}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|\mathcal{F}_p]. \tag{4.2.21}$$

4.3 Main results

Write $\lambda_{n,\tau}$ for the n-th largest eigenvalue of the symmetrized lag-τ sample auto-covariance matrix $\widehat{\Sigma}_\tau \widehat{\Sigma}_\tau^\top$. The main goal of our work is to establish the asymptotic normality of $\lambda_{n,\tau}$ for $n \leq K$ after appropriate centering and scaling. We will first in Section 4.3.1 establish the asymptotic location of the eigenvalue $\lambda_{n,\tau}$ as well as identify the correct centering for $\lambda_{n,\tau}$ in order to obtain a central limit theorem. The central limit theorem itself, which is the main result of our work, is stated in Theorem 4.5 of Section 4.3.2.

Due to its length, the proof of Theorem 4.5 will be divided into a series of propositions and technical lemmas, which are collected in Section 4.4 and Section 4.5. For the convenience of the reader, we will summarize the strategy of the proof of Theorem 4.5 and explain how the intermediate results are used at the end of Section 4.3.2.

4.3.1 Location of Spiked Eigenvalues

We first show that the spiked eigenvalue $\lambda_{n,\tau}$ is close to its population counterpart $\mu_{n,\tau}$ asymptotically. This will in particular give the asymptotic order of $\lambda_{n,\tau}$ as $T \to \infty$.

Theorem 4.2. *Under Assumption 4.1 and either Assumption 4.2 or 4.3, we have*

$$\frac{\lambda_{n,\tau}}{\mu_{n,\tau}} - 1 = O_p\bigg(\frac{1}{\gamma_n(\tau)\sqrt{T}}\bigg) + O_p\bigg(\frac{K}{\sigma_n^2\gamma_n(\tau)^2}\bigg), \quad n = 1, \ldots, K, \tag{4.3.1}$$

where $\mu_{n,\tau}$ and $\gamma_n(\tau)$ are defined in (4.2.6) and (4.2.3) respectively.

Proof. We shall write $\Lambda_n(A)$ for the n-th largest eigenvalue of a matrix A. Note that the non-zero eigenvalues of $\hat\Sigma_\tau\hat\Sigma_\tau^\top = Y_\tau Y_0^\top Y_0 Y_\tau^\top$ coincide with those of the matrix $Y_0^\top Y_0 Y_\tau^\top Y_\tau = (X_0^\top X_0 + E_0^\top E_0)(X_\tau^\top X_\tau + E_\tau^\top E_\tau)$. We first show that the eigenvalue $\Lambda_n(\hat\Sigma_\tau\hat\Sigma_\tau^\top)$ is close to $\Lambda_n(X_0^\top X_0 X_\tau^\top X_\tau)$. By Weyl's inequality (Lemma B.1 of [71]) we have

$$\begin{aligned}\Big|\Lambda_n(\hat\Sigma_\tau\hat\Sigma_\tau^\top) - \Lambda_n(X_0^\top X_0 X_\tau^\top X_\tau)\Big| &= \Big|\Lambda_n(Y_0^\top Y_0 Y_\tau^\top Y_\tau) - \Lambda_n(X_0^\top X_0 X_\tau^\top X_\tau)\Big| \\ &\le \|X_0^\top X_0 E_\tau^\top E_\tau + E_0^\top E_0 X_\tau^\top X_\tau + E_0^\top E_0 E_\tau^\top E_\tau\| = O_p(K\sigma_1^2),\end{aligned}$$

where the last equality follows from (4.2.17). Dividing by $\mu_{n,\tau} = \sigma_n^4\gamma_n(\tau)^2$ we have

$$\frac{\Lambda_n(\hat\Sigma_\tau\hat\Sigma_\tau^\top) - \Lambda_n(X_\tau X_0^\top X_0 X_\tau^\top)}{\sigma_n^4\gamma_n(\tau)^2} = O_p\bigg(\frac{K\sigma_1^2}{\sigma_n^4\gamma_n(\tau)^2}\bigg). \tag{4.3.2}$$

Next we compute $\Lambda_n(X_\tau X_0^\top X_0 X_\tau^\top)$ in more detail. It is shown in Lemma 4.14 that

$$(X_0X_\tau^\top)_{ij} = \mathbb{E}[(X_0X_\tau^\top)_{ij}] + O_{L^2}(\sigma_i\sigma_j T^{-1/2}), \tag{4.3.3}$$

where from equation (4.5.6) we know $\mathbb{E}[(X_0X_\tau^\top)_{ij}] = 1_{i=j}\sigma_i^2\gamma_i(\tau)$. Therefore for any $i \neq j$, the off-diagonal elements of $X_\tau X_0^\top X_0 X_\tau^\top$ can be written as

$$\begin{aligned}(X_\tau X_0^\top X_0 X_\tau^\top)_{ij} &= \sum_{k=1}^K (X_0X_\tau^\top)_{ki}(X_0X_\tau^\top)_{kj} \\ &= (X_0X_\tau^\top)_{ii}(X_0X_\tau^\top)_{ij} + (X_0X_\tau^\top)_{ji}(X_0X_\tau^\top)_{jj} + \sum_{k\neq i,j}(X_0X_\tau^\top)_{ki}(X_0X_\tau^\top)_{kj} \\ &= \Big(\sigma_i^2\gamma_i(\tau) + \sigma_j^2\gamma_j(\tau)\Big)O_{L^2}(\sigma_i\sigma_j T^{-1/2}) + O_{L^1}(\sigma_1^4 K T^{-1}) \\ &= O_{L^2}\Big(\sigma_1^4\gamma_1(\tau)T^{-1/2}\Big) + O_{L^1}\bigg(\frac{\sigma_1^4\gamma_1(\tau)}{\sqrt{T}}\frac{K}{\gamma_1(\tau)\sqrt{T}}\bigg) = O_{L_1}\bigg(\frac{\sigma_1^4\gamma_1(\tau)}{\sqrt{T}}\bigg),\end{aligned} \tag{4.3.4}$$

where the equality in the second last line follows from (4.3.3) and the last line follows from Assumption 4.2 and 4.3. Similarly, the diagonal elements of $X_\tau X_0^\top X_0 X_\tau^\top$ satisfy

$$
\begin{aligned}
(X_\tau X_0^\top X_0 X_\tau^\top)_{ii} &= (X_0 X_\tau^\top)_{ii}^2 + \sum_{k\neq i}(X_0X_\tau^\top)_{ki}^2 \\
&= \Big(\sigma_i^2\gamma_i(\tau) + O_{L^2}(\sigma_i^2 T^{-1/2})\Big)^2 + O_{L^1}(KT^{-1}) = \mu_{i,\tau} + O_{L_1}\left(\frac{\sigma_1^4\gamma_1(\tau)}{\sqrt{T}}\right). \qquad (4.3.5)
\end{aligned}
$$

Using (4.3.4), (4.3.5) and taking a union bound over i, j we finally obtain

$$
\|X_\tau X_0^\top X_0 X_\tau^\top - \operatorname{diag}(\mu_{i,\tau})\|_\infty = O_p\left(\frac{K^2\sigma_1^4\gamma_1(\tau)}{\sqrt{T}}\right) \qquad (4.3.6)
$$

or equivalently we may write

$$
X_\tau X_0^\top X_0 X_\tau^\top \operatorname{diag}(\mu_{i,\tau}^{-1}) = I_K + O_{p,\|\cdot\|_\infty}\left(\frac{K^2}{\sqrt{T}\gamma_1(\tau)}\right). \qquad (4.3.7)
$$

Let $\omega_1, \dots, \omega_K$ be the eigenvalues of $X_\tau X_0^\top X_0 X_\tau^\top$ arranged in decreasing order. Let ω be one of these eigenvalues. Define the function

$$
G(\omega) := \Big(X_\tau X_0^\top X_0 X_\tau^\top - \omega I_K\Big)\operatorname{diag}(\mu_{i,\tau}^{-1}),
$$

then clearly we have $0 = |X_\tau X_0^\top X_0 X_\tau^\top - \omega I_K| = |G(\omega)|$. From (4.3.7) we get

$$
\begin{aligned}
0 = |G(\omega)| &= \left|I_K + O_{p,\|\cdot\|_\infty}\left(\frac{K^2}{\sqrt{T}\gamma_1(\tau)}\right) - \omega\operatorname{diag}\Big(\mu_{i,\tau}^{-1}\Big)\right| \\
&= \left|I_K - \operatorname{diag}(\omega\mu_i^{-1}) + O_{p,\|\cdot\|_\infty}\left(\frac{K^2}{\sqrt{T}\gamma_1(\tau)}\right)\right|,
\end{aligned}
$$

and using Leibniz's formula analogous to the derivation of (4.4.42) we obtain

$$
0 = |G(\omega)| = \prod_{i=1}^K G(\omega)_{ii} + O_p\left(\frac{K^6}{\gamma_1(\tau)^2 T}\right). \qquad (4.3.8)
$$

Since $\prod_i G(\omega)_{ii} = o_p(1)$, there is at least one $i \in \{1, \dots, K\}$ such that $G(\omega)_{ii} = o_p(1)$. Now we show that in fact there can be only one such i. For any $i \neq j$, we have

$$
G(\omega)_{ii} - G(\omega)_{jj} = \omega(\mu_i^{-1} - \mu_j^{-1}) \geq \omega\mu_i^{-1}(1+\epsilon) \qquad (4.3.9)
$$

for some $\epsilon > 0$, where the last inequality follows from either Assumption 4.2 or 4.3. We first observe that $\omega\mu_i^{-1} \neq o_p(1)$ for any i as $T \to \infty$. Indeed, suppose for a contradiction that $\omega\mu_i^{-1} = o_p(1)$, since $\mu_i \asymp \mu_j$ for any $i = j$, we easily see that $G(\omega)_{ii} = 1 + o_p(1)$ for every i, which makes (4.3.8) impossible. Substituting back into (4.3.9) we see that $G(\omega)_{ii} - G(\omega)_{jj} \gtrsim 1+\epsilon$ for any $i \neq j$. Clearly, if $G(\omega)_{ii} = o_p(1)$, then $G(\omega)_{jj} \gtrsim 1+\epsilon+o_p(1)$ for any $j \neq i$, i.e. there can be only one i such that $G(\omega)_{ii} = o_p(1)$.

Therefore, for (4.3.8) to hold, there must exist some $i \in \{1, \dots, K\}$ such that

$$0 = G(\omega)_{ii} + O_p\left(\frac{K^6}{\gamma_1(\tau)^2 T}\right).$$

It then follows that the K solutions to $|G(\omega)| = 0$ satisfy the system of equations

$$0 = G(\omega)_{ii} + O_p\left(\frac{K^6}{\gamma_1(\tau)^2 T}\right), \quad i = 1, \dots, K.$$

Using (4.3.5), we see that each $G(\omega)_{ii}$ satisfies

$$G(\omega)_{ii} = \frac{(X_\tau X_0^\top X_0 X_\tau^\top)_{ii} - \omega}{\mu_{i,\tau}} = 1 - \frac{\omega}{\mu_{i,\tau}} + O_p\left(\frac{1}{\sqrt{T}\gamma_1(\tau)}\right),$$

which implies that the K solutions to $|G(\omega)| = 0$ satisfy the system of equations

$$\frac{\omega}{\mu_{i,\tau}} - 1 = O_p\left(\frac{1}{\gamma_1(\tau)\sqrt{T}}\right), \quad i = 1, \dots, K. \tag{4.3.10}$$

Note that by definition there are K possible choices of ω, which are the ordered eigenvalues of $X_\tau X_0^\top X_0 X_\tau^\top$. Since $\{\mu_{i,\tau}\}$ are ordered under Assumption 4.2 or asymptotically ordered under Assumption 4.3, we can easily conclude that

$$\frac{\Lambda_i(X_\tau X_0^\top X_0 X_\tau^\top)}{\mu_{i,\tau}} - 1 = O_p\left(\frac{1}{\gamma_1(\tau)\sqrt{T}}\right).$$

Combining this result with (4.3.2) we get

$$\frac{\Lambda_i(\hat{\Sigma}_\tau \hat{\Sigma}_\tau^\top)}{\mu_{i,\tau}} - 1 = O_p\left(\frac{1}{\sqrt{T}\gamma_1(\tau)} + \frac{K}{\sigma_i^2 \gamma_i(\tau)^2}\right)$$

which completes the proof. □

Remark 4.3. A closer inspection of the convergence rate in Theorem 4.2 shows that $\mu_{n,\tau}$ is not the appropriate centering constant for $\lambda_{n,\tau}$ for the purpose of obtaining a CLT. The first term on the right hand side of (4.3.1) can indeed be shown to be asymptotically normal at a scaling of $\gamma_n(\tau)\sqrt{T}$, which is the same scaling as in our main result Theorem 4.5. However, the second term in (4.3.1) is in general not negligible after scaling by $\gamma_n(\tau)\sqrt{T}$ since we do not impose assumptions on the rate of divergence of $\mu_{n,\tau}$.

If we were to impose stronger assumptions on the rate of $\mu_{n,\tau}$, for example assuming the rate $\mu_{n,\tau} \asymp p^{1-\delta}$ required in [104], then the second term in (4.3.1) indeed becomes negligible. Under such assumptions the K leading eigenvalues of the $(p+K) \times (p+K)$ dimensional matrix $\widehat{\Sigma}_\tau \widehat{\Sigma}_\tau^\top$ are extremely close to the eigenvalues of the $K \times K$ matrix $X_\tau X_0^\top X_0 X_\tau^\top$, as can be deduced from the proof of Theorem 4.2. The analysis of the matrix $\widehat{\Sigma}_\tau \widehat{\Sigma}_\tau^\top$ reduces to the analysis of the much simpler matrix $X_\tau X_0^\top X_0 X_\tau^\top$, which is essentially a low dimensional problem. In this case, the derivation of a CLT is much easier and does not require any techniques and results from random matrix theory.

As can be seen from the proof of Theorem 4.2, the second term in (4.3.1) represents the bias incurred when estimating $\mu_{n,\tau}$ using $\lambda_{n,\tau}$. In order to obtain a CLT, we need a more accurate centering term for $\lambda_{n,\tau}$ to remove or reduce this bias. This centering term, which we write as $\theta_{n,\tau}$, will be defined implicitly as the unique solution to the equation

$$1 = \mathbb{E}[B(\theta_{n,\tau})_{nn}1_{\mathcal{B}_0}] - \mathbb{E}[A(\theta_{n,\tau})_{nn}1_{\mathcal{B}_0}]^2\mathbb{E}[Q(\theta_{n,\tau})_{nn}^{-1}1_{\mathcal{B}_2}], \tag{4.3.11}$$

where the matrices $B(a)$, $A(a)$ and $Q(a)$ are defined in (4.2.13) and (4.2.14) for $a \in \mathbb{R}$.

To make this definition rigorous, we start with Proposition 4.4 which shows that (4.3.11) indeed has a unique solution for T large enough. Furthermore, this solution is shown to exist in some small interval containing $\mu_{n,\tau} = \sigma_n^4\gamma_n(\tau)^2$. This in particular establishes the asymptotic order of $\theta_{n,\tau}$.

Proposition 4.4. *Suppose Assumption 4.1 and either Assumption 4.2 or Assumption 4.3 hold. Fix $n \in \{1, \ldots, K\}$ and let $\epsilon \in (0, 1)$ be an arbitrary constant not related to p, T. Then there exists T_0 large enough such that for $T > T_0$, the function*

$$a \mapsto g(a) = 1 - \mathbb{E}[B(a)_{nn}1_{\mathcal{B}_0}] - \mathbb{E}[A(a)_{nn}1_{\mathcal{B}_0}]^2\mathbb{E}[Q(a)_{nn}^{-1}1_{\mathcal{B}_2}]$$

has a unique root in the interval

$$\sigma_n^4\gamma_n(\tau)^2[1-\epsilon, 1+\epsilon]. \tag{4.3.12}$$

Proof. We first consider the invertibility of the matrix $Q(a)$ defined in (4.2.14). Recall the matrix $R(a)$ from (4.2.9) and the event $\mathcal{B}_2$ from (4.2.16). Since $a \asymp \sigma_n^4\gamma_n(\tau)^2 \to \infty$ and $\|E_\tau^\top E_\tau E_0^\top E_0 1_{\mathcal{B}_2}\|$ is bounded by definition of $\mathcal{B}_2$, the matrix $I - a^{-1}E_\tau^\top E_\tau E_0^\top E_0$ is invertible under $\mathcal{B}_2$ and we have $\|R(a)\|1_{\mathcal{B}_2} = O(1)$. Therefore $\|a^{-1}X_0R_aE_\tau^\top E_\tau X_0^\top\|1_{\mathcal{B}_2} = O(\sigma_n^{-2}\gamma_n(\tau)^{-2}) = o(1)$ and thus $Q(a)$ is invertible under $\mathcal{B}_2$ for T large enough.

It will be shown in Lemma 4.15 that

$$\mathbb{E}[A(a)_{nn}1_{\mathcal{B}_0}] = \frac{\sigma_n^2\gamma_n(\tau)}{\sqrt{a}} + o(1), \quad \mathbb{E}[B(a)_{nn}1_{\mathcal{B}_0}] = o(1), \quad \mathbb{E}[Q(a)_{nn}^{-1}1_{\mathcal{B}_2}] = 1 + o(1).$$

From (4.3.12) we have $a^{-1/2}\sigma_n^2\gamma_n(\tau) \asymp O(1)$, using which we can obtain

$$g(a) = 1 - \mathbb{E}[A(a)1_{\mathcal{B}_0}]_{nn}^2\mathbb{E}[Q(a)_{nn}^{-1}1_{\mathcal{B}_2}] + o(1) = 1 - \frac{\sigma_n^4\gamma_n(\tau)^2}{a} + o(1).$$

Substituting the endpoints of the interval (4.3.12) into the function g, we have

$$g((1 \pm \epsilon)\sigma_n^4\gamma_n(\tau)^2) = 1 - \frac{1}{1 \pm \epsilon} + o(1) = \frac{\mp\epsilon}{1 \pm \epsilon} + o(1).$$

For T large enough, the signs of g differ at the two endpoints of the interval (4.3.12) and therefore g has a root inside the interval. It is not difficult to observe that g is a monotone function in a for T large enough which implies the root is unique. □

4.3.2 Central Limit Theorem for Spiked Eigenvalues

The constant $\theta_{n,\tau}$ defined via equation (4.3.11) turns out to be the appropriate centering constant for $\lambda_{n,\tau}$, in the sense that the second term in (4.3.1) becomes negligible after centering by $\theta_{n,\tau}$. We are ready to state the main result of our work. Define

$$\delta_{n,\tau} := \frac{\lambda_{n,\tau} - \theta_{n,\tau}}{\theta_{n,\tau}} = \frac{\lambda_{n,\tau}}{\theta_{n,\tau}} - 1. \tag{4.3.13}$$

Theorem 4.5. *Under Assumption 4.1 and either Assumption 4.2 or 4.3, we have*

$$\sqrt{T}\frac{\gamma_n(\tau)}{2v_{n,\tau}}\delta_{n,\tau} \Rightarrow N(0,1),$$

where $v_{n,\tau}$ is defined by

$$v_{i,\tau}^2 := \frac{1}{T}\mathrm{Var}(\mathbf{f}_{i0}^\top \mathbf{f}_{i\tau}) = \sum_{|k|<T-\tau}\left(1 - \frac{|k|}{T-\tau}\right)u_k, \tag{4.3.14}$$

and $(u_k)_{|k|<T-\tau}$ is a sequence of constants given by

$$u_k := \gamma_i(k)^2 + \gamma_i(k+\tau)\gamma_i(k-\tau) + (\mathbb{E}[z_{11}^4] - 3)\sum_{l=0}^{\infty}\varphi_{i,l}\varphi_{i,l+\tau}\varphi_{i,l+k}\varphi_{i,l+k+\tau}.$$

Remark 4.6. We remark that for generality as well as tidiness of presentation we choose to formulate Theorem 4.5 in a form that holds under either one of Assumption 4.2 and 4.3. A closer inspection shows that the two cases are quite different. In the case where $\tau \to \infty$, we observe that $\gamma_n(\tau) \to 0$ while the term $v_{i,\tau}^2$ defined by (4.3.14) can easily be shown to be bounded away from zero. This implies that the scaling of CLT in the two cases are not of the same order. In the case where τ is fixed, the variance $4v_{n,\tau}\gamma_n(\tau)^{-2}$ of $\sqrt{T}\delta_{n,\tau}$ is bounded both from above and away from zero from below. On the other hand, when $\tau \to \infty$, the variance of $\sqrt{T}\delta_{n,\tau}$ tends to infinity at a speed of $\gamma_n(\tau)^{-1}$.

This result might seem surprising since $\delta_{n,\tau} = (\lambda_{n,\tau} - \theta_{n,\tau})/\theta_{n,\tau}$ is already normalized in an obvious way so one might expect $\delta_{n,\tau}$ to be of order $T^{-1/2}$. One might be tempted to draw the conclusion that $\lambda_{n,\tau}$ is less accurate an estimator of $\theta_{n,\tau}$ for larger values of τ, since the variance of $\delta_{n,\tau}$ increases with τ. However, the opposite is true here. Since $\theta_{n,\tau} \asymp \sigma_n^4\gamma_n(\tau)^2$ by Proposition 4.4, this implies that in fact $\lambda_{n,\tau} - \theta_{n,\tau} = O_p(\sigma_n^4\gamma_n(\tau)T^{-1/2})$, which is faster than the rate $\lambda_{n,\tau} - \delta_{n,\tau} = O_p(\sigma_n^4 T^{-1/2})$ obtained in the case where τ is fixed. In practical situations where we deal with the auto-covariance matrix with a larger

τ, the CLT under Assumption 4.3 provides a much more accurate convergence speed and asymptotic variance than using fixed τ results.

To establish our main result Theorem 4.5, we need a collection of preliminary results in Propositions 4.7-4.10 and technical Lemmas 4.12-4.18. Due to the length of the proofs, we provide a summary of the key ideas involved. Before we proceed, we make some simplifications to commonly used notations.

Throughout the rest of the work, we will without loss of generality deal with the n-th largest eigenvalue $\lambda_{n,\tau}$ of $\widehat{\Sigma}_\tau \widehat{\Sigma}_\tau^\top$ for a chosen $n \in \{1, \ldots, K\}$. To avoid using too many layers of subscripts, we will routinely suppress the subscripts n, τ and write $\lambda := \lambda_{n,\tau}$. Recall the scaled resolvent $R(a)$ of the matrix $E_\tau^\top E_\tau E_0^\top E_0$ evaluated at a, as defined in (4.2.9). When we evaluate $R(\cdot)$ at the values $\lambda_{n,\tau}$ and $\theta_{n,\tau}$, we will simply write $R_\lambda := R_{\lambda_{n,\tau}}$ and $R := R_{\theta_{n,\tau}}$. Similarly, we will write $Q_\lambda, \overline{Q}_\lambda, Q$ and $\overline{Q}$ for $Q_{\lambda_{n,\tau}}, \overline{Q}_{\lambda_{n,\tau}}, Q_{\theta_{n,\tau}}$ and $\overline{Q}_{\theta_{n,\tau}}$ respectively. Using these notations we define the matrix $M := (M_{ij})_{ij}$ by

$$M := I_K - \frac{1}{\theta} X_\tau E_0^\top E_0 R X_\tau^\top - \frac{1}{\theta} X_\tau R^\top X_0^\top Q^{-1} X_0 R X_\tau^\top, \tag{4.3.15}$$

which will turn out to be the central object of our study.

The initial step in our analysis is to derive an expression for the eigenvalue λ and the quantity δ defined in (4.3.13). This is necessary because the eigenvalue λ of the matrix $\widehat{\Sigma}_\tau \widehat{\Sigma}_\tau^\top$ in general depends on its entries in complicated and non-linear ways. We take an approach commonly seen in the random matrix literature (e.g. [15, 52, 111]) and express δ as the solution to an equation involving the determinant of certain random matrices. This is established in Proposition 4.7 in which we express δ as the solution to the equation

$$\det\left(M + \frac{\delta}{\theta} X_\tau X_0^\top X_0 X_\tau^\top + \delta o_{p,\|\cdot\|}(1) \right) = 0. \tag{4.3.16}$$

The main idea is then to apply Leibniz's formula to compute this determinant. Doing so will express δ as a polynomial function of the entries of the matrices M and $\theta^{-1} X_\tau X_0^\top X_0 X_\tau^\top$ plus many higher order terms. After controlling the terms in this polynomial, it can be shown that the asymptotic normality of the ratio $\delta_{n,\tau}$ directly originates from the asymptotic normality of the corresponding entry M_{nn}. Specifically, as shown in the proof of Theorem 4.5, to establish the CLT, it suffices to show

$$\sqrt{T} \frac{M_{nn}}{2\gamma_n(\tau) v_{n,\tau}} \Rightarrow N(0,1), \quad M_{ii} = 1 + o_p(1), \quad \forall i \neq n,$$

establish a uniform bound of sufficient sharpness on the off-diagonals of M, and identify the limits in probability of the entries of $\theta^{-1} X_\tau X_0^\top X_0 X_\tau^\top$. From this, it is clear that M and hence the resolvents R and Q^{-1} appearing in the definition of M are the central objects of our analysis. To deal with the expression (4.3.15), we first construct an approximation to

M in Proposition 4.8 that preserves its asymptotic distribution.

We now discuss why this approximation is constructed in a seemingly unusual way. Observe that since θ diverges, the resolvents R and Q^{-1} are very close to identity matrices for large T, a fact used many times in our proofs. However one cannot simply replace them by identities to simplify (4.3.15). Indeed, it is easy to show that $R - I_{T-\tau} = O_{p,\|\cdot\|}(\theta^{-1})$, which converges to zero but not fast enough for obtaining a CLT after scaling by $\sqrt{T}$. This is because we allow θ to diverge at any rate and not as a specified function of T.

It can be shown however that this approximation error of order θ^{-1} appears only in the mean of the asymptotic distribution, see for instance (4.4.33). That is, we can use identity matrices to approximate R and Q^{-1} in (4.8) as long as we include an appropriate centering step to adjust the expectation of M before scaling by $\sqrt{T}$. This centering step for the matrix M results in (4.3.11) from which the centering $\theta_{n,\tau}$ for $\lambda_{n,\tau}$ is defined. Recall from our discussion following the statement of Theorem 4.2 that θ is a more accurate centering term for λ than μ is. From the discussion above, it can be seen that this is essentially due to the fact that R and Q^{-1} are not close enough (in spectral norm) to identity matrices to allow for the scaling of $\sqrt{T}$.

Instead of identity matrices, we therefore use more accurate approximations to R and Q^{-1}, which in our case turn out to be their expectations under certain events of high probability. To show that these expectations are close enough to the resolvents themselves, we establish the concentration of R around its expectation in Lemma 4.16, the concentration of Q^{-1} around a certain conditional expectation in Lemma 4.17, and estimates on the differences between conditional and unconditional expectations in 4.18.

Using these tools, we can show in Proposition 4.8 that after centering by a certain conditional expectation (which is later replaced by an unconditional one using Lemma 4.18), the asymptotic distribution of M can be obtained from the asymptotic distribution of the bilinear form $X_0 R X_\tau^\top$, up to adjustments in the expectations.

It therefore remains to establish the asymptotics of $X_0 R X_\tau^\top$. Using tools developed in Lemma 4.12-4.15, we study the bilinear form $X_0 R X_\tau^\top$ and establish its concentration around a certain conditional expectation. Using these results we show in the proof of Proposition 4.10 that the asymptotic normality for $X_0 R X_\tau^\top$ follows from the asymptotic normality of the much simpler auto-covariance matrix $X_0 X_\tau^\top$, again up to adjustments in the expectations. The CLT for this matrix $X_0 X_\tau^\top$ is established in Proposition 4.9. Finally Proposition 4.10 gives the CLT for diagonals of the matrix M and required estimates for the off-diagonals. The proof of Theorem 4.5 is then assembled from the above pieces.

To recapitulate, the quantity of interest δ is first shown to satisfy equation 4.7. Through a series of approximations we establish the asymptotic normality of the diagonals of the matrix M appearing in 4.7. The off-diagonals of M are bounded in probability and we establish the limit in probability of the matrix $X_\tau X_0^\top X_0 X_\tau^\top$ appearing in 4.7. Leibniz's formula is then applied to compute the determinant in 4.7, and our main result 4.5 follows.

4.4 Proofs

This section contains the statements and proofs of the four propositions described in Section 4.3. The proof of our main result Theorem 4.5 is given at the end of this section.

We first give an expression for $\delta := \delta_{n,\tau}$. Recall the matrix M from (4.3.15)

$$M := I_K - \frac{1}{\theta} X_\tau E_0^\top E_0 R X_\tau^\top - \frac{1}{\theta} X_\tau R^\top X_0^\top Q^{-1} X_0 R X_\tau^\top .$$

Proposition 4.7. *Suppose Assumption 4.1 and either Assumption 4.2 or 4.3 hold. Then the ratio δ is the solution to the following equation*

$$\det\left(M + \frac{\delta}{\theta} X_\tau X_0^\top X_0 X_\tau^\top + \delta o_{p,\|\cdot\|}(1) \right) = 0. \tag{4.4.1}$$

Proof. Suppose λ is an eigenvalue of $\widehat{\Sigma}_\tau \widehat{\Sigma}_\tau^\top$, then $\sqrt{\lambda}$ is a singular value of the matrix $\widehat{\Sigma}_\tau$, or equivalently an eigenvalue of the $(2p+2K) \times (2p+2K)$ matrix

$$\begin{pmatrix} 0 & \widehat{\Sigma}_\tau \\ \widehat{\Sigma}_\tau^\top & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & X_\tau X_0^\top & X_\tau E_0^\top \\ 0 & 0 & E_\tau X_0^\top & E_\tau E_0^\top \\ X_0 X_\tau^\top & X_0 E_\tau^\top & 0 & 0 \\ E_0 X_\tau^\top & E_0 E_\tau^\top & 0 & 0 \end{pmatrix}.$$

By definition the eigenvalue λ satisfies

$$0 = \left| \begin{pmatrix} \sqrt{\lambda} I_K & \mathbf{0} \\ \mathbf{0} & \sqrt{\lambda} I_K^\top \end{pmatrix} - \begin{pmatrix} 0 & \widehat{\Sigma}_\tau \\ \widehat{\Sigma}_\tau^\top & 0 \end{pmatrix} \right| = \left| \begin{pmatrix} \sqrt{\lambda} I_K & 0 & -X_\tau X_0^\top & -X_\tau E_0^\top \\ 0 & \sqrt{\lambda} I_p & -E_\tau X_0^\top & -E_\tau E_0^\top \\ -X_0 X_\tau^\top & -X_0 E_\tau^\top & \sqrt{\lambda} I_K & 0 \\ -E_0 X_\tau^\top & -E_0 E_\tau^\top & 0 & \sqrt{\lambda} I_p \end{pmatrix} \right|,$$

which, after interchanging the columns and rows, becomes

$$0 = \left| \begin{pmatrix} \sqrt{\lambda} I_K & -X_\tau X_0^\top & 0 & -X_\tau E_0^\top \\ -X_0 X_\tau^\top & \sqrt{\lambda} I_K & -X_0 E_\tau^\top & 0 \\ 0 & -E_\tau X_0^\top & \sqrt{\lambda} I_p & -E_\tau E_0^\top \\ -E_0 X_\tau^\top & 0 & -E_0 E_\tau^\top & \sqrt{\lambda} I_p \end{pmatrix} \right|. \tag{4.4.2}$$

From Theorem 4.2 we know that $\lambda \to \infty$ as $T \to \infty$. From Lemma 4.1 we recall that the spectral norm of $E_\tau E_0^\top$ is bounded with probability tending to 1 as $T \to \infty$. Therefore the bottom right sub-matrix $\begin{pmatrix} \sqrt{\lambda} I_p & -E_\tau E_0^\top \\ -E_0 E_\tau^\top & \sqrt{\lambda} I_p \end{pmatrix}$ is invertible with probability tending to 1. Using the matrix identity

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & (D - CA^{-1}B)^{-1}) \end{pmatrix}$$

we can compute the inverse of the submatrix $\left(\begin{smallmatrix} \sqrt{\lambda} I_p & -E_\tau E_0^\top \\ -E_0 E_\tau^\top & \sqrt{\lambda} I_p \end{smallmatrix}\right)$ and get

$$\begin{aligned}
&\begin{pmatrix} \sqrt{\lambda} I_p & -E_\tau E_0^\top \\ -E E_\tau^\top & \sqrt{\lambda} I_p \end{pmatrix}^{-1} \\
&\quad = \begin{pmatrix} \left(\sqrt{\lambda} I_p - \frac{1}{\sqrt{\lambda}} E_\tau E_0^\top E_0 E_\tau^\top\right)^{-1} & \frac{1}{\sqrt{\lambda}} E_\tau E_0^\top \left(\sqrt{\lambda} I_p - \frac{1}{\sqrt{\lambda}} E_0 E_\tau^\top E_\tau E_0^\top\right)^{-1} \\ \frac{1}{\sqrt{\lambda}} E_0 E_\tau^\top \left(\sqrt{\lambda} I_p - \frac{1}{\sqrt{\lambda}} E_\tau E_0^\top E_0 E_\tau^\top\right)^{-1} & \left(\sqrt{\lambda} I_p - \frac{1}{\sqrt{\lambda}} E_0 E_\tau^\top E_\tau E_0^\top\right)^{-1} \end{pmatrix} \\
&\quad = \begin{pmatrix} \sqrt{\lambda}\left(\lambda I_p - E_\tau E_0^\top E_0 E_\tau^\top\right)^{-1} & E_\tau E_0^\top \left(\lambda I_p - E_0 E_\tau^\top E_\tau E_0^\top\right)^{-1} \\ E_0 E_\tau^\top \left(\lambda I_p - E_\tau E_0^\top E_0 E_\tau^\top\right)^{-1} & \sqrt{\lambda}\left(\lambda I_p - E_0 E_\tau^\top E_\tau E_0^\top\right)^{-1} \end{pmatrix}.
\end{aligned}$$

Observe that

$$\begin{pmatrix} 0 & \alpha \\ \beta & 0 \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 0 & \beta^\top \\ \alpha^\top & 0 \end{pmatrix} = \begin{pmatrix} \alpha D \alpha^\top & \alpha C \beta^\top \\ \beta B \alpha^\top & \beta A \beta^\top \end{pmatrix}.$$

Substituting the above computations back into (4.4.2) we have

$$\begin{aligned}
0 &= \left| \begin{pmatrix} \sqrt{\lambda} I_K & -X_\tau X_0^\top \\ -X_0 X_\tau^\top & \sqrt{\lambda} I_K \end{pmatrix} - \begin{pmatrix} 0 & -X_\tau E_0^\top \\ -X_0 E_\tau^\top & 0 \end{pmatrix} \begin{pmatrix} \sqrt{\lambda} I_p & -E_\tau E_0^\top \\ -E_0 E_\tau^\top & \sqrt{\lambda} I_p \end{pmatrix}^{-1} \begin{pmatrix} 0 & -E_\tau X_0^\top \\ -E_0 X_\tau^\top & 0 \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} \sqrt{\lambda} I_K & -X_\tau X_0^\top \\ -X_0 X_\tau^\top & \sqrt{\lambda} I_K \end{pmatrix} - \begin{pmatrix} X_\tau E_0^\top \sqrt{\lambda}\left(\lambda I_p - E_0 E_\tau^\top E_\tau E_0^\top\right)^{-1} E_0 X_\tau^\top & X_\tau E_0^\top E_0 E_\tau^\top \left(\lambda I_p - E_\tau E_0^\top E_0 E_\tau^\top\right)^{-1} E_\tau X_0^\top \\ X_0 E_\tau^\top E_\tau E_0^\top \left(\lambda I_p - E_0 E_\tau^\top E_\tau E_0^\top\right)^{-1} E_0 X_\tau^\top & X_0 E_\tau^\top \sqrt{\lambda}\left(\lambda I_p - E_\tau E_0^\top E_0 E_\tau^\top\right)^{-1} E_\tau X_0^\top \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} \sqrt{\lambda}(I_K - X_\tau E_0^\top (\lambda I_p - E_0 E_\tau^\top E_\tau E_0^\top)^{-1} E_0 X_\tau^\top) & -X_\tau (I_{T-\tau} + E_0^\top E_0 E_\tau^\top (\lambda I_p - E_\tau E_0^\top E_0 E_\tau^\top)^{-1} E_\tau) X_0^\top \\ -X_0 (I_{T-\tau} + E_\tau^\top E_\tau E_0^\top (\lambda I_p - E_0 E_\tau^\top E_\tau E_0^\top)^{-1} E_0) X_\tau^\top & \sqrt{\lambda}(I_K - X_0 E_\tau^\top (\lambda I_p - E_\tau E_0^\top E_0 E_\tau^\top)^{-1} E_\tau X_0^\top) \end{pmatrix} \right|, \\
&= \left| \begin{pmatrix} \sqrt{\lambda}(I_K - X_\tau E_0^\top E_0 (\lambda I_{T-\tau} - E_\tau^\top E_\tau E_0^\top E_0)^{-1} X_\tau^\top) & -X_\tau (I_{T-\tau} + E_0^\top E_0 E_\tau^\top E_\tau (\lambda I_{T-\tau} - E_0^\top E_0 E_\tau^\top E_\tau)^{-1}) X_0^\top \\ -X_0 (I_{T-\tau} + (\lambda I_{T-\tau} - E_\tau^\top E_\tau E_0^\top E_0)^{-1} E_\tau^\top E_\tau E_0^\top E_0) X_\tau^\top & \sqrt{\lambda}(I_K - X_0 (\lambda I_{T-\tau} - E_\tau^\top E_\tau E_0^\top E_0)^{-1} E_\tau^\top E_\tau X_0^\top) \end{pmatrix} \right|,
\end{aligned}$$

where the last equality holds by (4.2.11). Recalling the notations we introduced in Section 4.2 and identity (4.2.10), we obtain

$$0 = \left| \begin{pmatrix} \sqrt{\lambda}\, \overline{Q}_\lambda & -X_\tau R_\lambda^\top X_0^\top \\ -X_0 R_\lambda X_\tau^\top & \sqrt{\lambda} Q_\lambda \end{pmatrix} \right| = \left| \overline{Q}_\lambda - \lambda^{-1} X_\tau R_\lambda^\top X_0^\top Q_\lambda^{-1} X_0 R_\lambda X_\tau^\top \right|. \tag{4.4.3}$$

Next, we center λ around the quantity θ defined in (4.3.11). Since λ and θ diverge, they are outside of the spectrum of $E_\tau^\top E_\tau E_0^\top E_0$ with probability tending to 1. Then

$$\begin{aligned}
\frac{1}{\lambda} R_\lambda - \frac{1}{\theta} R &= (\lambda I_{T-\tau} - E_\tau^\top E_\tau E_0^\top E_0)^{-1} - (\theta I_{T-\tau} - E_\tau^\top E_\tau E_0^\top E_0)^{-1} \\
&= (\theta - \lambda)(\lambda I_{T-\tau} - E_\tau^\top E_\tau E_0^\top E_0)^{-1}(\theta I_{T-\tau} - E_\tau^\top E_\tau E_0^\top E_0)^{-1} = -\frac{\delta}{\lambda} R_\lambda R.
\end{aligned}$$

Substituting back into itself, we obtain

$$\frac{1}{\lambda}R_\lambda = \frac{1}{\theta}R - \delta\left[\frac{1}{\theta}R - \frac{\delta}{\lambda}R_\lambda R\right]R = \frac{1}{\theta}R - \frac{\delta}{\theta}R^2 + \frac{\delta^2}{\lambda}R_\lambda R^2. \tag{4.4.4}$$

Using the bounds in (4.2.17) and (4.2.18) we have

$$R - I_{T-\tau} = \frac{1}{\theta}E_\tau^\top E_\tau E_0^\top E_0 R = O_{p,\|\cdot\|}(\theta^{-1}), \quad R^2 = I_{T-\tau} + O_{p,\|\cdot\|}(\theta^{-1}), \tag{4.4.5}$$

where the second equation follows from expanding $(R-I)^2$. By Theorem 4.2 we have $\delta = o_p(1)$. Substituting back into (4.4.4) we get

$$\frac{1}{\lambda}R_\lambda = \frac{1}{\theta}R - \frac{\delta}{\theta}R^2 + \delta o_{p,\|\cdot\|}(\lambda^{-1}) = \frac{1}{\theta}R - \frac{\delta}{\theta}I_{T-\tau} + \delta o_{p,\|\cdot\|}(\lambda^{-1}). \tag{4.4.6}$$

Using this we can get

$$\begin{aligned}\overline{Q}_\lambda - \overline{Q} &= (I_K - X_\tau E_0^\top E_0 \lambda^{-1} R_\lambda X_\tau^\top) - (I_K - X_\tau E_0^\top E_0 \theta^{-1} R X_\tau^\top) \\ &= X_\tau E_0^\top E_0(\theta^{-1}R - \lambda^{-1}R_\lambda)X_\tau^\top = X_\tau E_0^\top E_0\left[\frac{\delta}{\theta}I_{T-\tau} + \delta o_{p,\|\cdot\|}(\lambda^{-1})\right]X_\tau^\top.\end{aligned}$$

From (4.2.17) we recall that $\|X_\tau\|^2 = O(K\sigma_1^2)$. Using $\delta = o_p(1)$ again we get

$$\overline{Q}_\lambda = \overline{Q} + \frac{\delta}{\theta}X_\tau E_0^\top E_0 X_\tau^\top + \delta o_{p,\|\cdot\|}(K\sigma_1^2\lambda^{-1}) = \overline{Q} + \delta o_{p,\|\cdot\|}(1),$$

and similarly $Q_\lambda = Q + \delta o_{p,\|\cdot\|}(1)$. Finally, since $\|Q_\lambda^{-1}\| = O_p(1)$, we have

$$Q_\lambda^{-1} - Q^{-1} = Q_\lambda^{-1}(Q - Q_\lambda)Q^{-1} = o_{p,\|\cdot\|}(1). \tag{4.4.7}$$

Next we consider the matrix $X_0 R_\lambda X_\tau^\top$ appearing in (4.4.3). From (4.4.4) we have

$$\frac{\sqrt{\theta}}{\lambda}X_0 R_\lambda X_\tau^\top = \frac{1}{\sqrt{\theta}}X_0 R X_\tau^\top - \frac{\delta}{\sqrt{\theta}}X_0 R^2 X_\tau^\top + \frac{\delta^2\sqrt{\theta}}{\lambda}X_0 R_\lambda R^2 X_\tau^\top. \tag{4.4.8}$$

For the second term on the right hand side of (4.4.8), using (4.4.5) and (4.2.17) we have

$$\begin{aligned}\frac{\delta}{\sqrt{\theta}}X_0 R^2 X_\tau^\top &= \frac{\delta}{\sqrt{\theta}}X_0 X_\tau^\top + \frac{\delta}{\sqrt{\theta}}X_0(R^2 - I)X_\tau^\top \\ &= \frac{\delta}{\sqrt{\theta}}X_0 X_\tau^\top + \delta O_{p,\|\cdot\|}\left(\frac{K\sigma_1^2}{\theta^{3/2}}\right) = \frac{\delta}{\sqrt{\theta}}X_0 X_\tau^\top + \delta o_{p,\|\cdot\|}(1).\end{aligned}$$

Similarly the last term in (4.4.8) satisfies $\frac{\delta^2\sqrt{\theta}}{\lambda}X_0 R_\lambda R^2 X_\tau^\top = \delta o_{p,\|\cdot\|}(1)$. Therefore

$$\frac{\sqrt{\theta}}{\lambda}X_0 R_\lambda X_\tau^\top = \frac{1}{\sqrt{\theta}}X_0 R X_\tau^\top - \frac{\delta}{\sqrt{\theta}}X_0 X_\tau^\top + \delta o_{p,\|\cdot\|}(1). \tag{4.4.9}$$

To deal with the second term appearing in the determinant in (4.4.3), we first make the following computations. Using (4.4.8)-(4.4.9) as well as (4.4.7) we have

$$\begin{aligned}\frac{\theta}{\lambda^2}X_\tau R_\lambda^\top X_0^\top Q_\lambda^{-1}X_0R_\lambda X_\tau^\top &= \Big(\frac{1}{\sqrt{\theta}}X_\tau R^\top X_0^\top - \frac{\delta}{\sqrt{\theta}}X_\tau X_0^\top + \delta o_{p,\|\cdot\|}(1)\Big)\\ &\quad\times\Big(Q^{-1}+\delta o_{p,\|\cdot\|}(1)\Big)\Big(\frac{1}{\sqrt{\theta}}X_0RX_\tau^\top - \frac{\delta}{\sqrt{\theta}}X_0X_\tau^\top + \delta o_{p,\|\cdot\|}(1)\Big)\\ &= \Big(\frac{1}{\sqrt{\theta}}X_\tau R^\top X_0^\top - \frac{\delta}{\sqrt{\theta}}X_\tau X_0^\top\Big)Q^{-1}\Big(\frac{1}{\sqrt{\theta}}X_0RX_\tau^\top - \frac{\delta}{\sqrt{\theta}}X_0X_\tau^\top\Big) + \delta o_{p,\|\cdot\|}(1).\end{aligned}$$

Expanding the expression above and using (4.4.8)-(4.4.9) again we obtain

$$\begin{aligned}\frac{\theta}{\lambda^2}X_\tau R_\lambda^\top X_0^\top Q_\lambda^{-1}X_0R_\lambda X_\tau^\top &= \frac{1}{\theta}X_\tau R^\top X_0^\top Q^{-1}X_0RX_\tau^\top\\ &\quad - \frac{\delta}{\theta}\Big(X_\tau R^\top X_0^\top Q^{-1}X_0X_\tau^\top + X_\tau X_0^\top Q^{-1}X_0RX_\tau^\top\Big) + \delta o_{p,\|\cdot\|}(1)\\ &= \frac{1}{\theta}X_\tau R^\top X_0^\top Q^{-1}X_0RX_\tau^\top - \frac{2\delta}{\theta}X_\tau X_0^\top Q^{-1}X_0X_\tau^\top + \delta o_{p,\|\cdot\|}(1)\end{aligned}$$

Finally, recalling $\lambda/\theta = 1+\delta$, we can conclude

$$\begin{aligned}\frac{1}{\lambda}X_\tau R_\lambda^\top X_0^\top Q_\lambda^{-1}X_0R_\lambda X_\tau^\top &= (1+\delta)\frac{\theta}{\lambda^2}X_\tau^\top R_\lambda^\top X_0^\top Q_\lambda^{-1}X_0R_\lambda X_\tau^\top\\ &= \frac{1}{\theta}X_\tau R^\top X_0^\top Q^{-1}X_0RX_\tau^\top - \frac{2\delta}{\theta}X_\tau X_0^\top Q^{-1}X_0X_\tau^\top\\ &\quad + \delta\Big(\frac{1}{\theta}X_\tau R^\top X_0^\top Q^{-1}X_0RX_\tau^\top - \frac{2\delta}{\theta}X_\tau X_0^\top Q^{-1}X_0X_\tau^\top\Big) + \delta o_{p,\|\cdot\|}(1)\\ &= \frac{1}{\theta}X_\tau R^\top X_0^\top Q^{-1}X_0RX_\tau^\top - \frac{\delta}{\theta}X_\tau X_0^\top X_0X_\tau^\top + \delta o_{p,\|\cdot\|}(1),\end{aligned}$$

where in the last line we have used (4.4.5)-(4.4.9) again. To conclude, we have shown

$$\overline{Q}_\lambda = I_K - \frac{1}{\theta}X_\tau E_0^\top E_0RX_\tau^\top + \delta o_{p,\|\cdot\|}(1),$$

for the first term in the right hand side of (4.4.3) and

$$\frac{1}{\lambda}X_\tau R_\lambda^\top X_0^\top Q_\lambda^{-1}X_0R_\lambda X_\tau^\top = \frac{1}{\theta}X_\tau R^\top X_0^\top Q^{-1}X_0RX_\tau^\top - \frac{\delta}{\theta}X_\tau X_0^\top X_0X_\tau^\top + \delta o_{p,\|\cdot\|}(1)$$

for the second term. The claim then follows. □

We now work towards establishing the the asymptotic distribution of the matrix M from (4.3.15) with the help of Lemma 4.14-4.18. For notational convenience we define

$$A = \frac{1}{\sqrt{\theta}}X_0RX_\tau^\top, \quad B = \frac{1}{\theta}X_\tau E_0^\top E_0RX_\tau^\top, \tag{4.4.10}$$

so that $M = I_K - B - A^\top Q^{-1}A$. For each $i = 1, \ldots, K$, define

$$\overline{M}_{ii} := 1 - \mathbb{E}[B_{ii}1_{\mathcal{B}_0}] - \mathbb{E}[A_{ii}1_{\mathcal{B}_0}]^2\mathbb{E}[Q_{ii}^{-1}1_{\mathcal{B}_2}] \tag{4.4.11}$$

which serves as a deterministic centering for the i-th diagonal entry of M. We first give an approximation for $M_{ii} - \overline{M}_{ii}$ up to the scaling of $T^{-1/2}$.

Proposition 4.8. *Under Assumption 4.1 and either Assumption 4.2 or 4.3, we have*

$$M_{ii} - \overline{M}_{ii} = -2\Big(A_{ii} - \underline{\mathbb{E}}[A_{ii}]\Big)\mathbb{E}[A_{ii}1_{\mathcal{B}_0}]\mathbb{E}[Q_{ii}^{-1}1_{\mathcal{B}_2}] + o_p\bigg(\frac{1}{\sqrt{T}}\bigg), \tag{4.4.12}$$

for all $i = 1, \dots, K$, where $\underline{\mathbb{E}}[\,\cdot\,]$ is defined in (4.2.21). Furthermore,

$$\max_{i \neq j} |M_{ij}| = O_p\bigg(\frac{K^3}{\gamma_1(\tau)^2\sqrt{T}}\bigg).$$

Proof. We first recall from Lemma 4.15 and Assumption 4.1 that

$$\mathbb{E}[A_{ij}1_{\mathcal{B}_0}] = 1_{i=j}\bigg(\frac{\sigma_i^2\gamma_i(\tau)}{\theta^{1/2}} + o(1)\bigg), \quad \mathrm{Var}(A_{ij}1_{\mathcal{B}_0}) = O\bigg(\frac{\sigma_i^2\sigma_j^2}{\theta T}\bigg). \tag{4.4.13}$$

We also recall from Lemma 4.17 that

$$Q_{kk}^{-1}1_{\mathcal{B}_2} = \underline{\mathbb{E}}[Q_{kk}^{-1}1_{\mathcal{B}_2}] + o_{L^1}(T^{-1/2}), \quad Q_{ij}^{-1}1_{\mathcal{B}_2} = o_{L^2}(T^{-1}). \tag{4.4.14}$$

Recall that $M = I_K - B - A^\top Q^{-1}A$. We first consider the i-th diagonal of $A^\top Q^{-1}A$ and show that it is close to $A_{ii}^2\underline{E}[Q_{ii}^{-1}1_{\mathcal{B}_2}]$ under the event $\mathcal{B}_2$. Note that we can write

$$\begin{aligned}(A^\top Q^{-1}A)_{ii} &= \sum_{m,n} A_{mi}A_{ni}Q_{mn}^{-1} \\ &= A_{ii}^2Q_{ii}^{-1} + \sum_{m,n\neq i} A_{mi}A_{ni}Q_{mn}^{-1} + A_{ii}\bigg(\sum_{n\neq i} A_{ni}Q_{in}^{-1} + \sum_{m\neq i} A_{mi}Q_{mi}^{-1}\bigg).\end{aligned} \tag{4.4.15}$$

We will consider each term in (4.4.15) separately. Recall from (4.2.19) that $\|Q^{-1}1_{\mathcal{B}_2}\| = 1 + o(1)$ which implies $Q_{ij}^{-1}1_{\mathcal{B}_2} = 1 + o(1)$ for all $i, j \leq K$. Using the triangle inequality followed by the Cauchy Schwarz inequality we have

$$\begin{aligned}\mathbb{E}\Big|\sum_{m,n\neq i} A_{mi}A_{ni}Q_{mn}^{-1}1_{\mathcal{B}_2}\Big| &\lesssim \sum_{m,n\neq i} \mathbb{E}[A_{mi}^2 1_{\mathcal{B}_2}]^{\frac{1}{2}}\mathbb{E}[A_{ni}^2 1_{\mathcal{B}_2}]^{\frac{1}{2}} \\ &\lesssim \sum_{m,n\neq i} \mathbb{E}[A_{mi}^2 1_{\mathcal{B}_0}]^{\frac{1}{2}}\mathbb{E}[A_{ni}^2 1_{\mathcal{B}_0}]^{\frac{1}{2}} = O\bigg(\frac{K^2\sigma_i^2\sigma_m\sigma_n}{\theta T}\bigg),\end{aligned}$$

where the second inequality follows since $1_{\mathcal{B}_2} \leq 1_{\mathcal{B}_0}$ by definition and the last equality follows from (4.4.13). By Assumption 4.1 we then have

$$\sum_{m,n\neq i} A_{mi}A_{ni}Q^{-1}_{mn}1_{\mathcal{B}_2} = o_{L^1}\left(\frac{1}{\sqrt{T}}\right). \tag{4.4.16}$$

For the last term in (4.4.15), we note that $1_{\mathcal{B}_2} = 1_{\mathcal{B}_2}1_{\mathcal{B}_0}$ by definition. Using $\|Q^{-1}1_{\mathcal{B}_2}\| = 1 + o(1)$ and the triangle inequality we have

$$\mathbb{E}\Big|\sum_{n\neq i} A_{ii}A_{ni}Q^{-1}_{in}1_{\mathcal{B}_2}\Big| \lesssim \sum_{n\neq i}\mathbb{E}\Big|(A_{ii}1_{\mathcal{B}_0} - \mathbb{E}[A_{ii}1_{\mathcal{B}_0}])A_{ni}1_{\mathcal{B}_2}\Big| + \Big|\mathbb{E}[A_{ii}1_{\mathcal{B}_0}]\Big|\sum_{n\neq i}\mathbb{E}\Big|A_{ni}Q^{-1}_{in}1_{\mathcal{B}_2}\Big|.$$

By the Cauchy Schwarz inequality and (4.4.13) we have

$$\begin{aligned}\mathbb{E}\Big|A_{ii}\sum_{n\neq i} A_{ni}Q^{-1}_{in}1_{\mathcal{B}_2}\Big| &\lesssim \sum_{n\neq i}\mathrm{Var}(A_{ii}1_{\mathcal{B}_0})^{1/2}\mathbb{E}[A^2_{ni}1_{\mathcal{B}_0}]^{1/2} \\ &\quad + \mathbb{E}|A^2_{ii}1_{\mathcal{B}_0}|^{1/2}\sum_{n\neq i}\mathbb{E}[A^2_{ni}1_{\mathcal{B}_0}]^{1/2}\mathbb{E}[(Q^{-1})^2_{in}1_{\mathcal{B}_2}]^{1/2} = o\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

Combining with (4.4.16), substituting back into (4.4.15) and applying (4.4.14) we obtain

$$(A^\top Q^{-1}A)_{ii}1_{\mathcal{B}_2} = A^2_{ii}1_{\mathcal{B}_2}\underline{\mathbb{E}}[Q^{-1}_{ii}1_{\mathcal{B}_2}] + o_{L^1}\left(\frac{1}{\sqrt{T}}\right).$$

From Lemma 4.1 we know that $1_{\mathcal{B}_2} = 1 + o(1)$, therefore

$$(A^\top Q^{-1}A)_{ii} = A^2_{ii}\underline{\mathbb{E}}[Q^{-1}_{ii}1_{\mathcal{B}_2}] + o_p\left(\frac{1}{\sqrt{T}}\right). \tag{4.4.17}$$

Next, we expand A^2_{ii} around the conditional mean $\underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}]$. Note that

$$A^2_{ii}1_{\mathcal{B}_0} = \underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}]^2 + 2\underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}](A_{ii}1_{\mathcal{B}_0} - \underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}]) + (A_{ii}1_{\mathcal{B}_0} - \underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}])^2, \tag{4.4.18}$$

where by (c) of Lemma 4.14 and Assumption 4.1, the last term satisfies

$$(A_{ii}1_{\mathcal{B}_0} - \underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}])^2 = O_{L^1}\left(\frac{\sigma_i^4}{\theta T}\right) = o_{L^1}\left(\frac{1}{\sqrt{T}}\right).$$

Note that by definition of $\underline{\mathbb{E}}$ and $\mathcal{B}_0$, we have

$$A_{ii}1_{\mathcal{B}_0} - \underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}] = (A_{ii} - \underline{\mathbb{E}}[A_{ii}])1_{\mathcal{B}_0} = A_{ii} - \underline{\mathbb{E}}[A_{ii}] + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where the last equality follows from Lemma 4.1. Therefore from (4.4.18) we may obtain

$$A^2_{ii} = \underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}]^2 + 2\underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}](A_{ii} - \underline{\mathbb{E}}[A_{ii}]) + o_p\left(\frac{1}{\sqrt{T}}\right).$$

Substituting back into (4.4.17) we have

$$
\begin{aligned}
(A^\top Q^{-1}A)_{ii} &= \underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}]^2\underline{\mathbb{E}}[Q_{ii}^{-1}1_{\mathcal{B}_2}] + 2\underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}]\underline{\mathbb{E}}[Q_{ii}^{-1}1_{\mathcal{B}_2}](A_{ii} - \underline{\mathbb{E}}[A_{ii}]) + o_p\left(\frac{1}{\sqrt{T}}\right) \\
&= \mathbb{E}[A_{ii}1_{\mathcal{B}_0}]^2\mathbb{E}[Q_{ii}^{-1}1_{\mathcal{B}_2}] + 2\mathbb{E}[A_{ii}1_{\mathcal{B}_0}]\mathbb{E}[Q_{ii}^{-1}1_{\mathcal{B}_2}](A_{ii} - \underline{\mathbb{E}}[A_{ii}]) + o_p\left(\frac{1}{\sqrt{T}}\right),
\end{aligned}
$$

where in the last equality, Lemma 4.18 is used to replace the conditional expectations with the unconditional ones (except for the centering of A_{ii} where the conditional expectation is intentionally kept).

Finally, we recall $M_{ii} = 1 - B_{ii} - (A^\top Q^{-1}A)_{ii}$, so it remains to consider the matrix $B = \frac{1}{\theta}X_\tau E_0^\top E_0 R X_\tau^\top$ in the same manner as above. By Lemma 4.14, we have

$$
\mathbb{E}\left|B_{ij}1_{\mathcal{B}_0} - 1_{i=j}\underline{\mathbb{E}}[B_{ii}1_{\mathcal{B}_0}]\right|^2 \lesssim \frac{1}{\theta^2T^2}O(\sigma_i^2\sigma_j^2T) = o\left(\frac{1}{\sqrt{T}}\right). \tag{4.4.19}
$$

Using Lemma 4.18 to replace $\underline{\mathbb{E}}[B_{ii}1_{\mathcal{B}_0}]$ with $\mathbb{E}[B_{ii}1_{\mathcal{B}_0}]$ and $1_{\mathcal{B}_0} = 1 - o(1)$, we get

$$
B_{ij} = 1_{i=j}\mathbb{E}[B_{ii}1_{\mathcal{B}_0}] + o_p\left(\frac{1}{\sqrt{T}}\right).
$$

Combining the above computations, we get

$$
M_{ii} = \overline{M}_{ii} - 2\mathbb{E}[A_{ii}1_{\mathcal{B}_0}]\mathbb{E}[Q_{ii}^{-1}1_{\mathcal{B}_2}](A_{ii} - \underline{\mathbb{E}}[A_{ii}]) + o_p\left(\frac{1}{\sqrt{T}}\right)
$$

and the first claim follows.

For the off-diagonal elements, write

$$
\begin{aligned}
(A^\top Q^{-1}A)_{ij} = \sum_{m,n} A_{mi}A_{nj}Q_{mn}^{-1} &= A_{ii}A_{jj}Q_{ij}^{-1} + A_{ii}A_{ij}Q_{ii}^{-1} + A_{ji}A_{jj}Q_{jj}^{-1} \\
&+ \sum_{m\neq i, n\neq j, m\neq n} A_{mi}A_{nj}Q_{mn}^{-1} + A_{ii}\sum_{n\neq i,j} A_{nj}Q_{in}^{-1} + A_{jj}\sum_{m\neq i,j} A_{mi}Q_{mj}^{-1}.
\end{aligned} \tag{4.4.20}
$$

Observe that by definition of A_{ii}, Q_{ii} and the event $\mathcal{B}_2$ we have

$$
A_{ii}1_{\mathcal{B}_2} = O\left(\frac{K\sigma_1^2}{\sqrt{\theta}}\right) = O\left(\frac{K}{\gamma_1(\tau)}\right), \quad Q_{ij}^{-1}1_{\mathcal{B}_2} = O(1). \tag{4.4.21}
$$

Recall from (4.4.13), Lemma 4.17 and Lemma 4.15 that

$$
A_{ij}1_{\mathcal{B}_2} = O_{L^2}\left(\frac{1}{\gamma_1(\tau)\sqrt{T}}\right), \quad Q_{ij}^{-1}1_{\mathcal{B}_2} = O_{L^2}\left(\frac{1}{\gamma_1(\tau)^2\sigma_1^2\sqrt{T}}\right), \quad \forall i \neq j. \tag{4.4.22}
$$

Substituting (4.4.21) and (4.4.22) back into the terms in (4.4.20) we have

$$
A_{ii}A_{jj}Q_{ij}^{-1}1_{\mathcal{B}_2} = O_{L^2}\left(\frac{K^2}{\gamma_1(\tau)^4\sigma_1^2\sqrt{T}}\right), \quad A_{ii}A_{ij}Q_{ii}^{-1}1_{\mathcal{B}_2} = O_{L^2}\left(\frac{K}{\gamma_1(\tau)^2\sqrt{T}}\right).
$$

With similar computation, the rest of (4.4.20) satisfy

$$\sum_{m\neq i, n\neq j, m\neq n} A_{mi}A_{nj}Q^{-1}_{mn} = O_{L_1}\bigg(\frac{K^4}{\gamma_1(\tau)^2 T}\bigg), \quad A_{ii}\sum_{n\neq i,j} A_{nj}Q^{-1}_{in} = O_{L^1}\bigg(\frac{K^2}{\gamma_1(\tau)^4\sigma_1^2 T}\bigg),$$

both of which are of order $O_{L^1}(T^{-1/2})$ by Assumption 4.1. Substituting the above four estimates back into (4.4.20) we get

$$(A^\top Q^{-1}A)_{ij} = O_{L^1}\bigg(\frac{K}{\gamma_1(\tau)^2\sqrt{T}}\bigg),$$

which is uniform in i, j. Taking a union bound we obtain

$$\mathbb{P}\Big(\max_{i\neq j}(A^\top Q^{-1}A)_{ij} > \epsilon\Big) \leq \frac{1}{\epsilon}\sum_{i\neq j}\mathbb{E}|(A^\top Q^{-1}A)_{ij}| = \frac{1}{\epsilon}O\bigg(\frac{K^3}{\gamma_1(\tau)^2\sqrt{T}}\bigg),$$

or in other words we have the bound

$$\max_{i\neq j}|(A^\top Q^{-1}A)_{ij}| = O_p\bigg(\frac{K^3}{\gamma_1(\tau)^2\sqrt{T}}\bigg)$$

Lastly, since $M = I_K - B - A^\top Q^{-1}A$, it remains to bound the off-diagonals of B. Routine computations similar to (4.4.13) and the union bound above show that B_{ij} is of high order compared to $(A^\top Q^{-1}A)_{ij}$ and is thus negligible. This completes the proof. □

From Proposition 4.8 we can conclude that the CLT for M_{ii} is given by the CLT for A_{ii}, up to centering and scaling. This is what we compute next.

Proposition 4.9. *Suppose Assumption 4.1 and either Assumption 4.2 or 4.3 hold. For any $i = 1, \dots, K$ and $\tau \geq 0$, define*

$$v^2_{i,\tau} := \frac{1}{T-\tau}\mathrm{Var}\bigg(\sum_{t=1}^{T-\tau} f_{i,t}f_{i,t+\tau}\bigg).$$

a) For any i and τ, the quantity $v^2_{i,\tau}$ satisfies

$$v^2_{i,\tau} = \sum_{|k|<T-\tau}\bigg(1 - \frac{|k|}{T-\tau}\bigg)u_{i,k}(\tau), \tag{4.4.23}$$

where the sequence $(u_{i,k}(\tau))_k$ is given by

$$u_{i,k}(\tau) := \gamma_i(k)^2 + \gamma_i(k+\tau)\gamma_i(k-\tau) + (\mathbb{E}[z_{11}^4] - 3)\sum_{l=0}^{\infty}\varphi_{i,l}\varphi_{i,l+\tau}\varphi_{i,l+k}\varphi_{i,l+k+\tau}.$$

b) As $T \to \infty$, the sequence $(v^2_{i,\tau})$ tends to a limit

$$\lim_{T\to\infty} v_{i,\tau}^2 = (\mathbb{E}[z_{11}^4]-3)\gamma_i(\tau)^2 + \sum_{k\in\mathbb{Z}}\Big(\gamma_i(k)^2+\gamma_i(k+\tau)\gamma_i(k-\tau)\Big)$$

in the case where τ is a fixed constant, and

$$\lim_{T\to\infty} v_{i,\tau}^2 = \sum_{k\in\mathbb{Z}}\gamma_i(k)^2$$

in the case where $\tau=\tau_T\to\infty$ as $T\to\infty$.

c) In both cases where τ is fixed and where $\tau\to\infty$, we have

$$\frac{1}{\sqrt{T}v_{i,\tau}}\left(\sum_{t=1}^{T-\tau} f_{i,t}f_{i,t+\tau} - \mathbb{E}\left[\sum_{t=1}^{T-\tau} f_{i,t}f_{i,t+\tau}\right]\right) \Rightarrow N(0,1), \quad T\to\infty.$$

Proof. In the case where τ is fixed, the proof can be adapted from the arguments in section 7.3 of [43] so it remains to consider the case where $\tau\to\infty$. For concreteness, the following proof covers both the case where τ is finite and fixed and where τ is diverging as $T\to\infty$. For brevity of notation we will drop the subscript i (denoting the i-th factor) within the proof and write for instance $f_t := f_{it}$ and $\varphi_l := \varphi_{il}$.

With some adaptations to the computations in page 226-227 of [43], we may obtain

$$\begin{aligned} v_T^2 := \frac{1}{T-\tau}\mathrm{Var}\left(\sum_{t=1}^{T-\tau} f_t f_{t+\tau}\right) &= \frac{1}{T-\tau}\mathbb{E}\left[\sum_{t=1}^{T-\tau}\sum_{s=1}^{T-\tau} f_t f_{t+\tau} f_s f_{s+\tau}\right] - (T-\tau)\gamma(\tau)^2 \\ &= \sum_{|k|<T-\tau}\left(1-\frac{|k|}{T-\tau}\right)u_k(\tau), \end{aligned} \tag{4.4.24}$$

where $(u_k(\tau))_k$ is given by

$$u_k(\tau) := \gamma(k)^2 + \gamma(k+\tau)\gamma(k-\tau) + (\mathbb{E}[z_{11}^4]-3)\sum_{l=0}^{\infty}\varphi_l\varphi_{l+\tau}\varphi_{l+k}\varphi_{l+k+\tau}.$$

Note that the sequence $(\varphi_l)_l$ is summable and so is the sequence $(u_k(\tau))_k$. Taking the limit of (4.4.24) and invoking the dominated convergence theorem we conclude

$$v^2 := \lim_{T\to\infty} v_T^2 = \sum_{k\in\mathbb{Z}}\lim_{T\to\infty}\left(1-\frac{|k|}{T-\tau}\right)u_k(\tau).$$

In the case where τ is a fixed constant, we have, as in Proposition 7.3.1 of [43],

$$v^2 = (\mathbb{E}[z_{11}^4]-3)\gamma(\tau)^2 + \sum_{k\in\mathbb{Z}}\Big(\gamma(k)^2+\gamma(k+\tau)\gamma(k-\tau)\Big), \tag{4.4.25}$$

and in the case where τ is diverging, i.e. $\gamma(\tau)\to 0$ as $T\to\infty$, we easily see that

$$v^2 = \lim_{T\to\infty} v_T^2 = \sum_{k\in\mathbb{Z}} \gamma(k)^2. \tag{4.4.26}$$

This settles the first two claims of the proposition.

We first prove a version of the CLT for a truncated version of the factor $(f_t)_{t\geq 0}$. The truncation will be justified further below. Fix $L > 0$ and define $(f_t^{(L)})_{t=1,\dots,T}$ by

$$f_t^{(L)} := \sum_{l=0}^{L} \varphi_l z_{t-l},$$

Consider the stochastic process $(f_t^{(L)} f_{t+\tau}^{(L)})_{t=1,\dots,T-\tau}$. Clearly $(f_t^{(L)} f_{t+\tau}^{(L)})_t$ is an $(L+\tau)$-dependent process, i.e. $f_t^{(L)} f_{t+\tau}^{(L)}$ is independent from $f_s^{(L)} f_{s+\tau}^{(L)}$ whenever $|s-t| > L+\tau$. The mean is given by $\mathbb{E}[f_t^{(L)} f_{t+\tau}^{(L)}] = \gamma_L(\tau)$, where $\gamma_L(\cdot)$ is the auto-covariance function of the truncated process $(f_t^{(L)})$. Similar to (4.4.24)-(4.4.26) we may compute

$$v_{T,(L)}^2 := \frac{1}{T-\tau} \mathrm{Var}\left(\sum_{t=1}^{T-\tau} f_t^{(L)} f_{t+\tau}^{(L)} \right),$$

which has limits, in the case where τ is fixed:

$$v_{(L)}^2 := \lim_{T\to\infty} v_{T,(L)}^2 = (\mathbb{E}[z_{11}^4] - 3)\gamma_L(\tau)^2 + \sum_{k\in\mathbb{Z}} \Big(\gamma_L(k)^2 + \gamma_L(k+\tau)\gamma_L(k-\tau) \Big)$$

and in the case where $\tau \to \infty$:

$$v_{(L)}^2 = \lim_{T\to\infty} v_{T,(L)}^2 = \sum_{k\in\mathbb{Z}} \gamma_L(k)^2.$$

Note that in either case $V^{(L)}$ is a non-zero constant. It can easily be checked that, under Assumption 4.1 and either Assumption 4.2 or 4.3, the process $(f_t^{(L)} f_{t+\tau}^{(L)})_{t=1,\dots,T-\tau}$ (after centering) satisfies the conditions in [33], whose main theorem can be applied here to obtain

$$\sqrt{T-\tau} \left(\frac{1}{T-\tau} \sum_{t=1}^{T-\tau} f_t^{(L)} f_{t+\tau}^{(L)} - \gamma_L(\tau) \right) \Rightarrow N(0, v_{(L)}^2), \quad T \to \infty.$$

We now justify the truncation. Since $(\varphi_l)_l \in \ell_1$ and $(z_t)_t$ is uniformly bounded in L^4, it is easy to conclude that, for each fixed T, we have

$$\left\| \sum_{t=1}^{T-\tau} f_t^{(L)} f_{t+\tau}^{(L)} - \sum_{t=1}^{T-\tau} f_t f_{t+\tau} \right\|_{L^2} \to 0, \quad L \to \infty. \tag{4.4.27}$$

Consequently, we may conclude that $\gamma_L(\tau) \to \gamma(\tau)$ and $v_{(L)}^2 \to v^2$ as $L \to \infty$, since they are the first and second moments of the sums in (4.4.27). We may then follow the arguments on page 229 of [43] and apply Proposition 6.3.9 of [43] to obtain

$$\sqrt{T-\tau}\left(\frac{1}{T-\tau}\sum_{t=1}^{T-\tau} f_t f_{t+\tau} - \gamma(\tau)\right) \Rightarrow N(0, v^2), \quad T \to \infty. \tag{4.4.28}$$

Finally, using (4.4.25), (4.4.26) and $\frac{T-\tau}{T} \to 1$, by Slutsky's theorem we may conclude that

$$\frac{1}{\sqrt{T}v_T}\left(\sum_{t=1}^{T-\tau} f_t f_{t+\tau} - (T-\tau)\gamma_L(\tau)\right) \Rightarrow N(0,1).$$

It remains to observe that $(T-\tau)\gamma_L(\tau)$ is exactly the expectation of $\sum f_t f_{t+\tau}$ and the last claim of the proposition follows. □

Proposition 4.10. *Under Assumption 4.1 and either Assumption 4.2 or 4.3, we have*

$$\sqrt{T}\frac{\theta}{2\sigma_i^4\gamma_i(\tau)v_{i,\tau}}\left(M_{ii} - \overline{M}_{ii}\right) \Rightarrow N(0,1), \quad i = 1, \dots, K, \tag{4.4.29}$$

where $\overline{M}_{ii}$ is as defined in (4.4.11) and $v_{i,\tau}$ is defined as in (4.3.14).

Proof. For simplicity, within the current proof we will denote $\mathbf{x}_{i0} := \mathbf{x}_{i,[1:T-\tau]}, \mathbf{x}_{i\tau} := \mathbf{x}_{i,[\tau+1:T]}$ and similarly $\mathbf{f}_{i0} := \mathbf{f}_{i,[1:T-\tau]}, \mathbf{f}_{i\tau} := \mathbf{f}_{i,[\tau+1:T]}, \boldsymbol{\epsilon}_{i0} := \boldsymbol{\epsilon}_{i,[1:T-\tau]}, \boldsymbol{\epsilon}_{i\tau} := \boldsymbol{\epsilon}_{i,[\tau+1:T]}$.

Observe from (4.4.12) that the asymptotic distribution of M depends crucially on that of A. We first give the asymptotic distribution of A_{ii}. Recall from (4.2.10) that $R - I_{T-\tau} = \theta^{-1}E_\tau^\top E_\tau E_0^\top E_0 R$, using which we can write

$$\begin{aligned} A_{ii} &= \frac{1}{\sqrt{\theta T}}\mathbf{x}_{i0}^\top\mathbf{x}_{i\tau} + \frac{1}{\sqrt{\theta T}}\mathbf{x}_{i0}^\top(R - I_{T-\tau})\mathbf{x}_{i\tau} \\ &= \frac{1}{\sqrt{\theta T}}\mathbf{x}_{i0}^\top\mathbf{x}_{i\tau} + \frac{1}{\theta^{3/2}T}\mathbf{x}_{i0}^\top E_\tau^\top E_\tau E_0^\top E_0 R\mathbf{x}_{i\tau}. \end{aligned} \tag{4.4.30}$$

Applying Lemma 4.14 to the last term in (4.4.30) we have

$$\frac{1}{\theta^{3/2}T}\mathbf{x}_{i0}^\top E_\tau^\top E_\tau E_0^\top E_0 R\mathbf{x}_{i\tau}1_{\mathcal{B}_0} - \frac{1}{\theta^{3/2}T}\mathbb{E}[\mathbf{x}_{i0}^\top E_\tau^\top E_\tau E_0^\top E_0 R\mathbf{x}_{i\tau}1_{\mathcal{B}_0}] = O_{L^2}\left(\frac{\sigma_i^2}{\theta^{3/2}\sqrt{T}}\right).$$

Recalling from Theorem 4.2 that $\theta \asymp \sigma_1^4\gamma_1(\tau)^2$, we get

$$\begin{aligned} A_{ii}1_{\mathcal{B}_0} - \mathbb{E}[A_{ii}1_{\mathcal{B}_0}] &= \frac{1}{\sqrt{\theta T}}\left(\mathbf{x}_{i0}^\top\mathbf{x}_{i\tau}1_{\mathcal{B}_0} - \mathbb{E}[\mathbf{x}_{i0}^\top\mathbf{x}_{i\tau}1_{\mathcal{B}_0}]\right) + O_{L^2}\left(\frac{\sigma_i^2}{\theta^{3/2}\sqrt{T}}\right) \\ &= \frac{1}{\sqrt{\theta T}}\left(\mathbf{x}_{i0}^\top\mathbf{x}_{i\tau} - \mathbb{E}[\mathbf{x}_{i0}^\top\mathbf{x}_{i\tau}]\right)1_{\mathcal{B}_0} + O_{L^2}\left(\frac{1}{\sigma_1^4\gamma_1(\tau)^3\sqrt{T}}\right). \end{aligned} \tag{4.4.31}$$

Next, we recall that $\mathbf{x}_{i0} = \frac{1}{\sqrt{\theta T}}(\sigma_i\mathbf{f}_{i0} + \boldsymbol{\epsilon}_{i0})$ and $\mathbf{x}_{i\tau} = \frac{1}{\sqrt{\theta T}}(\sigma_i\mathbf{f}_{i\tau} + \boldsymbol{\epsilon}_{i\tau})$ so that

$$\mathbf{x}_{i0}^\top\mathbf{x}_{i\tau} = \sigma_i^2\mathbf{f}_{i0}^\top\mathbf{f}_{i\tau} + (\sigma_i\mathbf{f}_{i0}^\top\boldsymbol{\epsilon}_{i\tau} + \sigma_i\boldsymbol{\epsilon}_{i0}^\top\mathbf{f}_{i\tau} + \boldsymbol{\epsilon}_{i0}^\top\boldsymbol{\epsilon}_{i\tau}).$$

Applying Lemma 4.13 to the three terms in parenthesis on the right hand we get

$$\mathbf{x}_{i0}^\top \mathbf{x}_{i\tau} - \mathbb{E}[\mathbf{x}_{i0}^\top \mathbf{x}_{i\tau}] = \sigma_i^2 \mathbf{f}_{i0}^\top \mathbf{f}_{i\tau} - \sigma_i^2 \mathbb{E}[\mathbf{f}_{i0}^\top \mathbf{f}_{i\tau}] + O_{L^2}(\sigma_i \sqrt{T}). \tag{4.4.32}$$

Substituting back into (4.4.31) and using $\theta \asymp \sigma_1^4 \gamma_1(\tau)^2$ again we obtain

$$\begin{aligned} A_{ii} 1_{\mathcal{B}_0} - \mathbb{E}[A_{ii} 1_{\mathcal{B}_0}] &= \frac{\sigma_i^2}{\sqrt{\theta T}} \Big(\mathbf{f}_{i0}^\top \mathbf{f}_{i\tau} - \mathbb{E}[\mathbf{f}_{i0}^\top \mathbf{f}_{i\tau}] \Big) 1_{\mathcal{B}_0} + O_{L^2}\left(\frac{\sigma_i}{\sqrt{\theta T}} \right) + O_{L^2}\left(\frac{1}{\sigma_1^4 \sqrt{T} \gamma_1(\tau)^3} \right) \\ &= \frac{\sigma_i^2}{\sqrt{\theta T}} \Big(\mathbf{f}_{i0}^\top \mathbf{f}_{i\tau} - \mathbb{E}[\mathbf{f}_{i0}^\top \mathbf{f}_{i\tau}] \Big) 1_{\mathcal{B}_0} + O_{L^2}\left(\frac{1}{\sigma_1 \gamma_1(\tau) \sqrt{T}} \right). \end{aligned} \tag{4.4.33}$$

Rescaling and recalling that $1_{\mathcal{B}_0} = 1 - o_p(T^{-l})$ for any $l \geq 1$, we have

$$\sqrt{T} \frac{\sqrt{\theta}}{\sigma_i^2} (A_{ii} - \mathbb{E}[A_{ii}]) = \frac{1}{\sqrt{T}} \Big(\mathbf{f}_{i0}^\top \mathbf{f}_{i\tau} - \mathbb{E}[\mathbf{f}_{i0}^\top \mathbf{f}_{i\tau}] \Big) + O_p(\sigma_1^{-1}).$$

From this we observe that we can obtain a CLT for A_{ii} from a CLT for the auto-covariance function of $\mathbf{f}_i$. Indeed, by Proposition 4.9 we have

$$\sqrt{T} \frac{\sqrt{\theta}}{\sigma_i^2 v_{i,\tau}} (A_{ii} - \mathbb{E}[A_{ii}]) \Rightarrow N(0,1), \quad p, T \to \infty, \tag{4.4.34}$$

where $v_{i,\tau}$ is specified in the statement of Proposition 4.9.

Finally, we recall from Proposition 4.8 that

$$M_{ii} - \overline{M}_{ii} = -2 \Big(A_{ii} - \mathbb{E}[A_{ii}] \Big) \mathbb{E}[A_{ii} 1_{\mathcal{B}_0}] \mathbb{E}[Q_{ii}^{-1} 1_{\mathcal{B}_2}] + o_p\left(\frac{1}{\sqrt{T}} \right). \tag{4.4.35}$$

In order to apply the CLT in (4.4.34) to (4.4.35), we need to divide (4.4.35) by the coefficient of $A_{ii} - \mathbb{E}[A_{ii}]$, which requires it to be bounded away from zero. Indeed, we recall from (4.2.19) that $Q_{ii}^{-1} 1_{\mathcal{B}_2} = 1 + o(1)$. Furthermore, from Lemma 4.15 we have

$$\mathbb{E}[A_{ii} 1_{\mathcal{B}_0}] = \frac{\sigma_i^2 \gamma_i(\tau)}{\sqrt{\theta}} + o(1) \tag{4.4.36}$$

which is bounded from below as well for large T. Therefore from (4.4.35) we get

$$\frac{\sqrt{\theta}}{-2\sigma_i^2 \gamma_i(\tau)} \Big(M_{ii} - \overline{M}_{ii} \Big) = \Big(A_{ii} - \mathbb{E}[A_{ii}] \Big) (1 + o(1)) + o_p\left(\frac{1}{\sqrt{T}} \right),$$

and the claim follows then from the CLT in (4.4.34). □

The asymptotic distribution of M_{ii} proved in Proposition 4.10 is the last result we need to prove the main result of the paper, which we present below.

Proof of Theorem 4.5. Recall from Section 4.3 that up to now we have dealt with, without loss of generality, the n-th largest eigenvalue $\lambda := \lambda_n$ and the corresponding $\theta := \theta_n$ and $\delta := \delta_n := \lambda_n/\theta_n - 1$. Recall from Proposition 4.7 that δ_n satisfies

$$\det\left(M + \frac{\delta_n}{\theta_n} X_\tau X_0^\top X_0 X_\tau^\top + \delta o_{p,\|\cdot\|}(1)\right) = 0 \tag{4.4.37}$$

We first consider the asymptotic properties of the elements of the matrix M. From Proposition 4.4 and the definition of $\overline{M}_{ii}$ in (4.4.11), clearly we see $\overline{M}_{nn} = 0$. From Theorem 4.2 and Proposition we also recall that 4.4 that $\theta_n/(\sigma_n^4\gamma_n(\tau)^2) = \theta_n/\mu_{n,\tau}^2 = 1+o(1)$. Then, using Proposition 4.10 we immediately have

$$\sqrt{T}\frac{\gamma_n(\tau)}{2v_{n,\tau}} M_{nn} \Rightarrow N(0,1), \quad T \to \infty.$$

For $i \neq n$, we recall from (4.4.11) and Lemma 4.15 that

$$\overline{M}_{ii} = 1 - \frac{\sigma_i^4\gamma_i(\tau)^2}{\theta_n} + o(1) = 1 - \frac{\mu_{i,\tau}^2}{\mu_{n,\tau}^2} + o(1) \asymp 1,$$

where the last equality is due to Assumption 4.1. Using Proposition 4.8 we have

$$M_{ii} \asymp 1 + o_p(1), \quad \forall i \neq n, \quad \max_{i\neq j} |M_{ij}| = O_p\left(\frac{K^3}{\gamma_1(\tau)^2\sqrt{T}}\right).$$

Next, recall $\delta = o_p(1)$ from Theorem 4.2. From (4.3.6) recall that

$$\|\theta^{-1} X_\tau X_0^\top X_0 X_\tau^\top - \theta^{-1}\mathrm{diag}(\sigma_i^4\gamma_i(\tau)^2)\|_\infty = O_p\left(\frac{K^2\sigma_1^4\gamma_1(\tau)}{\theta\sqrt{T}}\right) = o_p(1).$$

This in particular implies $\delta\theta^{-1}(X_\tau X_0^\top X_0 X_\tau^\top)_{ii} = \delta\theta^{-1}\sigma_i^4\gamma_i(\tau)^2 + o_p(1) = o_p(1)$ and

$$\frac{\delta}{\theta}(X_\tau X_0^\top X_0 X_\tau^\top)_{ij} = o_p(\delta), \quad \forall i \neq j.$$

Combining the above, equation (4.4.37) becomes $\det(Q) = 0$, where Q is a matrix satisfying

$$Q_{nn} = M_{nn} + \delta_n \frac{\sigma_n^4\gamma_n(\tau)^2}{\theta_n} + \delta_n o_p(1),$$

for its n-th diagonal element, $Q_{ii} \asymp 1 + o_p(1), \quad \forall i \neq n$ and

$$\sup_{ij} Q_{ij} = O_p\left(\frac{K^3}{\gamma_1(\tau)^2\sqrt{T}}\right) + \delta o_p(1). \tag{4.4.38}$$

Using Leibniz's formula to compute $\det(Q)$, we have

$$0 = \det(Q) = \sum_{\pi\in S_K} \mathrm{sgn}(\pi) \prod_{i=1}^{K} Q_{i,\pi(i)}, \tag{4.4.39}$$

where $\mathrm{sgn}(\pi)$ is the sign of a permutation π in the symmetry group S_K. Next we show that $\prod_i Q_{ii}$ is the leading term in the sum in (4.4.39). Write $S_{K,k}$ for the subgroup of permutations that has exactly $K-k$ fixed points, i.e.

$$S_{K,k} = \{\pi \in S_K, i = \pi(i) \text{ for exactly } K-k \text{ such } i\text{'s}\}.$$

Using this notation we can rewrite (4.4.39) into

$$0 = \det(Q) = \sum_{k=0}^{K} \sum_{\pi \in S_{K,k}} \operatorname{sgn}(\pi) \prod_{i=1}^{K} Q_{i,\pi(i)}. \tag{4.4.40}$$

We recall that the order of $S_{K,k}$ is given by the rencontres numbers (see [149])

$$|S_{K,k}| = D_{K,K-k} := \frac{K!}{(K-k)!} \sum_{i=0}^{k} \frac{(-1)^i}{i!}.$$

Observe that $|S_{K,0}| = 1$ since $S_{K,0}$ contains only the identity permutation and $|S_{K,1}| = 0$ since for any non-identity permutation π, there exists at least two indices $i, j \in \{1, \ldots, K\}$, $i \neq j$ such that $i \neq \pi(i)$ and $j \neq \pi(j)$. Therefore (4.4.40) becomes

$$0 = \det(Q) = \prod_{i=1}^{K} Q_{ii} + \sum_{k=2}^{K} \sum_{\pi \in S_{K,k}} \operatorname{sgn}(\pi) \prod_{i=1}^{K} Q_{i,\pi(i)}. \tag{4.4.41}$$

Note that for any $k \geq 2$ and any permutation $\pi \in S_{K,k}$, the product $\prod_{i=1}^{K} Q_{i,\pi(i)}$ contains exactly k off-diagonal elements of Q. By (4.4.38) we have the estimate

$$\prod_{i=1}^{K} Q_{i,\pi(i)} = \left(O_p\left(\frac{K^3}{\gamma_1(\tau)^2 \sqrt{T}} \right) + \delta o_p(1) \right)^k.$$

Finally, after substituting back into (4.4.41) and a lengthy computation we have

$$0 = \det(Q) = \prod_{i=1}^{K} Q_{ii} + \delta o_p(1) + o_p(T^{-1/2}), \tag{4.4.42}$$

which shows that the product $\prod_{i=1}^{K} Q_{ii}$ is the leading term of $\det(Q)$.

Next, using (4.4.38) again we see that $\prod_{i=1}^{K} Q_{ii}$ can be written as

$$\prod_{i=1}^{K} Q_{ii} = \left(M_{nn} + \delta_n \frac{\sigma_n^4 \gamma_n(\tau)^2}{\theta_n} + \delta_n o_p(1) \right)(1 + o_p(1)) = (M_{nn} + \delta_n)(1 + o_p(1)),$$

which can be substituted back into (4.4.42) to obtain

$$0 = M_{nn}(1 + o_p(1)) + \delta_n(1 + o_p(1)) + \delta_n o_p(1) + o_p(T^{-1/2}).$$

Rearranging (and recalling $\theta_n \asymp \sigma_n^4 \gamma_n(\tau)^2$ by Proposition 4.4), we finally get

$$-\sqrt{T}\frac{\gamma_n(\tau)}{2v_{n,\tau}}\delta_n(1+o_p(1)) = \sqrt{T}\frac{\theta_n}{2\sigma_n^4\gamma_n(\tau)v_{n,\tau}}M_{nn}(1+o_p(1)) + o_p(1).$$

Applying Proposition (4.10) we immediately have

$$\sqrt{T}\frac{\gamma_n(\tau)}{2v_{n,\tau}}\delta_n \Rightarrow N(0,1), \quad T\to\infty$$

and the proof is complete. □

4.5 Technical Lemmas

4.5.1 Estimates on quadratic forms

We start with the proof of Lemma 4.1.

Proof of Lemma 4.1. (a) of the lemma can be found in [52], here we give a proof for (b). Since $X_0^\top X_0$ is symmetric and positive definite, we have

$$\|X_0^\top X_0\| \le \mathrm{tr}(X_0^\top X_0) = \frac{1}{T}\sum_{i=1}^K\sum_{t=1}^{T-\tau} x_{it}^2 \le \sum_{i=1}^K \frac{1}{T}\sum_{t=1}^T x_{it}^2.$$

By (a) of Lemma 4.14 (whose proof does not depend the current lemma) we have

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^T x_{it}^2\right] = \sigma_i^2+1, \quad \mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^T x_{it}^2\right) = O\left(\frac{\sigma_i^4}{T}\right).$$

Taking a union bound and applying Chebyshev's inequality we have

$$\begin{aligned}\mathbb{P}\left(\|X_0^\top X_0\| > 2\sum_{i=1}^K \sigma_i^2\right) &\le \mathbb{P}\left(\sum_{i=1}^K\frac{1}{T}\sum_{t=1}^T x_{it}^2 > 2\sum_{i=1}^K\sigma_i^2\right) \le \sum_{i=1}^K \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^T x_{it}^2 > 2\sigma_i^2\right)\\ &= \sum_{i=1}^K \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^T x_{it}^2 - (\sigma_i^2+1) > \sigma_i^2 - 1\right) = O\left(\frac{K\sigma_i^4}{(\sigma_i^2-1)^2 T}\right) = O\left(\frac{K}{T}\right)\end{aligned}$$

and the proof is complete. □

Lemma 4.11 (Sherman-Morrison formula)**.** *Suppose A and B are invertible matrices of the same dimension, such that $A-B$ is of rank one. Then*

$$A^{-1} - B^{-1} = -\frac{B^{-1}(A-B)B^{-1}}{1+\mathrm{tr}(B^{-1}(A-B))}. \tag{4.5.1}$$

Further more, if $A - B = \mathbf{u}\mathbf{v}^\top$, then

$$A^{-1}\mathbf{u} = \frac{B^{-1}\mathbf{u}}{1+\mathbf{v}^\top B^{-1}\mathbf{u}}, \quad \mathbf{v}^\top A^{-1} = \frac{\mathbf{v}^\top B^{-1}}{1+\mathbf{v}^\top B^{-1}\mathbf{u}}. \tag{4.5.2}$$

We first establish some concentration inequalities for quadratic forms of the random vector $\mathbf{x}$. To do so we will need to introduce some notations. We recall that

$$x_{it} = \sigma_i f_{it} + \epsilon_{it} = \sigma_i \sum_{l=0}^{\infty} \varphi_{il} z_{i,t-l} + \epsilon_{it}, \quad i = 1, \ldots, K, \quad t = 1, \ldots, T.$$

We truncate the series and define an approximation

$$x_{it}^{(L)} := \sigma_i f_{it}^{(L)} + \epsilon_{it} := \sigma_i \sum_{l=0}^{L} \varphi_{il} z_{i,t-l} + \epsilon_{it}, \quad L \geq 1, \tag{4.5.3a}$$

and write $\mathbf{x}_{i,[1:T]}^{(L)}$, $\mathbf{f}_{i,[1:T]}^{(L)}$ for $(x_{it}^{(L)})_{t=1,\ldots,T}$ and $(f_{it}^{(L)})_{t=1,\ldots,T}$. For each L, We write

$$\underline{\boldsymbol{\varphi}}_i^\top := \Big(\varphi_{iL}, \ldots, \varphi_{i0}, \mathbf{0}_{T-1}^\top\Big) \in \mathbb{R}^{T+L}. \tag{4.5.3b}$$

Let S be the right-shift operator on $\mathbb{R}^{T+L}$, i.e. $S\mathbf{e}_i = \mathbf{e}_{i+1}$. Define

$$\Phi_i := \Big(\underline{\boldsymbol{\varphi}}_i, S\underline{\boldsymbol{\varphi}}_i, \ldots, S^{T-1}\underline{\boldsymbol{\varphi}}_i\Big) \in \mathbb{R}^{(T+L)\times T}, \tag{4.5.3c}$$

then clearly we can write the approximation $\mathbf{f}_i^{(L)}$ into

$$\mathbf{f}_{i,[1:T]}^{(L)} = \mathbf{z}_{i,[1-L,T]}^\top \Phi_i. \tag{4.5.3d}$$

We note that the space of $n \times n$ matrices equipped with the Frobenius norm is isometrically isomorphic to $\mathbb{R}^{n\times n}$ with the Euclidean norm. For each $1 \leq i, j \leq K$, we define linear operators Ψ_n^{ij}, $n = 0, 1, 2$,

$$\Psi_n^{ij} : \mathbb{R}^{(T-\tau)\times(T-\tau)} \to \mathbb{R}^{(2T+L)\times(2T+L)}$$

by sending a $(T-\tau)\times(T-\tau)$ matrix B to the $(2T+L)\times(2T+L)$ matrices

$$\begin{aligned}
\Psi_0^{ij} B &:= \begin{pmatrix} \sigma_i \Phi_i \\ I_T \end{pmatrix} \begin{pmatrix} I_{T-\tau} \\ \mathbf{0}_{\tau\times(T-\tau)} \end{pmatrix} B \begin{pmatrix} I_{T-\tau} \\ \mathbf{0}_{\tau\times(T-\tau)} \end{pmatrix}^\top \begin{pmatrix} \sigma_j \Phi_j \\ I_T \end{pmatrix}^\top, \\
\Psi_1^{ij} B &:= \begin{pmatrix} \sigma_i \Phi_i \\ I_T \end{pmatrix} \begin{pmatrix} I_{T-\tau} \\ \mathbf{0}_{\tau\times(T-\tau)} \end{pmatrix} B \begin{pmatrix} \mathbf{0}_{\tau\times(T-\tau)} \\ I_{T-\tau} \end{pmatrix}^\top \begin{pmatrix} \sigma_j \Phi_j \\ I_T \end{pmatrix}^\top, \\
\Psi_2^{ij} B &:= \begin{pmatrix} \sigma_i \Phi_i \\ I_T \end{pmatrix} \begin{pmatrix} \mathbf{0}_{\tau\times(T-\tau)} \\ I_{T-\tau} \end{pmatrix} B \begin{pmatrix} \mathbf{0}_{\tau\times(T-\tau)} \\ I_{T-\tau} \end{pmatrix}^\top \begin{pmatrix} \sigma_j \Phi_j \\ I_T \end{pmatrix}^\top,
\end{aligned} \tag{4.5.4}$$

where $\Phi_i := (\underline{\boldsymbol{\varphi}}_i, S\underline{\boldsymbol{\varphi}}_i \ldots, S^{T-1}\underline{\boldsymbol{\varphi}}_i) \in \mathbb{R}^{(T+L)\times T}$ is as defined in (4.5.3c). We first give some estimates on the operators Ψ_n^{ij}.

Lemma 4.12. *The following estimates hold uniformly in $L \in \mathbb{N}$.*

a) The matrix $\Phi_i^\top \Phi_i$ is symmetric and (banded) Toeplitz with

$$\sup_i \|\Phi_i^\top \Phi_i\| \le 1 + \sup_i \|\boldsymbol{\varphi}_i\|_{\ell_1}^2 = O(1).$$

b) For $n = 0, 1, 2$, the operator norms of Ψ_n^{ij} be bounded by

$$\|\Psi_n^{ij}\|^2 \le \left(1 + \sigma_i^2 \|\boldsymbol{\varphi}_i\|_{\ell_1}^2\right)\left(1 + \sigma_j^2 \|\boldsymbol{\varphi}_j\|_{\ell_1}^2\right) = O(\sigma_i^2 \sigma_j^2).$$

c) For any $B \in \mathbb{R}^{T\times T}$, the trace of $\Psi_n^{ii} B$ can be bounded by

$$\left|\operatorname{tr}(\Psi_n^{ii} B)\right| \le (T - \tau)(1 + \sigma_i^2 \|\boldsymbol{\varphi}_i\|_{\ell_1}^2)\|B\| = O(\sigma_i^2 (T - \tau)\|B\|).$$

Proof. (a) From the definitions (4.5.3b) and (4.5.3c) we immediately have

$$(\Phi_i^\top \Phi_i)_{s,t} = 1_{|s-t|\le L} \underline{\boldsymbol{\varphi}}_i^\top S^{|s-t|} \underline{\boldsymbol{\varphi}}_i = 1_{|s-t|=k\le L} \sum_{l=0}^{L-k} \varphi_{i,l+k}\varphi_{i,l}.$$

It is clear that $\Phi_i^\top \Phi_i$ is a banded, symmetric Toeplitz matrix. The operator norm of $\Phi_i^\top \Phi_i$ is controlled by the supremum of its symbol over $\mathbb{C}$ (see [39]) and we have

$$\|\Phi_i^\top \Phi_i\| \le \sup_{\lambda\in\mathbb{C}} \left| \sum_{|k|=0}^{L} \underline{\boldsymbol{\varphi}}_i^\top S^{|k|} \underline{\boldsymbol{\varphi}}_i e^{\sqrt{-1}k\lambda} \right| \le \|\underline{\boldsymbol{\varphi}}_i\|_{\ell_2}^2 + \sum_{k=1}^{L} \sum_{l=0}^{L-k} |\varphi_{i,l+k}\varphi_{i,l}| \le 1 + \|\boldsymbol{\varphi}_i\|_{\ell_1}^2,$$

which is bounded uniformly in $i = 1, \dots, K$, due to Assumption (4.1).

(b) By the cyclic property of the trace and Cauchy-Schwarz inequality we get

$$\begin{aligned}
\|\Psi_1^{ij} B\|_F^2 &= \operatorname{tr}\Big((\Psi_1^{ij} B)(\Psi_1^{ij} B)^\top\Big) \\
&= \operatorname{tr}\Big((I_T + \sigma_i^2 \Phi_i^\top \Phi_i)(I_{T-\tau}, \mathbf{0})^\top B(\mathbf{0}, I_{T-\tau})(I_T + \sigma_j^2 \Phi_j^\top \Phi_j)(\mathbf{0}, I_{T-\tau})^\top B^\top (I_{T-\tau}, \mathbf{0})\Big). \\
&\le \Big|(I_T + \sigma_i^2 \Phi_i^\top \Phi_i)(I_{T-\tau}, \mathbf{0})^\top B(I_{T-\tau}, \mathbf{0})\Big|_{\mathrm{F}} \Big|(I_T + \sigma_j^2 \Phi_j^\top \Phi_j)(\mathbf{0}, I_{T-\tau})^\top B^\top (I_{T-\tau}, \mathbf{0})\Big|_{\mathrm{F}}.
\end{aligned}$$

Since $\|AB\|_F \le \|A\|\|B\|_F$, we have

$$\|\Psi_1^{ij} B\|_{\mathrm{F}}^2 \le \|I_T + \sigma_i^2 \Phi_i^\top \Phi_i\| \|I_T + \sigma_j^2 \Phi_j^\top \Phi_j\| \|B\|_F^2,$$

where $\|I_T + \sigma_i^2 \Phi_i^\top \Phi_i\| \le 1 + \sigma_i^2 \|\boldsymbol{\varphi}_i\|_{\ell_1}^2$ by the first claim of the Lemma. By identifying Ψ_1^{ij} as an operator between spaces of matrices equipped with the Frobenius norm, this translates to a bound on its spectral norm. The case of Ψ_0 and Ψ_2 hold analogously.

(c) For the last bound, similar computations give

$$|\operatorname{tr}(\Psi_0^{ii} B)| = \left|\operatorname{tr}\Big((I_T + \sigma_i^2 \Phi_i^\top \Phi_i)(I_{T-\tau}, \mathbf{0})^\top B(I_{T-\tau}, \mathbf{0})\Big)\right|$$

$$\leq \|I_T + \sigma_i^2 \Phi_i^\top \Phi_i\| \|B\|_F \leq (T-\tau)(1+\sigma_i^2 \|\boldsymbol{\varphi}_i\|_{\ell_1}^2)\|B\|$$

The rest of the claims hold similarly. □

Next, we state an easy extension to Lemma 2.7 of [13] suited to our needs.

Lemma 4.13. *Let $\mathbf{z} = (\mathbf{z}_1^\top, \mathbf{z}_2^\top)^\top$, where $\mathbf{z}_1 = (z_1, \dots, z_m)$ and $\mathbf{z}_2 = (\tilde{z}_1, \dots, \tilde{z}_n)$ are independent random vectors each with i.i.d. entries satisfying $\mathbb{E}[z_1] = \mathbb{E}[\tilde{z}_1] = 0$, $\mathbb{E}[z_1^2] = \mathbb{E}[\tilde{z}_1^2] = 1$, $\nu_q := \mathbb{E}|z_1|^q < \infty$ and $\tilde{\nu}_q := \mathbb{E}|\tilde{z}_1|^q < \infty$ for some $q \in [1, \infty)$.*

a) Let C be a deterministic $m \times n$ matrix, then

$$\mathbf{z}_1^\top C \mathbf{z}_2 = O_{L^q}(\|C\|_F),$$

where the constant in the estimate depends only on q and $\nu_q, \tilde{\nu}_q$.

b) Let M be a deterministic $(m+n) \times (m+n)$ matrix, then

$$\mathbf{z}^\top M \mathbf{z} - \mathrm{tr} M = O_{L^q}(\|M\|_F),$$

where the constant in the estimate depends only on q and $\nu_k, \tilde{\nu}_k$ for $k \leq 2q$.

Proof. (a) By Lemma 2.2 and Lemma 2.3 of [13] we have

$$\begin{aligned}
\mathbb{E}\Big|\mathbf{z}_1^\top C \mathbf{z}_2\Big|^q &= \mathbb{E}\Big|\sum_{i,j} z_i \tilde{z}_j C_{ij}\Big|^q \lesssim \mathbb{E}\Big|\sum_{i,j} z_i^2 \tilde{z}_j^2 C_{ij}^2\Big|^{q/2} \\
&\lesssim \Big(\sum_{i,j} \mathbb{E}[z_i^2 \tilde{z}_j^2 C_{ij}^2]\Big)^{q/2} + \sum_{i,j} \mathbb{E}[|z_i|^q |\tilde{z}_j|^q |C_{ij}|^q] \\
&= \Big(\sum_{i,j} M_{ij}^2\Big)^{q/2} + \nu_q \tilde{\nu}_q \sum_{i,j} |C_{ij}|^q \leq (1 + \nu_q \tilde{\nu}_q) \|C\|_F^q,
\end{aligned}$$

where the last inequality holds since $\sum |C_{ij}|^q \leq (\sum |C_{ij}|^2)^{q/2}$ for $q \geq 2$.
(b) Write $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ where A, B, C, D are of dimensions such that

$$\mathbf{z}^\top M \mathbf{z} = \mathbf{z}_1^\top A \mathbf{z}_1 + \mathbf{z}_1^\top B \mathbf{z}_2 + \mathbf{z}_2^\top C \mathbf{z}_1 + \mathbf{z}_2^\top D \mathbf{z}_2.$$

By Lemma 2.7 of Bai and Silverstein [13] we have

$$\begin{aligned}
\mathbb{E}\Big|\mathbf{z}_1^\top A \mathbf{z}_1 - \mathrm{tr} A\Big|^q &\lesssim (\nu_4^{q/2} + \nu_{2q}) \mathrm{tr}(AA^\top)^{q/2} \leq (\nu_4^{q/2} + \nu_{2q}) \|M\|_F^q, \\
\mathbb{E}\Big|\mathbf{z}_2^\top D \mathbf{z}_2 - \mathrm{tr} D\Big|^q &\lesssim (\tilde{\nu}_4^{q/2} + \tilde{\nu}_{2q}) \mathrm{tr}(DD^\top)^{q/2} \leq (\nu_4^{q/2} + \nu_{2q}) \|M\|_F^q.
\end{aligned}$$

Then we can write

$$\mathbb{E}|\mathbf{z}^\top M\mathbf{z} - \mathrm{tr}M|^q \lesssim \mathbb{E}\Big|\mathbf{z}_1^\top A\mathbf{z}_1 - \mathrm{tr}A\Big|^q + \mathbb{E}\Big|\mathbf{z}_2^\top B\mathbf{z}_2 - \mathrm{tr}B\Big|^q \\ + \mathbb{E}\Big|\mathbf{z}_1^\top C\mathbf{z}_2\Big|^q + \mathbb{E}\Big|\mathbf{z}_2^\top D\mathbf{z}_1\Big|^q.$$

and the claim follows from (a) of the lemma. □

Using Lemma 4.12 and Lemma 4.13 we can derive the following concentration inequalities for quadratic forms involving certain high probability events.

Lemma 4.14. *Let $\mathbf{x}, \mathbf{f}, \boldsymbol{\epsilon}$ be defined as in (4.2.7) and $q \leq 2$. Under Assumptions 4.1 and either Assumptions 4.2 or 4.3 we have*

a) For any (deterministic) square matrix B of size $T - \tau$, we have

$$\mathbf{x}_{i,[1:T-\tau]}^\top B\mathbf{x}_{j,[\tau+1:T]} - \mathbb{E}[\mathbf{x}_{i,[1:T-\tau]}^\top B\mathbf{x}_{j,[\tau+1:T]}] = O_{L^q}\Big(\sigma_i\sigma_j\sqrt{T}\|B\|\Big),$$

where the expectation is satisfies

$$\mathbb{E}[\mathbf{x}_{i,[1:T-\tau]}^\top B\mathbf{x}_{j,[\tau+1:T]}] = 1_{i=j}\mathrm{tr}\Big(\Psi_1^{ii}(B)\Big) = 1_{i=j}O(\sigma_i^2 T\|B\|).$$

b) For all i, j we have $\mathbb{E}[\mathbf{f}_{i,[1:T-\tau]}^\top B\boldsymbol{\epsilon}_{j,[\tau+1:T]}] = 0$ and

$$\mathbf{f}_{i,[1:T-\tau]}^\top B\boldsymbol{\epsilon}_{j,[\tau+1:T]} = O_{L^{2q}}(\sigma_i\sqrt{T}\|B\|).$$

c) Suppose $n \in \{1, \dots, K\}$ and c_1, c_2 are positive constants with $c_1 < c_2$. Pick any

$$a \in [c_1, c_2]\mu_{n,\tau}^2.$$

Recall from (4.2.9) the resolvent $R(a) := (I_{T-\tau} - a^{-1}E_\tau^\top E_\tau E^\top E)^{-1}$, then

$$\mathbf{x}_{i,[1,T-\tau]}^\top R(a)^k\mathbf{x}_{j,[\tau+1:T]}1_{\mathcal{B}_0} - \underline{\mathbb{E}}[\mathbf{x}_{i,[1,T-\tau]}^\top R(a)^k\mathbf{x}_{j,[\tau+1:T]}1_{\mathcal{B}_0}] = O_{L^q}(\sigma_i\sigma_j\sqrt{T}),$$

for all $k \in \mathbb{N}$, where $\underline{\mathbb{E}}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|\mathcal{F}_p]$ is defined in (4.2.21). In particular,

$$\mathbf{x}_{i,[1,T-\tau]}^\top R(a)^k\mathbf{x}_{j,[\tau+1:T]} - \underline{\mathbb{E}}[\mathbf{x}_{i,[1,T-\tau]}^\top R(a)^k\mathbf{x}_{j,[\tau+1:T]}1_{\mathcal{B}_0}] = O_p(\sigma_i\sigma_j\sqrt{T}).$$

d) Parts (a)-(c) of the lemma remain true if the vector $\mathbf{x}_{j,[\tau+1:T]}$ is replaced by $\mathbf{x}_{j,[1:T-\tau]}$ and the operator Ψ_1^{ii} is replaced by Ψ_0^{ii}.

Proof. (a) We apply the truncation procedure as described in (4.5.3a). Recalling (4.5.3a), (4.5.3d) and (4.5.4) we may write

$$\mathbf{x}_{i,[1:T-\tau]}^{(L)\top} B\mathbf{x}_{j,[\tau+1:T]}^{(L)} = (\sigma_i\mathbf{f}_{i,[1:T]}^{(L)} + \boldsymbol{\epsilon}_{i,[1:T]})^\top \begin{pmatrix} I_{T-\tau} \\ \mathbf{0} \end{pmatrix} B \begin{pmatrix} \mathbf{0} \\ I_{T-\tau} \end{pmatrix}^\top (\sigma_j\mathbf{f}_{j,[1:T]}^{(L)} + \boldsymbol{\epsilon}_{j,[1:T]})$$

$$= (\mathbf{z}_{i,[1-L:T]}^\top, \boldsymbol{\epsilon}_{i,[1:T]}^\top)(\Psi_1^{ij}(B))(\mathbf{z}_{i,[1-L:T]}^\top, \boldsymbol{\epsilon}_{i,[1:T]}^\top)^\top.$$

Applying (b) of Lemma 4.13 to the above quadratic form gives

$$\mathbb{E}\Big|\mathbf{x}_{i,[1:T-\tau]}^{(L)\top} B\mathbf{x}_{j,[\tau+1:T]}^{(L)} - \mathbb{E}\Big[\mathbf{x}_{i,[1:T-\tau]}^{(L)\top} B\mathbf{x}_{j,[\tau+1:T]}^{(L)}\Big]\Big|^q \lesssim \|\Psi_1^{ij}(B)\|_F^q, \tag{4.5.5}$$

where $\mathbb{E}[\mathbf{x}_{i,[1:T]}^{(L)\top} B\mathbf{x}_{j,[1:T]}^{(L)}] = 1_{i=j}\mathrm{tr}(\Psi_1^{ii}(B))$. Using Lemma 4.12 we see that

$$\|\Psi_0^{ij}(B)\|_F^q = O\Big(\sigma_i^q\sigma_j^q\Big)\|B\|_F^q = (T-\tau)^{q/2}O\Big(\sigma_i^q\sigma_j^q\Big)\|B\|^q = T^{q/2}O\Big(\sigma_i^q\sigma_j^q\Big)\|B\|^q,$$

and

$$\mathbb{E}[\mathbf{x}_{i,[1:T]}^{(L)\top} B\mathbf{x}_{j,[1:T]}^{(L)}] = 1_{i=j}O(\sigma_i^2(T-\tau))\|B\| = 1_{i=j}O(\sigma_i^2 T)\|B\|,$$

both of which are uniform in L.

Since $(\varphi_{il})_l$ is summable and (z_{it}) have uniformly bounded 4-th moments, it is clear that $\mathbf{x}_i^{(L)}/\sigma_i$ converges to $\mathbf{x}_i/\sigma_i$ in L^4 as $L\to\infty$, for each fixed T. By the dominated convergence theorem with (4.5.5) as an upper-bound, we can take the limit as $L\to\infty$ inside the expectation in (4.5.5) and the claim follows.

(b) follows from similar computations as in (a) and is omitted.

(c) Note that $E_\tau^\top E_\tau E^\top E$ has bounded operator norm under the event $\mathcal{B}_0$ defined in (4.2.16). Since $a \asymp \sigma_n^4\gamma_n(\tau)^2$ diverges as $T\to\infty$, the resolvent $R(a)$ is well-defined under $\mathcal{B}_0$ and $\|R(a)^k 1_{\mathcal{B}_0}\| = O(1)$. After conditioning on the σ-algebra $\mathcal{F}$ defined in (4.2.20), we can then apply (a) of the Lemma and get

$$\underline{\mathbb{E}}\Big|\mathbf{x}_{i,[1:T-\tau]}^\top R(a)^k\mathbf{x}_{j,[\tau+1:T]}1_{\mathcal{B}_0} - \underline{\mathbb{E}}\Big[\mathbf{x}_{i,[1:T-\tau]}^\top R(a)^k\mathbf{x}_{j,[\tau+1:T]}1_{\mathcal{B}_0}\Big]\Big|^q \lesssim T^{q/2}O(\sigma_i^q\sigma_j^q).$$

Taking expectations again to remove the conditioning, we obtain

$$\mathbb{E}\Big|\mathbf{x}_{i,[1:T-\tau]}^\top R(a)^k\mathbf{x}_{j,[\tau+1:T]}1_{\mathcal{B}_0} - \underline{\mathbb{E}}\Big[\mathbf{x}_{i,[1:T-\tau]}^\top R(a)^k\mathbf{x}_{j,[\tau+1:T]}1_{\mathcal{B}_0}\Big]\Big|^q \lesssim T^{q/2}O(\sigma_i^q\sigma_j^q).$$

Note that $\underline{\mathbb{E}}[\mathbf{x}_{i,[1:T-\tau]}^\top R(a)^k\mathbf{x}_{j,[\tau+1:T]}1_{\mathcal{B}_0}] = 0$ for all $i\neq j$ by (a) of the Lemma. So

$$\mathbf{x}_{i,[1,T-\tau]}^\top R(a)^k\mathbf{x}_{j,[\tau+1:T]}1_{\mathcal{B}_0} = 1_{i=j}\underline{\mathbb{E}}[\mathbf{x}_{i,[1,T-\tau]}^\top R(a)^k\mathbf{x}_{i,[\tau+1:T]}1_{\mathcal{B}_0}] + O_{L^q}(\sigma_i\sigma_j\sqrt{T}).$$

By Lemma 4.1 we have $1_{\mathcal{B}_0} = 1 - o_p(1)$, from which the last claim follows.

(d) follows from similar computations to the above and is omitted. □

Note that the expectations appearing in the previous lemma are conditional on the noise series $\boldsymbol{\epsilon}$. The following lemma gives a preliminary computation on the unconditional moments of certain quadratic forms. Recall matrices $B(a)$, $A(a)$ and $Q(a)$:

$$A(a) := \frac{1}{\sqrt{a}}X_0R(a)X_\tau^\top, \quad B(a) := \frac{1}{a}X_\tau E_0^\top E_0 R(a)X_\tau^\top,$$

$$Q(a) := I_K - a^{-1} X_0 R_a E_\tau^\top E_\tau X_0^\top.$$

Lemma 4.15. *Under the same setting as (c) of Lemma 4.14, we have*

$$\mathbb{E}[A(a)_{ij} 1_{\mathcal{B}_0}] = 1_{i=j}\bigg(\frac{\sigma_i^2 \gamma_i(\tau)}{a^{1/2}} + o(1)\bigg)$$
$$\mathrm{Var}(A(a)_{ij} 1_{\mathcal{B}_0}) = O\bigg(\frac{\sigma_i^2 \sigma_j^2}{aT}\bigg),$$
$$\mathbb{E}[B(a)_{ij} 1_{\mathcal{B}_0}] = 1_{i=j} o(1), \quad \mathbb{E}[Q(a)_{ij}^{-1} 1_{\mathcal{B}_2}] = 1_{i=j} + o(1).$$

Proof. Since $\mathbf{x}_i = \sigma_i \mathbf{f}_i + \boldsymbol{\epsilon}_i$, we first observe that

$$\frac{1}{\sqrt{a}T}\mathbb{E}[\mathbf{x}_{i,[1,T-\tau]}^\top \mathbf{x}_{j,[\tau+1:T]}] = \frac{1}{\sqrt{a}T}\mathbb{E}[\sigma_i^2 \mathbf{f}_{i,[1,T-\tau]}^\top \mathbf{f}_{j,[\tau+1:T]}] = 1_{i=j}\frac{\sigma_i^2 \gamma_i(\tau)}{\sqrt{a}}. \tag{4.5.6}$$

By definition, the event $\mathcal{B}_0$ is independent from the vector $\mathbf{x}$. Therefore

$$\begin{aligned}\mathbb{E}[A(a)_{ij} 1_{\mathcal{B}_0}] &= \frac{1}{\sqrt{a}T}\mathbb{E}[\mathbf{x}_{i,[1,T-\tau]}^\top \mathbf{x}_{j,[\tau+1:T]} 1_{\mathcal{B}_0}] + \frac{1}{\sqrt{a}T}\mathbb{E}[\mathbf{x}_{i,[1,T-\tau]}^\top (R(a) - I) 1_{\mathcal{B}_0} \mathbf{x}_{j,[\tau+1:T]}] \\ &= 1_{i=j}\frac{\sigma_i^2 \gamma_i(\tau)}{\sqrt{a}} \mathbb{P}(\mathcal{B}_0) + \frac{1}{\sqrt{a}T}\mathbb{E}[\mathbf{x}_{i,[1,T-\tau]}^\top (R(a) - I) 1_{\mathcal{B}_0} \mathbf{x}_{j,[\tau+1:T]}] \\ &= 1_{i=j}\bigg(\frac{\sigma_i^2 \gamma_i(\tau)}{\sqrt{a}} + o(1)\bigg) + \frac{1}{\sqrt{a}T}\mathbb{E}[\mathbf{x}_{i,[1,T-\tau]}^\top (R(a) - I) 1_{\mathcal{B}_0} \mathbf{x}_{j,[\tau+1:T]}],\end{aligned}$$

where the last equality follows since $\mathbb{P}(\mathcal{B}_0) = 1 + o(1)$ by Lemma 4.1. It remains to compute the last expectation above. Recall from (4.2.10) that the resolvent $R(a)$ satisfies $R(a) - I = a^{-1} E_\tau^\top E_\tau E^\top E R(a)$. By definition of $\mathcal{B}_0$ we have $\|E_\tau^\top E_\tau E^\top E 1_{\mathcal{B}_0}\| = O(1)$ and $\|R(a) 1_{\mathcal{B}_0}\| = O(1)$. Therefore

$$(R(a) - I) 1_{\mathcal{B}_0} = O_{\|\cdot\|}(a^{-1}). \tag{4.5.7}$$

Using (4.5.7) and (a) of Lemma 4.14 and taking iterated expectations we obtain

$$\begin{aligned}\frac{1}{\sqrt{a}T}\mathbb{E}[\mathbf{x}_{i,[1,T-\tau]}^\top (R(a) - I) 1_{\mathcal{B}_0} \mathbf{x}_{j,[\tau+1:T]}] &= \frac{1}{\sqrt{a}T}\mathbb{E}\Big[\underline{\mathbb{E}}[\mathbf{x}_{i,[1,T-\tau]}^\top (R(a) - I) 1_{\mathcal{B}_0} \mathbf{x}_{j,[\tau+1:T]}]\Big] \\ &= 1_{i=j}\frac{1}{\sqrt{a}T} O(\sigma_i^2 T) \mathbb{E}[\|R(a) - I\| 1_{\mathcal{B}_0}] = 1_{i=j} o(1).\end{aligned}$$

For the second moment, using $(a-b)^2 = (a-c)^2 + (c-b)^2 + 2(a-c)(c-b)$, we write

$$\begin{aligned}&(A(a)_{ij} 1_{\mathcal{B}_0} - \mathbb{E}[A(a)_{ij} 1_{\mathcal{B}_0}])^2 \\ &\quad = (A(a)_{ij} 1_{\mathcal{B}_0} - \underline{\mathbb{E}}[A(a)_{ij} 1_{\mathcal{B}_0}])^2 + (\underline{\mathbb{E}}[A(a)_{ij} 1_{\mathcal{B}_0}] - \mathbb{E}[A(a)_{ij} 1_{\mathcal{B}_0}])^2 \\ &\quad\quad + 2(A(a)_{ij} 1_{\mathcal{B}_0} - \underline{\mathbb{E}}[A(a)_{ij} 1_{\mathcal{B}_0}])(\underline{\mathbb{E}}[A(a)_{ij} 1_{\mathcal{B}_0}] - \mathbb{E}[A_{ij} 1_{\mathcal{B}_0}]).\end{aligned} \tag{4.5.8}$$

where by (c) of Lemma 4.14 we have

$$\mathbb{E}\Big[\big(A(a)_{ij}1_{\mathcal{B}_0} - \underline{\mathbb{E}}[A(a)_{ij}1_{\mathcal{B}_0}]\big)^2\Big] = \frac{1}{aT^2}O\big(\sigma_i^2\sigma_j^2T\big) = O\bigg(\frac{\sigma_i^2\sigma_j^2}{aT}\bigg),$$

and from Lemma 4.18 (whose proof does not depend on the current lemma) we recall

$$\mathbb{E}\Big[\big(\underline{\mathbb{E}}[A(a)_{ij}1_{\mathcal{B}_0}] - \mathbb{E}[A(a)_{ij}1_{\mathcal{B}_0}]\big)^2\Big] = O\Big(\frac{1}{aT}\Big).$$

Taking the expectation of (4.5.8) and using the Cauchy Schwarz inequality we have

$$\mathbb{E}[(A(a)_{ij}1_{\mathcal{B}_0} - \mathbb{E}[A(a)_{ij}1_{\mathcal{B}_0}])^2] = O\bigg(\frac{\sigma_i^2\sigma_j^2}{aT}\bigg).$$

The expectation of $B(a)$ can be computed based on the same ideas and is omitted.

Lastly, under the event $\mathcal{B}_2$, the matrix $Q(a)$ is invertible with $\|Q(a)1_{\mathcal{B}_2}\| = O(1)$. We recall from (4.2.15) that the inverse of $Q(a)$ satisfies

$$Q(a)^{-1} = I_K + \frac{1}{a}Q(a)^{-1}X_0R(a)E_\tau^\top E_\tau X_0^\top.$$

By definition of $\mathcal{B}_2$ we know $1_{\mathcal{B}_2}Q(a)^{-1}X_0R(a)E_\tau^\top E_\tau X_0^\top = O_{\|\cdot\|}(\sigma_1^2)$ and therefore

$$Q(a)^{-1}1_{\mathcal{B}_2} = 1_{\mathcal{B}_2}I_K + o_{\|\cdot\|}(1)$$

and the last claim follows after taking expectations. □

4.5.2 Estimates on resolvents

Define the following families of σ-algebras $(\mathcal{F}_i)_{i=1}^p$ and $(\underline{\mathcal{F}}_i)_{i=1}^K$ by

$$\mathcal{F}_i := \sigma\Big(\boldsymbol{\epsilon}_{[K+1,K+i],[1:T]}\Big), \quad \underline{\mathcal{F}}_i := \sigma\Big(\mathbf{x}_{[1:i],[1:T]}, \boldsymbol{\epsilon}_{[K+1,K+p],[1:T]}\Big),$$

i.e. $\mathcal{F}_i$ is the σ-algebra generated by the first i coordinates of the noise series $\boldsymbol{\epsilon}$ and $\underline{\mathcal{F}}_i$ is generated by all p coordinates of $\boldsymbol{\epsilon}$ plus the first i coordinates of the series $\mathbf{x}$.

Throughout the appendix we will write

$$\mathbb{E}_i[\,\cdot\,] := \mathbb{E}_i[\,\cdot\,|\mathcal{F}_i], \quad \underline{\mathbb{E}}_i[\,\cdot\,] := \mathbb{E}_i[\,\cdot\,|\underline{\mathcal{F}}_i]. \tag{4.5.9}$$

Note that by definition $\mathbb{E}_0[\,\cdot\,] = \mathbb{E}[\,\cdot\,]$ and $\mathbb{E}_p[\,\cdot\,] = \underline{\mathbb{E}}_0[\,\cdot\,] = \underline{\mathbb{E}}[\,\cdot\,]$.

We first develope a concentration inequality for normalized traces of the resolvent R.

Lemma 4.16. *For any matrix B with $T-\tau$ columns, we have*

(a) $$\frac{1}{T}\mathrm{tr}\Big(B(R1_{\mathcal{B}_0} - \mathbb{E}[R1_{\mathcal{B}_0}])\Big) = O_{L^2}\left(\frac{\|B\|}{\theta\sqrt{T}}\right),$$

(b) $$\frac{1}{T}\mathrm{tr}\Big(B(E_0^\top E_0 R1_{\mathcal{B}_0} - \mathbb{E}[E_0^\top E_0 R1_{\mathcal{B}_0}])\Big) = O_{L^2}\left(\frac{\|B\|}{\sqrt{T}}\right).$$

Proof. (a) Similar to Lemma 4.17 the proof is based on a martingale difference decomposition of $R1_{\mathcal{B}_0} - \mathbb{E}[R1_{\mathcal{B}_0}]$. We first setup the necessary notations and carry out some preliminary computations.

Recall that the k-th row of E_0 is equal to $T^{-1/2}\boldsymbol{\epsilon}_{K+k,[1:T-\tau]}^\top$. For brevity of notation we will adopt the following notation

$$\underline{\boldsymbol{\epsilon}}_{k0} := \boldsymbol{\epsilon}_{K+k,[1:T-\tau]}, \quad \underline{\boldsymbol{\epsilon}}_{k\tau} := \boldsymbol{\epsilon}_{K+k,[\tau+1:T]}. \tag{4.5.10}$$

Let E_{k0} and $E_{k\tau}$ be the matrices E_0 and E_τ with the k-th row replaced by zeros, i.e. $E_{k0} := E_0 - T^{-1/2}\mathbf{e}_k\underline{\boldsymbol{\epsilon}}_{k0}$ and $E_{k\tau} := E_0 - T^{-1/2}\mathbf{e}_k\underline{\boldsymbol{\epsilon}}_{k\tau}$. Define

$$\underline{R}_k := \left(I_T - \frac{1}{\theta}E_{k\tau}^\top E_{k\tau}E_0^\top E_0\right)^{-1}, \quad R_k := \left(I_T - \frac{1}{\theta}E_{k\tau}^\top E_{k\tau}E_{k0}^\top E_{k0}\right)^{-1},$$

where R_k is not to be confused with R_θ and R_λ defined previously. Then

$$E_0^\top E_0 - E_{k0}^\top E_{k0} = \frac{1}{T}\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top, \quad E_\tau^\top E_\tau - E_{k\tau}^\top E_{k\tau} = \frac{1}{T}\underline{\boldsymbol{\epsilon}}_{k\tau}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top,$$

from which we can compute

$$\begin{aligned}
R^{-1} - \underline{R}_k^{-1} &= -\frac{1}{\theta}(E_\tau^\top E_\tau - E_{k\tau}^\top E_{k\tau})E_0^\top E_0 = -\frac{1}{\theta T}\underline{\boldsymbol{\epsilon}}_{k\tau}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0 \\
\underline{R}_k^{-1} - R_k^{-1} &= -\frac{1}{\theta}E_{k\tau}^\top E_{k\tau}(E_0^\top E_0 - E_{k0}^\top E_{k0}) = -\frac{1}{\theta T}E_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top.
\end{aligned}$$

We furthermore define scalars

$$\begin{aligned}
\underline{\beta}_k &= \frac{1}{1 + \mathrm{tr}(\underline{R}_k(R^{-1} - \underline{R}_k^{-1}))} = \frac{1}{1 - \frac{1}{\theta T}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0 \underline{R}_k \underline{\boldsymbol{\epsilon}}_{k\tau}}, \\
\beta_k &= \frac{1}{1 + \mathrm{tr}(R_k(\underline{R}_k^{-1} - R_k^{-1}))} = \frac{1}{1 - \frac{1}{\theta T}\underline{\boldsymbol{\epsilon}}_{k0}^\top R_k E_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0}},
\end{aligned}$$

both of which are clearly of order $1 + o(1)$ under the event $\mathcal{B}_0$. Using (4.5.1) we get

$$R - \underline{R}_k = -\underline{\beta}_k\underline{R}_k(R^{-1} - \underline{R}_k^{-1})\underline{R}_k = \frac{\underline{\beta}_k}{\theta T}\underline{R}_k\underline{\boldsymbol{\epsilon}}_{k\tau}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0\underline{R}_k, \tag{4.5.11a}$$

$$\underline{R}_k - R_k = -\beta_k R_k(\underline{R}_k^{-1} - R_k^{-1})R_k = \frac{\beta_k}{\theta T}R_k E_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top R_k. \tag{4.5.11b}$$

Substituting (4.5.11b) back into (4.5.11a) we get

$$R - \underline{R}_k = \frac{\underline{\beta}_k}{\theta T}\Big(R_k + \frac{\beta_k}{\theta T} R_k E_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top R_k\Big)\underline{\boldsymbol{\epsilon}}_{k\tau}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0\Big(R_k + \frac{\beta_k}{\theta T} R_k E_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top R_k\Big),$$

and so we have

$$R - R_k = (\underline{R}_k - R_k) + (R - \underline{R}_k) =: U_1 + U_2 + U_3 + U_4 + U_5, \tag{4.5.12}$$

where we have defined

$$\begin{aligned} U_1 &:= \frac{\beta_k}{\theta T} R_k E_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top R_k, \quad U_2 := \frac{\underline{\beta}_k}{\theta T} R_k \underline{\boldsymbol{\epsilon}}_{k\tau}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0 R_k, \\ U_3 &:= \frac{\underline{\beta}_k \beta_k}{\theta^2 T^2} R_k \underline{\boldsymbol{\epsilon}}_{k\tau}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0 R_k E_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top R_k, \\ U_4 &:= \frac{\underline{\beta}_k \beta_k}{\theta^2 T^2} R_k E_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top R_k \underline{\boldsymbol{\epsilon}}_{k\tau}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0 R_k, \\ U_5 &:= \frac{\underline{\beta}_k \beta_k^2}{\theta^3 T^3} R_k E_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top R_k \underline{\boldsymbol{\epsilon}}_{k\tau}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0 R_k E_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top R_k. \end{aligned} \tag{4.5.13}$$

Recall the event $\mathcal{B}_0$ from (4.1). Define

$$\mathcal{B}_0^k := \Big\{ \|E_{k0}^\top E_{k0}\| + \|E_{k\tau}^\top E_{k\tau}\| \le 4\Big(1 + \frac{p}{T}\Big)\Big\}, \quad k = 1, \dots, p. \tag{4.5.14}$$

Clearly $\|E_{k0}^\top E_{k0}\| \le \|E_0^\top E_0\|$ which implies $\mathcal{B}_0 \subseteq \mathcal{B}_0^k$ and so $1_{\mathcal{B}_0} \le 1_{\mathcal{B}_0^k}$. Recall the family of conditional expectations $\mathbb{E}_i[\,\cdot\,]$ defined in (4.5.9). Then

$$\begin{aligned} \frac{1}{T}\mathrm{tr}(B(R1_{\mathcal{B}_0} - \mathbb{E}[R1_{\mathcal{B}_0}])) &= \frac{1}{T}\sum_{k=1}^p (\mathbb{E}_k - \mathbb{E}_{k-1})\mathrm{tr}(BR1_{\mathcal{B}_0}) \\ &= \frac{1}{T}\sum_{k=1}^p (\mathbb{E}_k - \mathbb{E}_{k-1})\Big(\mathrm{tr}(BR1_{\mathcal{B}_0}) - \mathrm{tr}(BR_k 1_{\mathcal{B}_0^k})\Big) \\ &= \frac{1}{T}\sum_{k=1}^p (\mathbb{E}_k - \mathbb{E}_{k-1})\mathrm{tr}(B(R - R_k)1_{\mathcal{B}_0}) - \frac{1}{T}\sum_{k=1}^p (\mathbb{E}_k - \mathbb{E}_{k-1})\mathrm{tr}(BR_k(1_{\mathcal{B}_0^k} - 1_{\mathcal{B}_0})) \\ &=: I_1 + I_2, \end{aligned} \tag{4.5.15}$$

where the second equality holds since $\mathbb{E}_k[\mathrm{tr}(BR_k 1_{\mathcal{B}_0^k})] = \mathbb{E}_{k-1}[\mathrm{tr}(BR_k 1_{\mathcal{B}_0^k})]$ and the third equality is purely algebraic computations. We first deal with the second term in (4.5.15). Using $\mathrm{tr}(BR_k) \le p\|BR_k\|$ and $\|BR_k 1_{\mathcal{B}_0^k}\| = O(\|B\|)$ we have

$$\begin{aligned} \mathbb{E}|I_2|^2 &= \frac{1}{T^2}\sum_{k=1}^p \mathbb{E}\Big|(\mathbb{E}_k - \mathbb{E}_{k-1})\mathrm{tr}(BR_k(1_{\mathcal{B}_0^k} - 1_{\mathcal{B}_0}))\Big|^2 \le \frac{4p^2}{T^2}\sum_{k=1}^p \mathbb{E}\Big|\|BR_k\|(1_{\mathcal{B}_0^k} - 1_{\mathcal{B}_0})\Big|^2 \\ &= O\Big(\frac{p^2}{T^2}\|B\|^2\Big)\sum_{k=1}^p \mathbb{E}\Big|1_{\mathcal{B}_0^k} - 1_{\mathcal{B}_0}\Big|^2 = O\Big(\frac{p^2}{T^2}\|B\|^2\Big)\sum_{k=1}^p \mathbb{P}(\mathcal{B}_0^c) = o(T^{-l}\|B\|^2), \end{aligned}$$

for any $l \in \mathbb{N}$ by Lemma 4.1. For the first term in (4.5.15), since I_1 is a sum of a martingale difference sequence, using (4.5.12) and $\mathcal{B}_0 \subseteq \mathcal{B}_0^k$ we have

$$\mathbb{E}|I_1|^2 \leq \frac{1}{T^2}\sum_{k=1}^{p}\mathbb{E}\Big|(\mathbb{E}_k-\mathbb{E}_{k-1})\mathrm{tr}(B(R-R_k)1_{\mathcal{B}_0^k})\Big|^2$$
$$\leq \frac{4}{T^2}\sum_{k=1}^{p}\mathbb{E}\Big|\mathrm{tr}(B(R-R_k)1_{\mathcal{B}_0^k})\Big|^2 \leq \frac{20}{T^2}\sum_{k=1}^{p}\sum_{n=1}^{5}\mathbb{E}\Big|\mathrm{tr}(BU_n1_{\mathcal{B}_0^k})\Big|^2,$$

and it remains to bound the second moment of each $\mathrm{tr}(BU_n1_{\mathcal{B}_0^k})$. Since $\{\epsilon_{it}\}$ are assumed to be i.i.d. standard Gaussian, we have the following moment estimate

$$\mathbb{E}[\|\underline{\boldsymbol{\epsilon}}_{k0}\|^n] = \mathbb{E}\left[\Big(\sum_{t=1}^{T-\tau}\epsilon_{kt}^2\Big)^{n/2}\right] \lesssim (T-\tau)^{n/2-1}\sum_{t=1}^{T-\tau}\mathbb{E}|\epsilon_{kt}|^n = O(T^{n/2}). \tag{4.5.16}$$

Using $\beta_k 1_{\mathcal{B}_0^k} = 1+o(1)$ and the trivial inequality $x^\top Ax \leq \|x\|^2\|A\|$ we obtain

$$\mathbb{E}\Big|\mathrm{tr}(BU_11_{\mathcal{B}_0^k})\Big|^2 \lesssim \frac{1}{\theta^2T^2}\mathbb{E}\Big[(\underline{\boldsymbol{\epsilon}}_{k0}^\top R_kBR_kE_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0})^2 1_{\mathcal{B}_0^k}\Big] \tag{4.5.17a}$$
$$\leq \frac{1}{\theta^2T^2}\mathbb{E}\Big[\|\underline{\boldsymbol{\epsilon}}_{k0}\|^4\|R_k\|^4\|E_{k\tau}^\top E_{k\tau}\|^2 1_{\mathcal{B}_0^k}\Big]\|B\|^2 \lesssim \frac{1}{\theta^2}\|B\|^2.$$

The second term U_2 can be dealt with in exactly the same way to obtain

$$\mathbb{E}\Big|\mathrm{tr}(BU_21_{\mathcal{B}_0^k})\Big|^2 \lesssim \frac{1}{\theta^2}\|B\|^2, \tag{4.5.17b}$$

and we omit the details. For U_3, similar computations gives

$$\mathbb{E}\Big|\mathrm{tr}(BU_31_{\mathcal{B}_0^k})\Big|^2 \lesssim \frac{1}{\theta^4T^4}\mathbb{E}\Big[(\underline{\boldsymbol{\epsilon}}_{k0}^\top R_kBR_k\underline{\boldsymbol{\epsilon}}_{k\tau}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0R_kE_{k\tau}^\top E_{k\tau}\underline{\boldsymbol{\epsilon}}_{k0})^2 1_{\mathcal{B}_0^k}\Big]$$
$$\leq \frac{1}{\theta^4T^4}\mathbb{E}\Big[\|\underline{\boldsymbol{\epsilon}}_{k0}\|^4\|\underline{\boldsymbol{\epsilon}}_{k\tau}\|^4\|R_k\|^4\|E_0^\top E_0R_kE_{k\tau}^\top E_{k\tau}\|^2 1_{\mathcal{B}_0^k}\Big]\|B\|^2,$$

since $x^\top Ay \leq \|x\|\|y\|\|A\|$. Therefore

$$\mathbb{E}\Big|\mathrm{tr}(BU_3)\Big|^2 \lesssim \frac{1}{\theta^4T^4}\mathbb{E}\Big[\|\underline{\boldsymbol{\epsilon}}_{k0}\|^8\Big]^{1/2}\mathbb{E}\Big[\|\underline{\boldsymbol{\epsilon}}_{k\tau}\|^8\Big]^{1/2} \lesssim \frac{1}{\theta^4}\|B\|^2. \tag{4.5.17c}$$

Once again U_4 can be bounded in the same way to obtain

$$\mathbb{E}\Big|\mathrm{tr}(BU_4)1_{\mathcal{B}_0^k}\Big|^2 \leq \frac{1}{\theta^4}\|B\|^2. \tag{4.5.17d}$$

With the same approach but more laborious computations we can obtain

$$\mathbb{E}\Big|\mathrm{tr}(BU_5)1_{\mathcal{B}_0^k}\Big|^2 \lesssim \frac{1}{\theta^6}\|B\|^2. \tag{4.5.17e}$$

Note that the estimates (4.5.17a)-(4.5.17e) are uniform in $k=1,\dots,p$. We then conclude

$$\mathbb{E}\Big|\frac{1}{T}\mathrm{tr}(B(R1_{\mathcal{B}_0^k}-\mathbb{E}[R1_{\mathcal{B}_0^k}]))\Big|^2 = O\Big(\frac{p}{T^2\theta^2}\Big)\|B\|^2 = O\Big(\frac{1}{T\theta^2}\Big)\|B\|^2,$$

and the conclusion follows.

(b) Similar to (a), via a martingale difference decomposition we obtain

$$\mathbb{E}\Big|\frac{1}{T}\mathrm{tr}(B(E_0^\top E_0 R 1_{\mathcal{B}_0^k} - \mathbb{E}[E_0^\top E_0 R 1_{\mathcal{B}_0^k}]))\Big|^2 \lesssim \frac{1}{T^2}\sum_{k=1}^{p}\mathbb{E}\Big|\mathrm{tr}(B(E_0^\top E_0 R - E_{k0}^\top E_{k0} R_k))1_{\mathcal{B}_0^k}\Big|^2,$$

where, recalling the U_n's defined in the proof of (a), we have

$$\begin{aligned} E_0^\top E_0 R - E_{k0}^\top E_{k0} R_k &= \frac{1}{T}\boldsymbol{\epsilon}_{k0}\boldsymbol{\epsilon}_{k0}^\top R_k + \frac{1}{T}\boldsymbol{\epsilon}_{k0}\boldsymbol{\epsilon}_{k0}^\top (R - R_k) + E_{k0}^\top E_{k0}(R - R_k) \\ &= \frac{1}{T}\boldsymbol{\epsilon}_{k0}\boldsymbol{\epsilon}_{k0}^\top R_k + \frac{1}{T}\sum_{n=1}^{5}\boldsymbol{\epsilon}_{k0}\boldsymbol{\epsilon}_{k0}^\top U_n + \sum_{n=1}^{5} E_{k0}^\top E_{k0} U_n. \end{aligned} \tag{4.5.18}$$

We deal with the first two term in (4.5.18) to illustrate the ideas of the proof, the other terms can be dealth with similarly. Using (4.5.16) and $p \asymp T$, clearly we have

$$\frac{1}{T^2}\sum_{k=1}^{p}\mathbb{E}\Big|\frac{1}{T}\mathrm{tr}(B\boldsymbol{\epsilon}_{k0}\boldsymbol{\epsilon}_{k0}^\top R_k)1_{\mathcal{B}_0^k}\Big|^2 \lesssim \frac{1}{T^4}\sum_{k=1}^{p} T^2\mathbb{E}[\|BR_k 1_{\mathcal{B}_0^k}\|^2] = O\Big(\frac{1}{T}\Big)\|B\|^2. \tag{4.5.19}$$

Similar to the computations in (4.5.17a), we can get

$$\begin{aligned} \mathbb{E}\Big|\frac{1}{T}\mathrm{tr}(B(\boldsymbol{\epsilon}_{k0}\boldsymbol{\epsilon}_{k0}^\top U_1))1_{\mathcal{B}_0^k}\Big|^2 &\lesssim \frac{1}{\theta^2 T^2}\frac{1}{T^2}\mathbb{E}[(\boldsymbol{\epsilon}_{k0}^\top R_k B\boldsymbol{\epsilon}_{k0}\boldsymbol{\epsilon}_{k0}^\top R_k E_{k\tau}^\top E_{k\tau}\boldsymbol{\epsilon}_{k0})^2 1_{\mathcal{B}_0^k}] \\ &\le \frac{1}{\theta^2 T^4}\mathbb{E}\Big[\|\boldsymbol{\epsilon}_{k0}\|^8\|R_k\|^4\|E_{k\tau}^\top E_{k\tau}\|^2 1_{\mathcal{B}_0^k}\Big]\|B\|^2 \lesssim \frac{1}{\theta^2}\|B\|^2, \end{aligned}$$

which immediately gives

$$\frac{1}{T^2}\sum_{k=1}^{p}\mathbb{E}\Big|\frac{1}{T}\mathrm{tr}(B(\boldsymbol{\epsilon}_{k0}\boldsymbol{\epsilon}_{k0}^\top U_1))1_{\mathcal{B}_0^k}\Big|^2 = O\Big(\frac{p}{\theta^2 T^2}\Big)\|B\|^2 = O\Big(\frac{1}{\theta^2 T}\Big)\|B\|^2.$$

Note that this term is negligible in comparison to (4.5.19). Using the same ideas, it is routine to check that the other 9 terms in (4.5.18) are negligible as well, and we omit the details. The bound therefore follows from (4.5.19). □

Next recall that $Q = I_K - \frac{1}{\theta}X_0 R E_\tau^\top E_\tau X_0^\top$. We now state a concentration inequality for entries of the matrix Q^{-1}, under the event $\mathcal{B}_2$.

Lemma 4.17. *Write $Q_{ij}^{-1} := (Q^{-1})_{ij}$. Then*

a) For all $k = 1, \ldots, K$, we have

$$Q_{kk}^{-1}1_{\mathcal{B}_2} - \mathbb{E}[Q_{kk}^{-1}1_{\mathcal{B}_2}] = o_{L^1}\Big(\frac{1}{\sqrt{T}}\Big).$$

b) The off-diagonal elements of Q^{-1} satisfy

$$Q_{ij}^{-1} 1_{\mathcal{B}_2} = O_{L^2}\left(\frac{1}{\gamma_1(\tau)^2\sigma_1^2\sqrt{T}}\right)$$

uniformly in $i, j = 1, \ldots, K$, $i \neq j$.

Proof. (a) Recalling the event $\mathcal{B}_2$, we note that the matrix Q is invertible with probability tending to 1. The proof relies on expressing $Q_{kk}^{-1} 1_{\mathcal{B}_2} - \mathbb{E}[Q_{kk}^{-1} 1_{\mathcal{B}_2}]$ as a sum of martingale differences. We first setup the notations necessary.

Let $T^{-1/2}\mathbf{x}_i := T^{-1/2}\mathbf{x}_{i,[1:T-\tau]}$ be the (column vector) of the i-th row of X_0, i.e. we can write $X_0 = T^{-1/2}\sum_{i=1}^K \mathbf{e}_i \mathbf{x}_i^\top$. Define $X_{i0} := X_0 - \frac{1}{\sqrt{T}}\mathbf{e}_i\mathbf{x}_i^\top$, and

$$Q_{(i)} := I_K - \frac{1}{\theta} X_{i0} R E_\tau^\top E_\tau X_0^\top, \quad Q_{(ii)} := I_K - \frac{1}{\theta} X_{i0} R E_\tau^\top E_\tau X_{i0}^\top,$$

from which we can immediately compute

$$Q - Q_{(i)} = -\frac{1}{\theta\sqrt{T}} \mathbf{e}_i \mathbf{x}_i^\top R E_\tau^\top E_\tau X_0^\top, \quad Q_{(i)} - Q_{(ii)} = -\frac{1}{\theta\sqrt{T}} X_{i0} R E_\tau^\top E_\tau \mathbf{x}_i \mathbf{e}_i^\top.$$

Note that all elements on the i-th row of $Q_{(i)}$ are equal to zero except for the diagonal which is equal to 1, i.e. $Q_{(i)}$ is equal to the identity when restricted to the i-th coordinate. Then the inverse $Q_{(i)}^{-1}$, whenever it exists, must also equal to the identity when restricted to the i-th coordinate. A similar observation can be made for the matrix $Q_{(ii)}$ and it is not hard to observe that

$$\begin{aligned} \mathbf{e}_i^\top (Q_{(ii)})^{-1}\mathbf{e}_i = 1, \quad \mathbf{e}_i^\top (Q_{(i)})^{-1}\mathbf{e}_j = 0, \quad \forall j \neq i, \\ \mathbf{e}_i^\top (Q_{(ii)})^{-1}\mathbf{e}_j = \mathbf{e}_j^\top (Q_{(ii)})^{-1}\mathbf{e}_i = 0, \quad \forall j \neq i. \end{aligned} \tag{4.5.20}$$

To compute the difference $Q^{-1} - Q_{(i)}^{-1}$, which will turn out to be the central focus of the proof, we first define the following scalars

$$\begin{aligned} b_i : &= \frac{1}{1 + \operatorname{tr}(Q_{(i)}^{-1}(Q - Q_{(i)}))} = \frac{1}{1 - \frac{1}{\theta\sqrt{T}}\mathbf{x}_i^\top R E_\tau^\top E_\tau X_0^\top Q_{(i)}^{-1}\mathbf{e}_i} \\ b_{ii} : &= \frac{1}{1 + \operatorname{tr}(Q_{(ii)}^{-1}(Q_{(i)} - Q_{(ii)}))} = \frac{1}{1 - \frac{1}{\theta\sqrt{T}}\mathbf{e}_i^\top Q_{(ii)}^{-1} X_{i0} R E_\tau^\top E_\tau \mathbf{x}_i} = 1, \end{aligned} \tag{4.5.21}$$

where the last equality holds by (4.5.20). Then using the identity (4.5.1) we have

$$Q^{-1} - Q_{(i)}^{-1} = \frac{b_i}{\theta\sqrt{T}} Q_{(i)}^{-1}\mathbf{e}_i\mathbf{x}_i^\top R E_\tau^\top E_\tau X_0^\top Q_{(i)}^{-1}, \tag{4.5.22a}$$

$$Q_{(i)}^{-1} - Q_{(ii)}^{-1} = \frac{1}{\theta\sqrt{T}} Q_{(ii)}^{-1} X_{i0} R E_\tau^\top E_\tau \mathbf{x}_i \mathbf{e}_i^\top Q_{(ii)}^{-1}. \tag{4.5.22b}$$

We observe that the matrices $Q_{(i)}^{-1}$ and $Q_{(ii)}^{-1}$ differ only on off-diagonal elements on the i-th column. Indeed, from (4.5.20) and (4.5.22b), if $n \neq i$ or if $n = m = i$ then

$$\mathbf{e}_m^\top(Q_{(i)}^{-1} - Q_{(ii)}^{-1})\mathbf{e}_n = \frac{1}{\theta\sqrt{T}}\mathbf{e}_m^\top Q_{(ii)}^{-1} X_{i0} R E_\tau^\top E_\tau \mathbf{x}_i \mathbf{e}_i^\top Q_{(ii)}^{-1}\mathbf{e}_n = 0. \tag{4.5.23}$$

Then, substituting (4.5.22b) back into (4.5.22a) we obtain

$$\begin{aligned}
\mathbf{e}_k^\top(Q^{-1} - Q_{(i)}^{-1})\mathbf{e}_k &= \frac{b_i}{\theta\sqrt{T}}\mathbf{e}_k^\top Q_{(i)}^{-1}\mathbf{e}_i\mathbf{x}_i^\top R E_\tau^\top E_\tau X_0^\top Q_{(i)}^{-1}\mathbf{e}_k \\
&= \frac{b_i}{\theta\sqrt{T}}\mathbf{e}_k^\top Q_{(i)}^{-1}\mathbf{e}_i\mathbf{x}_i^\top R E_\tau^\top E_\tau X_0^\top Q_{(ii)}^{-1}\mathbf{e}_k \\
&\qquad + \frac{b_i}{\theta^2 T}\mathbf{e}_k^\top Q_{(i)}^{-1}\mathbf{e}_i\mathbf{x}_i^\top R E_\tau^\top E_\tau X_0^\top Q_{(ii)}^{-1} X_{i0} R E_\tau^\top E_\tau \mathbf{x}_i \mathbf{e}_i^\top Q_{(ii)}^{-1}\mathbf{e}_k \\
&= \frac{b_i}{\theta\sqrt{T}}\mathbf{e}_k^\top Q_{(ii)}^{-1}\mathbf{e}_i\mathbf{x}_i^\top R E_\tau^\top E_\tau X_0^\top Q_{(ii)}^{-1}\mathbf{e}_k \\
&\qquad + \frac{b_i}{\theta^2 T}\mathbf{e}_k^\top Q_{(i)}^{-1}\mathbf{e}_i\mathbf{x}_i^\top R E_\tau^\top E_\tau X_0^\top Q_{(ii)}^{-1} X_{i0} R E_\tau^\top E_\tau \mathbf{x}_i \mathbf{e}_i^\top Q_{(ii)}^{-1}\mathbf{e}_k \\
&\qquad + \frac{b_i}{\theta^2 T}\mathbf{e}_k^\top Q_{(ii)}^{-1} X_{i0} R E_\tau^\top E_\tau \mathbf{x}_i \mathbf{e}_i^\top Q_{(ii)}^{-1}\mathbf{e}_i\mathbf{x}_i^\top R E_\tau^\top E_\tau X_0^\top Q_{(ii)}^{-1}\mathbf{e}_k \\
&=: I_1 + I_2 + I_3.
\end{aligned} \tag{4.5.24}$$

To simplify this expression further, define the following quadratic forms

$$\begin{aligned}
\xi_i &:= \frac{1}{\theta T}\mathbf{x}_i^\top R E_\tau^\top E_\tau \mathbf{x}_i, \quad \eta_i := \frac{1}{\theta^2 T}\mathbf{x}_i^\top R E_\tau^\top E_\tau X_{i0}^\top Q_{(ii)}^{-1} X_{i0} R E_\tau^\top E_\tau \mathbf{x}_i, \\
\zeta_{ik} &:= \frac{1}{\theta^2 T}\mathbf{x}_i^\top R E_\tau^\top E_\tau X_{i0}^\top Q_{(ii)}^{-1}\mathbf{e}_k\mathbf{e}_k^\top Q_{(ii)}^{-1} X_{i0} R E_\tau^\top E_\tau \mathbf{x}_i,
\end{aligned} \tag{4.5.25}$$

then using (4.5.20), we can easily write I_1, I_2 and I_3 into

$$\begin{aligned}
I_1 &= 1_{i=k}\frac{b_k}{\theta\sqrt{T}}\mathbf{x}_k^\top R E_\tau^\top E_\tau X_0^\top Q_{(kk)}^{-1}\mathbf{e}_k = 1_{i=k} b_k \xi_k, \qquad (4.5.26)\\
I_2 &= 1_{i=k}\frac{b_k}{\theta^2 T}\mathbf{x}_k^\top R E_\tau^\top E_\tau X_0^\top Q_{(kk)}^{-1} X_{k0} R E_\tau^\top E_\tau \mathbf{x}_k = 1_{i=k} b_k \eta_k, \\
I_3 &= 1_{i\neq k}\frac{b_i}{\theta^2 T}\mathbf{e}_k^\top Q_{(ii)}^{-1} X_{i0} R E_\tau^\top E_\tau \mathbf{x}_i \mathbf{x}_i^\top R E_\tau^\top E_\tau X_0^\top Q_{(ii)}^{-1}\mathbf{e}_k = 1_{i\neq k} b_i \zeta_{ik}.
\end{aligned}$$

We first state some estimates on ξ and η under appropriate events. Recall from (4.2.16) the event $\mathcal{B}_1 := \left\{\|X_0^\top X_0\| \le 2\sum_{i=1}^K \sigma_1^2\right\}$. Define the event

$$\mathcal{B}_1^i := \left\{\|X_{i0}^\top X_{i0}\| \le 2\sum_{i=1}^K \sigma_1^2\right\}, \quad i = 1, \dots, K \tag{4.5.27}$$

and write $\mathcal{B}_2^i := \mathcal{B}_0 \cap \mathcal{B}_1^i$. Then clearly $\mathcal{B}_2^i \subseteq \mathcal{B}_2$. Define

$$\overline{\xi}_i := \frac{1}{\theta T}\mathrm{tr}\Big(\Psi_0^{ii}(R E_\tau^\top E_\tau)\Big), \quad \overline{\eta}_i := \frac{1}{\theta^2 T}\mathrm{tr}\Big(\Psi_0^{ii}(R E_\tau^\top E_\tau X_{i0}^\top Q_{(ii)}^{-1} X_{i0} R E_\tau^\top E_\tau)\Big),$$

where Ψ_0^{ii} is defined in (4.5.4). Write

$$\underline{\xi}_i := \xi_i - \overline{\xi}_i, \quad \underline{\eta}_i := \eta_i - \overline{\eta}_i. \tag{4.5.28}$$

Using Lemma 4.14 and taking iterated expectations we have

$$\mathbb{E}\Big[\underline{\xi}_i^2 1_{\mathcal{B}_2^i}\Big] = \mathbb{E}\Big[\underline{\mathbb{E}}[\underline{\xi}_i^2 1_{\mathcal{B}_2^i}]\Big] = \frac{1}{\theta^2 T^2} O(\sigma_i^4 T)\mathbb{E}\|RE_\tau^\top E 1_{\mathcal{B}_2^i}\|^2 = O\bigg(\frac{\sigma_i^4}{\theta^2 T}\bigg),$$
$$\mathbb{E}\Big[\underline{\eta}_i^2 1_{\mathcal{B}_2^i}\Big] = \frac{1}{\theta^4 T^2} O(\sigma_i^4 T)\mathbb{E}\|RE_\tau^\top E_\tau X_{i0}^\top Q_{(ii)}^{-1} X_{i0} RE_\tau^\top E_\tau 1_{\mathcal{B}_2^i}\|^2 = O\bigg(\frac{\sigma_i^4 \sum_{j=1}^K \sigma_j^4}{\theta^4 T}\bigg).$$

Using $\theta = \theta_k \asymp \sigma_k^4 \gamma_k(\tau)^2$ from Proposition 4.4, Lemma 4.12 to deal with Ψ_0^{ii} and Assumptions 4.1 to compare the different speeds, we conclude

$$\mathbb{E}\Big[\underline{\xi}_i^2 1_{\mathcal{B}_2^i}\Big] = O\bigg(\frac{1}{\gamma_1(\tau)^2\theta T}\bigg), \quad \mathbb{E}\Big[\underline{\eta}_i^2 1_{\mathcal{B}_2^i}\Big] = O\bigg(\frac{K}{\gamma_1(\tau)^4\theta^2 T}\bigg), \tag{4.5.29}$$
$$\overline{\xi}_i 1_{\mathcal{B}_2^i} = O(\sigma_i^2\theta^{-1}), \quad \overline{\eta}_i 1_{\mathcal{B}_2^i} = O(K\sigma_i^4\theta^{-2}).$$

We then consider the scalar b_i defined in (4.5.21). From (4.5.20) and (4.5.22b) we observe

$$\frac{1}{\theta\sqrt{T}}\mathbf{x}_i^\top RE_\tau^\top E_\tau X_0^\top Q_{(i)}^{-1}\mathbf{e}_i = \frac{1}{\theta\sqrt{T}}\mathbf{x}_i^\top RE_\tau^\top E_\tau X_0^\top Q_{(ii)}^{-1}\mathbf{e}_i$$
$$+ \frac{1}{\theta^2 T}\mathbf{x}_i^\top RE_\tau^\top E_\tau X_0^\top Q_{(ii)}^{-1} X_{i0} RE_\tau^\top E_\tau \mathbf{x}_i \mathbf{e}_i^\top Q_{(ii)}^{-1}\mathbf{e}_i = \xi_i + \eta_i.$$

Substituting back into (4.5.21) we can simplify to obtain

$$b_i = (1 - \xi_i - \eta_i)^{-1}. \tag{4.5.30}$$

Define $\overline{b}_i = (1 - \overline{\xi}_i - \overline{\eta}_i)^{-1}$ so that subtracting the two we get

$$b_i = (1 - \xi_i - \eta_i)^{-1} = \overline{b}_i - b_i\overline{b}_i(\underline{\xi}_i + \underline{\eta}_i). \tag{4.5.31}$$

Finally, from the expression (4.5.30) and the bounds (4.5.29) we clearly have

$$b_i 1_{\mathcal{B}_2} = 1 + o(1), \quad \overline{b}_i 1_{\mathcal{B}_2^i} = 1 + o(1). \tag{4.5.32}$$

We can now carry out the main idea of the proof. Recall notations $\underline{\mathbb{E}}[\,\cdot\,]$ and $\underline{\mathbb{E}}_i[\,\cdot\,]$ from (4.5.9). By definition of $Q_{(ii)}$ and $\mathcal{B}_2^i$ we have

$$\mathbf{e}_k^\top\Big(Q^{-1}1_{\mathcal{B}_2} - \underline{\mathbb{E}}[Q^{-1}1_{\mathcal{B}_2}]\Big)\mathbf{e}_k = \sum_{i=1}^K(\underline{\mathbb{E}}_i - \underline{\mathbb{E}}_{i-1})\Big(\mathbf{e}_k^\top Q^{-1}1_{\mathcal{B}_2}\mathbf{e}_k - \mathbf{e}_k^\top Q_{(ii)}^{-1}1_{\mathcal{B}_2^i}\mathbf{e}_k\Big)$$
$$= \sum_{i=1}^K(\underline{\mathbb{E}}_i - \underline{\mathbb{E}}_{i-1})\Big(\mathbf{e}_k^\top Q^{-1}1_{\mathcal{B}_2}\mathbf{e}_k - \mathbf{e}_k^\top Q_{(i)}^{-1}1_{\mathcal{B}_2^i}\mathbf{e}_k\Big),$$

where the last equality follows from (4.5.23). Similar to how we dealt with the second term in (4.5.15) in the proof of Lemma 4.16, using Lemma 4.1 we may obtain

$$\begin{aligned}\mathbf{e}_k^\top\Big(Q^{-1}1_{\mathcal{B}_2}-\mathbb{E}[Q^{-1}1_{\mathcal{B}_2}]\Big)\mathbf{e}_k&=\sum_{i=1}^K(\mathbb{E}_i-\mathbb{E}_{i-1})\mathbf{e}_k\Big(Q^{-1}-Q_{(ii)}^{-1}\Big)1_{\mathcal{B}_2}\mathbf{e}_k+o_{L^2}\left(\frac{1}{\sqrt{T}}\right)\\&=\sum_{i=1}^K(\mathbb{E}_i-\mathbb{E}_{i-1})(I_1+I_2+I_3)1_{\mathcal{B}_2}+o_{L^2}\left(\frac{1}{\sqrt{T}}\right),\end{aligned}\tag{4.5.33}$$

where the second equality holds by (4.5.23).

As will be shown, the term involving I_1 is the leading term of (4.5.33), this is what we consider now. Using the identity (4.5.26) we simply have

$$\sum_{i=1}^K(\mathbb{E}_i-\mathbb{E}_{i-1})I_1 1_{\mathcal{B}_2^i}=(\mathbb{E}_k-\mathbb{E}_{k-1})b_k\xi_k 1_{\mathcal{B}_2},$$

which, recalling (4.5.28) and using (4.5.31), can be written into

$$\begin{aligned}(\mathbb{E}_k-\mathbb{E}_{k-1})b_k\xi_k1_{\mathcal{B}_2}&=(\mathbb{E}_k-\mathbb{E}_{k-1})\Big(\bar{b}_k-b_k\bar{b}_k(\underline{\xi}_k+\underline{\eta}_k)\Big)\Big(\bar{\xi}_k+\underline{\xi}_k\Big)1_{\mathcal{B}_2}\\&=(\mathbb{E}_k-\mathbb{E}_{k-1})\Big[\bar{b}_k\bar{\xi}_k+\bar{b}_k\underline{\xi}_k-b_k\bar{b}_k(\underline{\xi}_k+\underline{\eta}_k)(\bar{\xi}_k+\underline{\xi}_k)\Big]1_{\mathcal{B}_2}.\end{aligned}\tag{4.5.34}$$

We consider the three terms in the square bracket in (4.5.34) separately. For the first term, we note that $(\mathbb{E}_k-\mathbb{E}_{k-1})\bar{b}_k\bar{\xi}_k1_{\mathcal{B}_2^k}=0$ by definition of $\bar{b}_k\bar{\xi}_k$ and $\mathcal{B}_2^k$. Using this, we have

$$(\mathbb{E}_k-\mathbb{E}_{k-1})\bar{b}_k\bar{\xi}_k1_{\mathcal{B}_2}=0-(\mathbb{E}_k-\mathbb{E}_{k-1})\bar{b}_k\bar{\xi}_k(1_{\mathcal{B}_2^k}-1_{\mathcal{B}_2}).$$

Recalling (4.5.29) and (4.5.32) and using Assumptions 4.1 we have

$$\mathbb{E}\Big|(\mathbb{E}_k-\mathbb{E}_{k-1})\bar{b}_k\bar{\xi}_k1_{\mathcal{B}_2}\Big|\le 2\mathbb{E}\Big|\bar{b}_k\bar{\xi}_k(1_{\mathcal{B}_2^k}-1_{\mathcal{B}_2})\Big|=O(\sigma_i^2\theta^{-1})\mathbb{E}|1_{\mathcal{B}_2^k}-1_{\mathcal{B}_2}|=o(T^{-1}),$$

where the last equality follows from the fact that $\mathcal{B}_2\subseteq\mathcal{B}_2^k$ and Lemma 4.1. For the second term in (4.5.34), using $\mathcal{B}_2\subseteq\mathcal{B}_2^k$, (4.5.32) and (4.5.29) we have

$$\mathbb{E}\Big|(\mathbb{E}_k-\mathbb{E}_{k-1})\bar{b}_k\underline{\xi}_k1_{\mathcal{B}_2}\Big|^2\lesssim 4\mathbb{E}\Big|\bar{b}_k\underline{\xi}_k1_{\mathcal{B}_2^k}\Big|^2=o(T^{-1/2}).$$

Similarly the third term of (4.5.34) is bounded by

$$\mathbb{E}\Big|(\mathbb{E}_k-\mathbb{E}_{k-1})\Big[b_k\bar{b}_k(\underline{\xi}_k+\underline{\eta}_k)(\bar{\xi}_k+\underline{\xi}_k)1_{\mathcal{B}_2}\Big]\Big|\lesssim 2\mathbb{E}\Big|(\underline{\xi}_k+\underline{\eta}_k)(\bar{\xi}_k+\underline{\xi}_k)1_{\mathcal{B}_2^k}\Big|.$$

Expanding, applying the Cauchy-Schwarz inequality and using (4.5.32) and (4.5.29), we may obtain a bound of order $o_{L^1}(T^{-1/2})$; we omit the repetitive details. Substituting the above bounds back into equation (4.5.34) we obtain

$$\mathbb{E}\Big|\sum_{i=1}^{K}(\mathbb{E}_i - \mathbb{E}_{i-1})I_1\Big| = \mathbb{E}\Big|(\mathbb{E}_k - \mathbb{E}_{k-1})b_k\xi_k\Big| = o\bigg(\frac{1}{\sqrt{T}}\bigg).$$

The cases of I_2 and I_3 can be dealt with with similar approaches and we omit the details. In fact, from the definitions in (4.5.26) it is not difficult to see that η and ζ are higher order terms relative to ξ under the event $\mathcal{B}_2$. It can therefore be shown that the term involving I_1 is the leading term in (4.5.33) and the claim follows.

(b) Define $\overline{Q} := I_T - \theta^{-1}X_0^\top X_0RE_\tau^\top E_\tau$ so that similar to (4.2.15) we have

$$Q^{-1} - I_K = \frac{1}{\theta}X_0RE_\tau^\top E_\tau X_0^\top(I_K - \theta^{-1}X_0RE_\tau^\top E_\tau X_0^\top)^{-1} = \frac{1}{\theta}X_0RE_\tau^\top E_\tau\overline{Q}^{-1}X_0^\top.$$

Recall that we have $X_0^\top X_0 = T^{-1}\sum_{i=1}^K \mathbf{x}_i\mathbf{x}_i^\top$, define the matrices

$$\overline{Q}_{(j)} := I_T - \frac{1}{\theta T}\sum_{k\neq j}\mathbf{x}_k\mathbf{x}_k^\top RE_\tau^\top E_\tau, \quad \overline{Q}_{(ij)} := I_T - \frac{1}{\theta T}\sum_{k\neq i,j}\mathbf{x}_k\mathbf{x}_k^\top RE_\tau^\top E_\tau,$$

so that $\overline{Q} - \overline{Q}_{(j)} = -\frac{1}{\theta T}\mathbf{x}_j\mathbf{x}_j^\top RE_\tau^\top E_\tau$ and $\overline{Q}_{(j)} - \overline{Q}_{(ij)} = -\frac{1}{\theta T}\mathbf{x}_i\mathbf{x}_i^\top RE_\tau^\top E_\tau$. Let

$$a_j := \frac{1}{1 + \mathrm{tr}(\overline{Q}_{(j)}^{-1}(\overline{Q} - \overline{Q}_{(j)}))} = \frac{1}{1 - \frac{1}{\theta T}\mathbf{x}_j^\top RE_\tau^\top E_\tau\overline{Q}_{(j)}^{-1}\mathbf{x}_j},$$
$$a_{ij} := \frac{1}{1 + \mathrm{tr}(\overline{Q}_{(ij)}^{-1}(\overline{Q}_{(j)} - \overline{Q}_{(ij)}))} = \frac{1}{1 - \frac{1}{\theta T}\mathbf{x}_i^\top RE_\tau^\top E_\tau\overline{Q}_{(ij)}^{-1}\mathbf{x}_i},$$

then by (4.5.2) we have

$$\overline{Q}^{-1}\mathbf{x}_j = a_j\overline{Q}_{(j)}^{-1}\mathbf{x}_j, \quad \mathbf{x}_i^\top RE_\tau^\top E_\tau\overline{Q}_{(i)}^{-1} = a_{ij}\mathbf{x}_i^\top RE_\tau^\top E_\tau\overline{Q}_{(ij)}^{-1}.$$

We can therefore write

$$Q_{ij}^{-1} = \frac{1}{\theta T}\mathbf{x}_i^\top RE_\tau^\top E_\tau\overline{Q}^{-1}\mathbf{x}_j = \frac{a_ja_{ij}}{\theta T}\mathbf{x}_i^\top RE_\tau^\top E_\tau\overline{Q}_{(ij)}^{-1}\mathbf{x}_j.$$

Now define $X_{ij0} := X_0 - T^{-1/2}(\mathbf{e}_i\mathbf{x}_i^\top + \mathbf{e}_j\mathbf{x}_j)$ and events $\mathcal{B}_1^{ij}$ and $\mathcal{B}_2^{ij}$ analogous to (4.5.27) with X_{i0} replaced by X_{ij0}. Similar to (a) of the Lemma we have $a_j = 1 + o(1)$ and $a_{ij} = 1 + o(1)$ under the event $\mathcal{B}_2$. Therefore we have

$$\mathbb{E}\Big|Q_{ij}^{-1}1_{\mathcal{B}_2}\Big|^2 \lesssim \frac{1}{\theta^2T^2}\mathbb{E}\Big|\mathbf{x}_i^\top RE_\tau^\top E_\tau\overline{Q}_{(ij)}^{-1}\mathbf{x}_j1_{\mathcal{B}_2}\Big|^2 \leq \frac{1}{\theta^2T^2}\mathbb{E}\Big|\mathbf{x}_i^\top RE_\tau^\top E_\tau\overline{Q}_{(ij)}^{-1}\mathbf{x}_j1_{\mathcal{B}_2^{ij}}\Big|^2.$$

By Lemma 4.14 and Assumptions 4.1 we have

$$\mathbb{E}\Big|Q_{ij}^{-1}1_{\mathcal{B}_2}\Big|^2 = \frac{1}{\theta^2T^2}O(\sigma_i^2\sigma_j^2T) = O\bigg(\frac{1}{\gamma_1(\tau)^4\sigma_1^4T}\bigg).$$

Finally we remark that the uniformity of this bound in $i, j = 1, \ldots, K$ should be obvious from the proof since all σ_i's are of the same order. $\square$

We finally show that the conditional expectations of diagonal elements of A, B and Q^{-1}, defined in (4.2.13), are sufficiently close to the unconditional expectations.

Lemma 4.18. *For each $i = 1, \dots, K$, we have*

$$\underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}] - \mathbb{E}[A_{ii}1_{\mathcal{B}_0}] = O_{L^2}\left(\frac{1}{\sqrt{\theta T}}\right), \quad \underline{\mathbb{E}}[B_{ii}1_{\mathcal{B}_0}] - \mathbb{E}[B_{ii}1_{\mathcal{B}_0}] = O_{L^2}\left(\frac{1}{\sqrt{\theta T}}\right),$$
$$\underline{\mathbb{E}}[Q_{ii}^{-1}1_{\mathcal{B}_2}] - \mathbb{E}[Q_{ii}^{-1}1_{\mathcal{B}_2}] = o_{L^2}\left(\frac{1}{\sqrt{T}}\right).$$

Proof. From (a) of Lemma 4.14 we recall that

$$\underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}] = \frac{1}{\sqrt{\theta T}}\underline{\mathbb{E}}[\mathbf{x}_{i,[1:T-\tau]}^\top R 1_{\mathcal{B}_0}\mathbf{x}_{i,[\tau+1:T]}] = \frac{1}{\sqrt{\theta T}}\mathrm{tr}(\Psi_1^{ii}(R))1_{\mathcal{B}_0} \tag{4.5.35}$$

where, using (4.5.4) and the cyclic property of the trace, we have

$$\mathrm{tr}(\Psi_1^{ii}(R)) = \mathrm{tr}\Big((\mathbf{0}, I_{T-\tau})(\sigma_i^2\Phi_i^\top\Phi_i + I_T)(I_{T-\tau}, \mathbf{0})^\top R\Big) =: \mathrm{tr}(GR). \tag{4.5.36}$$

Furthermore, using (a) of Lemma 4.12, we see that

$$G := (\mathbf{0}, I_{T-\tau})(\sigma_i^2\Phi_i^\top\Phi_i + I_T)(I_{T-\tau}, \mathbf{0})^\top = O_{\|\cdot\|}(\sigma_i^2). \tag{4.5.37}$$

From (4.5.35) we have $\mathbb{E}[A_{ii}1_{\mathcal{B}_0}] = \mathbb{E}[\underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}]] = \frac{1}{\sqrt{\theta T}}\mathbb{E}[\mathrm{tr}(\Psi_1^{ii}(R))1_{\mathcal{B}_0}]$ and so

$$\begin{aligned}\underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}] - \mathbb{E}[A_{ii}1_{\mathcal{B}_0}] &= \frac{1}{\sqrt{\theta T}}\Big(\mathrm{tr}(\Psi_1^{ii}(R))1_{\mathcal{B}_0} - \mathbb{E}[\mathrm{tr}(\Psi_1^{ii}(R))1_{\mathcal{B}_0}]\Big)\\ &= \frac{1}{\sqrt{\theta T}}(\mathrm{tr}(GR)1_{\mathcal{B}_0} - \mathbb{E}[\mathrm{tr}(GR)1_{\mathcal{B}_0}]) = \frac{1}{\sqrt{\theta T}}\mathrm{tr}\Big(G(R1_{\mathcal{B}_0} - \mathbb{E}[R1_{\mathcal{B}_0}])\Big),\end{aligned}$$

by linearity of the expectation and the trace. By (a) of Lemma 4.16 we have

$$\underline{\mathbb{E}}[A_{ii}1_{\mathcal{B}_0}] - \mathbb{E}[A_{ii}1_{\mathcal{B}_0}] = \frac{1}{\sqrt{\theta}}O_{L^2}\left(\frac{\|G\|}{\theta\sqrt{T}}\right) = O_{L^2}\left(\frac{1}{\sqrt{\theta T}}\right),$$

where the last equality follows from (4.5.37) and Assumption 4.1. For the case of B, similar computations and (b) of Lemma 4.16 give

$$\begin{aligned}\underline{\mathbb{E}}[B_{ii}1_{\mathcal{B}_0}] - \mathbb{E}[B_{ii}1_{\mathcal{B}_0}] &= \frac{1}{\theta T}\Big(\mathrm{tr}(\Psi_1^{ii}(E_0^\top E_0 R))1_{\mathcal{B}_0} - \mathbb{E}[\mathrm{tr}(\Psi_1^{ii}(E_0^\top E_0 R))1_{\mathcal{B}_0}]\Big)\\ &= \frac{1}{\theta T}\mathrm{tr}\Big(G(E_0^\top E_0 R 1_{\mathcal{B}_0} - \mathbb{E}[E_0^\top E_0 R 1_{\mathcal{B}_0}]\Big) = \frac{1}{\theta}O_{L^2}\left(\frac{\|G\|}{\sqrt{T}}\right) = O_{L^2}\left(\frac{1}{\sqrt{\theta T}}\right).\end{aligned}$$

It remains to consider $\underline{\mathbb{E}}[Q_{ii}^{-1}]$. We recall from (4.2.15) that

$$Q := I_K - \frac{1}{\theta} X_0 R E_\tau^\top E_\tau X_0^\top. \tag{4.5.38}$$

The strategy of the proof, similar to that of Lemma 4.16 and Lemma 4.17, is to express $\underline{\mathbb{E}}[Q_{ii}^{-1} 1_{\mathcal{B}_2}] - \mathbb{E}[Q_{ii}^{-1} 1_{\mathcal{B}_2}]$ as a sum of martingale differences. We first introduce the necessary notations and carry out some algebraic computations.

Similar to (4.5.10), we will define $\underline{\boldsymbol{\epsilon}}_{k0} := \boldsymbol{\epsilon}_{K+k,[1:T-\tau]}$ and $\underline{\boldsymbol{\epsilon}}_{k\tau} := \boldsymbol{\epsilon}_{K+k,[\tau+1:T]}$. Recall from (4.5.12) that $R - R_k = \sum_{n=1}^5 U_n$, where the U_n's are defined in (4.5.13). Similar to the computations in (4.5.18), we may obtain

$$R E_\tau^\top E_\tau - R_k E_{k\tau}^\top E_{k\tau} = \frac{1}{T} R_k \underline{\boldsymbol{\epsilon}}_{k\tau} \underline{\boldsymbol{\epsilon}}_{k\tau}^\top + \sum_{n=1}^5 U_n E_\tau^\top E_\tau =: V + W,$$

where we defined

$$\begin{aligned} V &:= \frac{1}{T} R_k \underline{\boldsymbol{\epsilon}}_{k\tau} \underline{\boldsymbol{\epsilon}}_{k\tau}^\top + (U_2 + U_3) E_\tau^\top E_\tau, \\ W &:= (U_1 + U_4 + U_5) E_\tau^\top E_\tau. \end{aligned}$$

Define matrices $V_1, V_2, V_3, W_1, W_2, W_3$ by,

$$\begin{aligned} V_1 &:= I_T, \quad V_2 := \frac{\underline{\beta}_k}{\theta} E_0^\top E_0 R_k E_\tau^\top E_\tau, \quad V_3 := \frac{\underline{\beta}_k \beta_k}{\theta^2 T} E_0^\top E_0 R_k E_{k\tau}^\top E_{k\tau} \boldsymbol{\epsilon}_{k0} \boldsymbol{\epsilon}_{k0}^\top R_k E_\tau^\top E_\tau \\ W_1 &:= \beta_k R_k E_\tau^\top E_\tau, \quad W_2 := \frac{\underline{\beta}_k \beta_k}{\theta T} R_k \underline{\boldsymbol{\epsilon}}_{k\tau} \underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0 R_k E_\tau^\top E_\tau, \\ W_3 &:= \frac{\underline{\beta}_k \beta_k^2}{\theta^2 T^2} R_k \underline{\boldsymbol{\epsilon}}_{k\tau} \underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0 R_k E_{k\tau}^\top E_{k\tau} \underline{\boldsymbol{\epsilon}}_{k0} \underline{\boldsymbol{\epsilon}}_{k0}^\top R_k E_\tau^\top E_\tau, \end{aligned}$$

so that using (4.5.13) we can decompose V and W into

$$V = \frac{1}{T} R_k \underline{\boldsymbol{\epsilon}}_{k\tau} \underline{\boldsymbol{\epsilon}}_{k\tau}^\top (V_1 + V_2 + V_3) \tag{4.5.39a}$$

$$W = \frac{1}{\theta T} R_k E_{k\tau}^\top E_{k\tau} \underline{\boldsymbol{\epsilon}}_{k0} \underline{\boldsymbol{\epsilon}}_{k0}^\top (W_1 + W_2 + W_3). \tag{4.5.39b}$$

It is clear that V and W are matrices of rank one. We define

$$\begin{aligned} \underline{Q}_{(k)} &: = I_K - \frac{1}{\theta} X_0 (R E_\tau^\top E_\tau - V) X_0^\top, \\ \underline{Q}_{(kk)} &: = I_K - \frac{1}{\theta} X_0 (R E_\tau^\top E_\tau - V - W) X_0^\top, \end{aligned}$$

then from (4.5.38) we can write $Q - \underline{Q}_{(k)} = -\theta^{-1} X_0 V X_0^\top$ and $Q_{(k)} - \underline{Q}_{(kk)} = -\theta^{-1} X_0 W X_0^\top$. Define the following scalar quantities

$$\alpha_k := \frac{1}{1 - \theta^{-1} \mathrm{tr}(Q_{(k)}^{-1} X_0 V X_0^\top)}, \quad \alpha_{kk} := \frac{1}{1 - \theta^{-1} \mathrm{tr}(Q_{(kk)}^{-1} X_0 W X_0^\top)},$$

then using (4.5.1) we obtain

$$Q^{-1} = \underline{Q}_{(k)}^{-1} + \frac{1}{\theta}\underline{Q}_{(k)}^{-1} X_0 V X_0^\top \underline{Q}_{(k)}^{-1}, \quad \underline{Q}_{(k)}^{-1} = \underline{Q}_{(kk)}^{-1} + \frac{1}{\theta}\underline{Q}_{(kk)}^{-1} X_0 W X_0^\top \underline{Q}_{(kk)}^{-1}.$$

Substituting the second identity into the first gives

$$\begin{aligned} Q^{-1} - \underline{Q}_{(kk)}^{-1} &= \frac{1}{\theta}\underline{Q}_{(kk)}^{-1} X_0 W X_0^\top \underline{Q}_{(kk)}^{-1} \\ &+ \frac{1}{\theta}\Big(\underline{Q}_{(kk)}^{-1} + \frac{1}{\theta}\underline{Q}_{(kk)}^{-1} X_0 W X_0^\top \underline{Q}_{(kk)}^{-1}\Big) X_0 V X_0^\top \Big(\underline{Q}_{(kk)}^{-1} + \frac{1}{\theta}\underline{Q}_{(kk)}^{-1} X_0 W X_0^\top \underline{Q}_{(kk)}^{-1}\Big), \end{aligned}$$

which after simplifying becomes

$$\begin{aligned} Q^{-1} - \underline{Q}_{(kk)}^{-1} &= \frac{1}{\theta}\underline{Q}_{(kk)}^{-1} X_0 V X_0^\top \underline{Q}_{(kk)}^{-1} + \frac{1}{\theta}\underline{Q}_{(kk)}^{-1} X_0 W X_0^\top \underline{Q}_{(kk)}^{-1} \\ &+ \frac{1}{\theta^2}\underline{Q}_{(kk)}^{-1} X_0 V X_0^\top \underline{Q}_{(kk)}^{-1} X_0 W X_0^\top \underline{Q}_{(kk)}^{-1} + \frac{1}{\theta^2}\underline{Q}_{(kk)}^{-1} X_0 W X_0^\top \underline{Q}_{(kk)}^{-1} X_0 V X_0^\top \underline{Q}_{(kk)}^{-1}. \end{aligned} \tag{4.5.40}$$

Before we proceed with the proof we first prove some moment estimates for the terms in (4.5.40). We start with some informal observations. By comparing (4.5.39a) and (4.5.39a), we see that the matrix W is smaller in magnitude in comparison to V by a factor of θ^{-1}. This suggests that the first term in (4.5.40) is the leading term while the rest are high order terms in comparison and we will therefore only deal with first term in detail below. The same arguments can be applied to the rest of (4.5.40) to make the above argument rigorous, but we omit the repetitive details.

Recall the family of event $\{\mathcal{B}_0^k, k = 1, \dots, p\}$ from (4.5.14) and define $\mathcal{B}_2^k := \mathcal{B}_0^k \cap \mathcal{B}_1$. From definition we note that $\mathcal{B}_2 \subseteq \mathcal{B}_2^k$. Furthermore, from Lemma 4.1 we have

$$1_{\mathcal{B}_2^k} - 1_{\mathcal{B}_2} \le 1 - 1_{\mathcal{B}_2} = o_p(T^{-l}), \quad \forall l \in \mathbb{N}. \tag{4.5.41}$$

In the computations below, we will often substitute $1_{\mathcal{B}_2^k}$ with $1_{\mathcal{B}_2}$ and vice versa in expectations. Whenever we do so, we may use (4.5.41) and a similar argument to how we dealt with (4.5.15) to show that the error term of such a substitution is negligible for the purpose of the proof. Hence from now on we will use the two indicators $1_{\mathcal{B}_2^k}$ and $1_{\mathcal{B}_2}$ interchangeably below without further justifications.

Since we can write $X_0^\top = \frac{1}{\sqrt{T}}\sum_{l=1}^K \mathbf{x}_l \mathbf{e}_l^\top$, the first term in (4.5.40) can be expressed as

$$\frac{1}{\theta}\mathbf{e}_i^\top \underline{Q}_{(kk)}^{-1} X_0 V X_0^\top \underline{Q}_{(kk)}^{-1} \mathbf{e}_i = \frac{1}{\theta T}\sum_{l=1}^K \sum_{m=1}^K \mathbf{e}_i^\top \underline{Q}_{(kk)}^{-1} \mathbf{e}_l (\mathbf{x}_l^\top V \mathbf{x}_m) \mathbf{e}_m^\top \underline{Q}_{(kk)}^{-1} \mathbf{e}_i, \tag{4.5.42}$$

where, recalling (4.5.39a), we have

$$\mathbf{x}_l^\top V \mathbf{x}_m = \frac{1}{T}\sum_{n=1}^3 \mathbf{x}_l^\top R_k \underline{\boldsymbol{\epsilon}}_{k\tau} \underline{\boldsymbol{\epsilon}}_{k\tau}^\top V_n \mathbf{x}_m. \tag{4.5.43}$$

Using (4.5.42)-(4.5.43) and the inequality $(\sum_{i=1}^n x_i)^p \lesssim n^{p-1}\sum_{i=1}^n x_i^p$ we have

$$\mathbb{E}\Big|\frac{1}{\theta}\mathbf{e}_i^\top \underline{Q}_{(kk)}^{-1}X_0 V X_0^\top \underline{Q}_{(kk)}^{-1}\mathbf{e}_i 1_{\mathcal{B}_2}\Big|^2 \lesssim \frac{K^2}{\theta^2 T^2}\sum_{l=1}^K\sum_{m=1}^K \mathbb{E}\Big|\mathbf{e}_i^\top \underline{Q}_{(kk)}^{-1}\mathbf{e}_l(\mathbf{x}_l^\top V\mathbf{x}_m)\mathbf{e}_m^\top \underline{Q}_{(kk)}^{-1}\mathbf{e}_i 1_{\mathcal{B}_2}\Big|^2$$
$$\lesssim \frac{3K^2}{\theta^2 T^4}\sum_{l=1}^K\sum_{m=1}^K\sum_{n=1}^3 \mathbb{E}\Big|\mathbf{x}_l^\top R_k \underline{\boldsymbol{\epsilon}}_{k\tau}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top V_n \mathbf{x}_m \|\underline{Q}_{(kk)}^{-1}\|^2 1_{\mathcal{B}_2}\Big|^2. \tag{4.5.44}$$

Note that under the event $\mathcal{B}_2$, we can easily see that $\|Q_{(kk)}^{-1}1_{\mathcal{B}_2}\| = O(1)$. Therefore by the Cauchy Schwarz inequality we can obtain

$$(4.5.44) \lesssim \frac{K^2}{\theta^2 T^4}\sum_{l=1}^K\sum_{m=1}^K\sum_{n=1}^3 \mathbb{E}\Big[(\mathbf{x}_l^\top R_k \underline{\boldsymbol{\epsilon}}_{k\tau})^4 1_{\mathcal{B}_2}\Big]^{1/2}\mathbb{E}\Big[(\underline{\boldsymbol{\epsilon}}_{k\tau}^\top V_n \mathbf{x}_m)^4 1_{\mathcal{B}_2}\Big]^{1/2}. \tag{4.5.45}$$

Note that $\mathcal{B}_2 = \mathcal{B}_1 \cap \mathcal{B}_0 \subseteq \mathcal{B}_0 \subseteq \mathcal{B}_0^k$ so that $1_{\mathcal{B}_2} \le 1_{\mathcal{B}_0^k}$. We can then condition on R_k and apply (a) of Lemma 4.13 to the first quadratic form in (4.5.44) to get

$$\mathbb{E}|\mathbf{x}_l^\top R_k \underline{\boldsymbol{\epsilon}}_{k\tau} 1_{\mathcal{B}_2}|^4 \le \mathbb{E}\left|\begin{pmatrix}\mathbf{z}_{l,[1:T]}\\ \boldsymbol{\epsilon}_{l,[1:T]}\end{pmatrix}^\top \begin{pmatrix}\sigma_l\Phi_l\\ I_T\end{pmatrix}\begin{pmatrix}I_{T-\tau}\\ \mathbf{0}_{\tau\times(T-\tau)}\end{pmatrix} R_k \underline{\boldsymbol{\epsilon}}_{k\tau} 1_{\mathcal{B}_0^k}\right|^4$$
$$\lesssim \mathbb{E}\Big[\mathrm{tr}(R_k^2(\sigma_l^2\Phi_l^\top\Phi_l + I_T))^2 1_{\mathcal{B}_0^k}\Big] = O(\sigma_l^4 T^2),$$

where the last equality follows from using $tr(R) \le T\|R\|$ and applying Lemma 4.12. Similarly, for the quadratic involving V_1 in (4.5.45), we have

$$\mathbb{E}|\underline{\boldsymbol{\epsilon}}_{k\tau}^\top V_1 \mathbf{x}_m|^4 = \mathbb{E}|\underline{\boldsymbol{\epsilon}}_{k\tau}^\top \mathbf{x}_m|^4 \lesssim \mathrm{tr}(\sigma_m^2\Phi_m^\top\Phi_m + I_T)^2 = O(\sigma_m^4 T^2).$$

We observe here that that the matrices V_2 and V_3 are smaller in magnitude in comparison to V_1 by a factor of θ^{-1} under the event $\mathcal{B}_2$. Hence it is to be expected that the quadratic forms involving V_2 and V_3 in (4.5.45) should be negligible in comparison to the one involving V_1. To be more concrete, we sketch here how bound the quadratic form involving V_2; the case of V_3 can be dealt with in a similar manner. Recall that the matrix $E_0^\top E_0$ can be written as $E_0^\top E_0 = E_{k0}^\top E_{k0} + \frac{1}{T}\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top$. Then we can write

$$\mathbb{E}|\underline{\boldsymbol{\epsilon}}_{k\tau}^\top V_2 \mathbf{x}_m 1_{\mathcal{B}_2}|^4 = \frac{1}{\theta^4}\mathbb{E}\Big|\underline{\beta}_k \underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_0^\top E_0 R_k E_\tau^\top E_\tau \mathbf{x}_m 1_{\mathcal{B}_2}\Big|^4$$
$$\lesssim \frac{1}{\theta^4}\mathbb{E}\Big|\underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_{k0}^\top E_{k0} R_k E_{k\tau}^\top E_{k\tau}\mathbf{x}_m 1_{\mathcal{B}_2}\Big|^4 + \frac{1}{\theta^4}\mathbb{E}\Big|\frac{1}{T}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top\underline{\boldsymbol{\epsilon}}_{k0}\frac{1}{T}\underline{\boldsymbol{\epsilon}}_{k0}^\top R_k \underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top \mathbf{x}_m 1_{\mathcal{B}_2}\Big|^4$$
$$+ \frac{1}{\theta^4}\mathbb{E}\Big|\frac{1}{T}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top E_{k0}^\top E_{k0} R_k \underline{\boldsymbol{\epsilon}}_{k\tau}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top \mathbf{x}_m 1_{\mathcal{B}_2}\Big|^4 + \frac{1}{\theta^4}\mathbb{E}\Big|\frac{1}{T}\underline{\boldsymbol{\epsilon}}_{k\tau}^\top\underline{\boldsymbol{\epsilon}}_{k0}\underline{\boldsymbol{\epsilon}}_{k0}^\top R_k E_{k0}^\top E_{k0}\mathbf{x}_m 1_{\mathcal{B}_2}\Big|^4.$$

At this point we recognize that the four terms above have a similar structure as in the case of V_1. Namely they all involve quadratic forms where the matrix in the middle is independent from the vectors on each side. Using the same approach as we did in the case of V_1 we can indeed show that this is a negligible term in comparison. The case of V_3 is similar albeit more tedious, and we omit the details.

After the above arguments, we may conclude that

$$\mathbb{E}\Big|\frac{1}{\theta}\mathbf{e}_i^\top \underline{Q}_{(kk)}^{-1} X_0 V X_0^\top \underline{Q}_{(kk)}^{-1} \mathbf{e}_i 1_{\mathcal{B}_2}\Big|^2 \lesssim \frac{K^2}{\theta^2 T^4}\sum_{l=1}^{K}\sum_{m=1}^{K}\sigma_l^2\sigma_m^2 T^2 = o\bigg(\frac{1}{\sqrt{\theta}T^2}\bigg). \tag{4.5.46}$$

The same strategy described above can then be repeated for each of the remaining three terms in (4.5.40) to show that they are negligible (c.f. the remark right below (4.5.40)). We may therefore conclude that

$$\mathbb{E}\Big|\mathbf{e}_i^\top (Q^{-1} - \underline{Q}_{(kk)}^{-1})\mathbf{e}_i 1_{\mathcal{B}_2}\Big|^2 = o\bigg(\frac{1}{\sqrt{\theta}T^2}\bigg). \tag{4.5.47}$$

Finally, we can decompose $\underline{\mathbb{E}}[\mathbf{e}_i^\top Q^{-1}\mathbf{e}_i 1_{\mathcal{B}_2}] - \mathbb{E}[\mathbf{e}_i^\top Q^{-1}\mathbf{e}_i 1_{\mathcal{B}_2}]$ into

$$\begin{aligned}\underline{\mathbb{E}}[\mathbf{e}_i^\top Q^{-1}\mathbf{e}_i 1_{\mathcal{B}_2}] - \mathbb{E}[\mathbf{e}_i^\top Q^{-1}\mathbf{e}_i 1_{\mathcal{B}_2}] &= \sum_{k=1}^{p}(\mathbb{E}_i - \mathbb{E}_{i-1})\mathbf{e}_i^\top Q^{-1}\mathbf{e}_i 1_{\mathcal{B}_2} \\ &= \sum_{k=1}^{p}(\mathbb{E}_i - \mathbb{E}_{i-1})\mathbf{e}_i^\top (Q^{-1} - Q_{(kk)}^{-1})\mathbf{e}_i 1_{\mathcal{B}_2},\end{aligned}$$

where for the last equality we refer to (4.5.41) and the remark immediately below it. Using the bound (4.5.47) we immediately have

$$\begin{aligned}&\mathbb{E}\Big|\underline{\mathbb{E}}[\mathbf{e}_i^\top Q^{-1}\mathbf{e}_i 1_{\mathcal{B}_2}] - \mathbb{E}[\mathbf{e}_i^\top Q^{-1}\mathbf{e}_i 1_{\mathcal{B}_2}]\Big|^2 \\ &\qquad \le 4\sum_{k=1}^{p}\mathbb{E}|\mathbf{e}_i^\top (Q^{-1} - \underline{Q}_{(kk)}^{-1} 1_{\mathcal{B}_2})\mathbf{e}_i|^2 = o\bigg(\frac{p}{\sqrt{\theta}T^2}\bigg),\end{aligned}$$

from which the claim follows. □

4.6 Conclusion and Future Work

In this chapter we focused on the asymptotic theory of a high dimensional time series arising from a factor model. We established the asymptotic normality of the spiked eigenvalues of the product symmetrized sample auto-covariance matrix of the data.

Our work serves as a first step in understanding the asymptotic distributions of eigenvalues of the auto-covariance matrix. So far we have dealt exclusively with spiked eigenvalues which diverge as the dimension and sample size tend to infinity; a natural next step is to study the asymptotic distributions of the non-spiked eigenvalues. Based on what is known about spiked covariance matrices, it is reasonable to suspect that the non-spiked eigenvalues are not asymptotically normal, but should tend to the Tracy-Widom distribution instead. Establishing this result is of both theoretical and practical interest.

Bibliography

[1] Fernando Albiac and Nigel John Kalton. *Topics in Banach Space Theory*, volume 233. Springer, 2016.

[2] Theodore Wilbur Anderson. Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics*, 34(1):122–148, 1963.

[3] D. Applebaum. Levy processes and stochastic integrals in Banach spaces. *Probability and Mathematical Statistics- Wroclaw University*, 27:75–88, 2007. ISSN 0208-4147.

[4] David Applebaum. Ornstein-Uhlenbeck Processes with Jumps in Hilbert Space. page 18.

[5] David Applebaum. Martingale-valued measures, Ornstein-Uhlenbeck processes with jumps and operator self-decomposability in Hilbert space. In *In Memoriam Paul-André Meyer*, pages 171–196. Springer, 2006.

[6] David Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, Cambridge; New York, 2009.

[7] David Applebaum. Infinite dimensional Ornstein-Uhlenbeck processes driven by Lévy processes. *Probability Surveys*, 12(0):33–54, 2015. ISSN 1549-5787. doi: 10.1214/14-PS249.

[8] David Applebaum and Markus Riedle. Cylindrical Lévy processes in Banach spaces. *Proceedings of the London Mathematical Society*, 101(3):697–726, November 2010. ISSN 00246115. doi: 10.1112/plms/pdq004.

[9] Aloisio Araujo and Evarist Gine. Type, Cotype and Levy Measures in Banach Spaces. *The Annals of Probability*, 6(4):637–643, 1978. ISSN 0091-1798.

[10] Jushan Bai. Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, 71(1):135–171, 2003. ISSN 1468-0262. doi: 10.1111/1468-0262.00392.

[11] Jushan Bai and Kunpeng Li. Statistical analysis of factor models of high dimension. *Ann. Statist.*, 40(1):436–465, February 2012. ISSN 0090-5364, 2168-8966. doi: 10.1214/11-AOS966.

[12] Jushan Bai and Serena Ng. Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1):191–221, 2002. ISSN 1468-0262. doi: 10.1111/1468-0262.00273.

[13] Z. D. Bai and Jack W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *The Annals of Probability*, 26(1):316–345, January 1998. doi: 10.1214/aop/1022855421.

[14] Zhidong Bai and Jack W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*, volume 20. Springer, 2010.

[15] Zhidong Bai and Jian-feng Yao. Central limit theorems for eigenvalues in a spiked population model. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 44(3):447–474, June 2008. ISSN 0246-0203. doi: 10.1214/07-AIHP118.

[16] Jinho Baik and Jack W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, July 2006. ISSN 0047259X. doi: 10.1016/j.jmva.2005.08.003.

[17] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, September 2005. ISSN 0091-1798. doi: 10.1214/009117905000000233.

[18] Zhigang Bao, Xiucai Ding, and Ke Wang. Singular vector and singular subspace distribution for the matrix denoising model. *arXiv preprint arXiv:1809.10476*, 2018.

[19] Zhigang Bao, Xiucai Ding, Jingming Wang, and Ke Wang. Statistical inference for principal components of spiked covariance matrices. *arXiv:2008.11903 [math, stat]*, September 2020.

[20] Ole E. Barndorff-Nielsen and Neil Shephard. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241, January 2001. ISSN 1369-7412. doi: 10.1111/1467-9868.00282.

[21] Ole E. Barndorff-Nielsen, Jan Pedersen, and Ken-Iti Sato. Multivariate subordination, self-decomposability and stability. *Advances in Applied Probability*, 33(01):160–187, 2001.

[22] Ole E. Barndorff-Nielsen, Fred Espen Benth, and Almut ED Veraart. Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. *arXiv preprint arXiv:1210.1354*, 2012.

[23] Ole E. Barndorff-Nielsen, Fred Espen Benth, Almut ED Veraart, et al. Modelling energy spot prices by volatility modulated Lévy-driven Volterra processes. *Bernoulli*, 19(3):803–845, 2013.

[24] Ole E. Barndorff-Nielsen, Fred Espen Benth, and Almut ED Veraart. Turbulence Modelling. In *Ambit Stochastics*, pages 303–332. Springer, 2018.

[25] Michał Barski and Jerzy Zabczyk. Heath–Jarrow–Morton–Musiela equation with Lévy perturbation. *Journal of Differential Equations*, 253(9):2657–2697, November 2012. ISSN 0022-0396. doi: 10.1016/j.jde.2012.06.022.

[26] Anirban Basak, Arup Bose, and Sanchayan Sen. Limiting spectral distribution of sample autocovariance matrices. *Bernoulli*, 20(3):1234–1259, 2014.

[27] A. Behme. *Generalized Ornstein-Uhlenbeck Processes and Extensions*. PhD Thesis, Doctoral Thesis, TU Braunschweig, 2011.

[28] Anita Behme and Alexander Lindner. Multivariate generalized Ornstein–Uhlenbeck processes. *Stochastic Processes and their Applications*, 122(4):1487–1518, April 2012. ISSN 0304-4149. doi: 10.1016/j.spa.2012.01.002.

[29] Anita Behme, Alexander Lindner, and Ross Maller. Stationary solutions of the stochastic differential equation dVt=Vt-dUt+dLt with Lévy noise. *Stochastic Processes and their Applications*, 121(1):91–108, January 2011. ISSN 03044149. doi: 10.1016/j.spa.2010.09.003.

[30] Albert Benassi, Serge Cohen, and Jacques Istas. Identification and properties of real harmonizable fractional Lévy motions. *Bernoulli*, 8(1):97–115, February 2002. ISSN 1350-7265.

[31] Albert Benassi, Serge Cohen, and Jacques Istas. On roughness indices for fractional fields. *Bernoulli*, pages 357–373, 2004.

[32] Christian Bender and Tina Marquardt. Stochastic calculus for convoluted Lévy processes. *Bernoulli*, 14(2):499–518, 2008. ISSN 1350-7265. doi: 10.3150/07-BEJ115.

[33] Kenneth N. Berk. A central limit theorem for m-dependent random variables with unbounded m. *The Annals of Probability*, pages 352–354, 1973.

[34] Ben S. Bernanke and Jean Boivin. Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3):525–546, 2003.

[35] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, June 2013. ISBN 978-1-118-62596-5.

[36] Alex Bloemendal, Antti Knowles, Horng-Tzer Yau, and Jun Yin. On the principal components of sample covariance matrices. *Probability theory and related fields*, 164 (1-2):459–552, 2016.

[37] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[38] Arup Bose and Monika Bhattacharjee. *Large Covariance and Autocovariance Matrices*. CRC Press, 2018.

[39] Albrecht Böttcher and Bernd Silbermann. *Introduction to Large Truncated Toeplitz Matrices*. Universitext. Springer New York, New York, NY, 1999. ISBN 978-1-4612-7139-0 978-1-4612-1426-7. doi: 10.1007/978-1-4612-1426-7.

[40] Philippe Bougerol and Nico Picard. Stationarity of GARCH processes and of some nonnegative time series. *Journal of econometrics*, 52(1-2):115–127, 1992.

[41] Svetlana Boyarchenko and Sergei Levendorskiĭ. Barrier options and touch-and-out options under regular Lévy processes of exponential type. *The Annals of Applied Probability*, 12(4):1261–1298, November 2002. doi: 10.1214/aoap/1037125863.

[42] David R. Brillinger. Some aspects of modern population mathematics. *Canadian Journal of Statistics*, 9(2):173–194, 1981.

[43] Peter J. Brockwell, Richard A. Davis, and Stephen E. Fienberg. *Time Series: Theory and Methods: Theory and Methods*. Springer Science & Business Media, 1991.

[44] Zdzisław Brzeźniak and Erika Hausenblas. Maximal regularity for stochastic convolutions driven by Lévy processes. *Probability Theory and Related Fields*, 145(3-4):615–637, November 2009. ISSN 0178-8051, 1432-2064. doi: 10.1007/s00440-008-0181-7.

[45] Zdzisław Brzeźniak and Jerzy Zabczyk. Regularity of Ornstein–Uhlenbeck Processes Driven by a Lévy White Noise. *Potential Analysis*, 32(2):153–188, February 2010. ISSN 0926-2601, 1572-929X. doi: 10.1007/s11118-009-9149-1.

[46] Boris Buchmann and Gernot Müller. Limit experiments of GARCH. *Bernoulli*, 18 (1):64–99, 2012. ISSN 1350-7265. doi: 10.3150/10-BEJ328.

[47] Boris Buchmann, Benjamin Kaehler, Ross Maller, and Alexander Szimayer. Multivariate subordination using generalised Gamma convolutions with applications to Variance Gamma processes and option pricing. *Stochastic Processes and their Applications*, 127(7):2208–2242, July 2017. ISSN 03044149. doi: 10.1016/j.spa.2016.10.008.

[48] Boris Buchmann, Kevin W. Lu, and Dilip B. Madan. Calibration for Weak Variance-Alpha-Gamma Processes. *Methodology and Computing in Applied Probability*, pages 1–14, 2018.

[49] Boris Buchmann, Kevin W. Lu, and Dilip B. Madan. Weak subordination of multivariate Lévy processes and variance generalised gamma convolutions. *Bernoulli*, 25(1):742–770, 2019.

[50] Boris Buchmann, Kevin W. Lu, and Dilip B. Madan. Self-decomposability of weak variance generalised gamma convolutions. *Stochastic Processes and their Applications*, 2019.

[51] Tony Cai, Zongming Ma, and Yihong Wu. Optimal estimation and rank detection for sparse spiked covariance matrices. *Probability theory and related fields*, 161(3-4): 781–815, 2015.

[52] Tony Cai, Xiao Han, and Guangming Pan. Limiting Laws for Divergent Spiked Eigenvalues and Largest Non-spiked Eigenvalue of Sample Covariance Matrices. *arXiv:1711.00217 [math, stat]*, November 2017.

[53] B.B. Chen and G.M. Pan. Convergence of the largest eigenvalue of normalized sample covariance matrices when \$p\$ and \$n\$ both tend to infinity with their ratio converging to zero. *Bernoulli*, 18(4):1405–1420, November 2012. ISSN 1350-7265. doi: 10.3150/11-BEJ381.

[54] Binbin Chen and Guangming Pan. CLT for linear spectral statistics of normalized sample covariance matrices with the dimension much larger than the sample size. *Bernoulli*, 21(2):1089–1133, 2015.

[55] Yong Chen and Hongjun Gao. Well-posedness and large deviations for a class of SPDEs with Lévy noise. *Journal of Differential Equations*, 263(9):5216–5252, November 2017. ISSN 00220396. doi: 10.1016/j.jde.2017.06.016.

[56] Erhan Çınlar. *Probability and Stochastics*, volume 261. Springer Science & Business Media, 2011.

[57] Rama Cont. Empirical properties of asset returns: Stylized facts and statistical issues. 2001.

[58] Valentina Corradi. Reconsidering the continuous time limit of the GARCH(1,1) process. *Journal of Econometrics*, 96(1):145–153, May 2000. doi: 10.1016/S0304-4076(99)00053-6.

[59] M. H. A. Davis and M. R. Pistorius. Explicit solution of an inverse first-passage time problem for Lévy processes and counterparty credit risk. *Ann. Appl. Probab.*, 25(5): 2383–2415, October 2015. ISSN 1050-5164, 2168-8737. doi: 10.1214/14-AAP1051.

[60] Lukasz Delong and Claudia Klüppelberg. Optimal investment and consumption in a Black–Scholes market with Lévy-driven stochastic coefficients. *Ann. Appl. Probab.*, 18(3):879–908, June 2008. ISSN 1050-5164, 2168-8737. doi: 10.1214/07-AAP475.

[61] E. Dettweiler. Banach space valued processes with independent increments and stochastic integration. In Anatole Beck and Konrad Jacobs, editors, *Probability in Banach Spaces IV*, volume 990, pages 54–83. Springer Berlin Heidelberg, Berlin, Heidelberg, 1983. ISBN 978-3-540-12295-1 978-3-540-39870-7. doi: 10.1007/BFb0064263.

[62] Odo Diekmann, Stephan A. van Gils, Sjoerd M. V. Lunel, and Hans-Otto Walther. *Delay Equations: Functional-, Complex-, and Nonlinear Analysis*. Springer Science & Business Media, 2012. ISBN 978-1-4612-4206-2.

[63] Zhao Dong and Yingchao Xie. Ergodicity of stochastic 2D Navier–Stokes equation with Lévy noise. *Journal of Differential Equations*, 251(1):196–222, July 2011. ISSN 00220396. doi: 10.1016/j.jde.2011.03.015.

[64] JC Duan. Augmented GARCH (p, q) process and its diffusion limit. *Journal of Econometrics*, 1997.

[65] William Dunsmuir, Ben Goldys, and Cuong Viet Tran. On Limits of Continuous Time Bilinear Processes (working paper). .

[66] William Dunsmuir, Ben Goldys, and Cuong Viet Tran. Stochastic Delay Differential Equations as Weak Limits of GARCH processes (working paper). .

[67] Noureddine El Karoui. Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *The Annals of Probability*, 35(2):663–714, 2007.

[68] Robert F. Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 1982.

[69] Ghysels Eric and Joann Jasiak. GARCH for Irregularly Spaced Financial Data: The ACD-GARCH Model. *Studies in Nonlinear Dynamics & Econometrics*, 2(4):1–19, 1998.

[70] Jianqing Fan, Jinchi Lv, and Lei Qi. Sparse high-dimensional models in economics. 2011.

[71] Jianqing Fan, Yuan Liao, and Martina Mincheva. Large Covariance Estimation by Thresholding Principal Orthogonal Complements. *J R Stat Soc Series B Stat Methodol*, 75(4), September 2013. ISSN 1369-7412. doi: 10.1111/rssb.12016.

[72] Pani W. Fernando, E. Hausenblas, and Paul A. Razafimandimby. Irreducibility and Exponential Mixing of Some Stochastic Hydrodynamical Systems Driven by Pure Jump Noise. *Communications in Mathematical Physics*, 348(2):535–565, December 2016. ISSN 0010-3616, 1432-0916. doi: 10.1007/s00220-016-2693-9.

[73] Christian Andrés Fonseca Mora. *Stochastic Analysis with Lévy Noise in the Dual of a Nuclear Space.* PhD thesis, University of Sheffield, October 2015.

[74] Mario Forni and Lucrezia Reichlin. Let's get real: A factor analytical approach to disaggregated business cycle dynamics. *The Review of Economic Studies*, 65(3): 453–473, 1998.

[75] Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82(4):540–554, 2000.

[76] Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. Coincident and leading indicators for the euro area. *The Economic Journal*, 111(471):C62–C85, 2001.

[77] Eric Ghysels and Joanna Jasiak. GARCH for irregularly spaced financial data: The ACD-GARCH model. *Studies in Nonlinear Dynamics & Econometrics*, 2(4), 1998.

[78] Iosif I. Gikhman and Anatoli V. Skorokhod. *The Theory of Stochastic Processes I.* Springer, March 2015. ISBN 978-3-642-61943-4.

[79] Jack Hale and Sjoerd M Verduyn Lunel. *Introduction to Functional Differential Equations.* Springer Science & Business Media, 2013.

[80] Edward James Hannan. *Multiple Time Series*, volume 38. John Wiley & Sons, 2009.

[81] Andrew C. Harvey and Albert Jaeger. Detrending, stylized facts and the business cycle. *Journal of applied econometrics*, 8(3):231–247, 1993.

[82] S. Haug, C. Klüppelberg, A. Lindner, and M. Zapp. Method of moment estimation in the COGARCH(1,1) model. *The Econometrics Journal*, 10(2):320–341, 2007. ISSN 1368-4221, 1368-423X. doi: 10.1111/j.1368-423X.2007.00210.x.

[83] J. Hoffmann-Jørgensen and G. Pisier. The Law of Large Numbers and the Central Limit Theorem in Banach Spaces. *The Annals of Probability*, 4(4):587–599, 1976. ISSN 0091-1798.

[84] Jørgen Hoffmann-Jørgensen. Sums of independent Banach space valued random variables. *Studia Mathematica*, 52(2):159–186, 1974. ISSN 0039-3223.

[85] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. Analysis in Banach spaces. Vol. II, volume 67 of. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 3.

[86] T. Hytönen, J. Van Neerven, M. Veraar, and L. Weis. *Analysis in Banach Spaces, Vol. I, A Series of Modern Surveys in Mathematics 63*. Springer, Cham, 2016.

[87] K. Ito and Makiko Nisio. On stationary solutions of a stochastic differential equation. *J. Math. Kyoto Univ*, 4(1):1–75, 1964.

[88] Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido. Affine Volterra processes. *arXiv:1708.08796*, 2017.

[89] Jean Jacod and Albert Shiryaev. *Limit Theorems For Stochastic Processes*. Springer, 2nd ed edition, 2013.

[90] I. M. Johnstone and D. Paul. PCA in High Dimensions: An Orientation. *Proceedings of the IEEE*, 106(8):1277–1292, August 2018. ISSN 0018-9219. doi: 10.1109/JPROC.2018.2846730.

[91] Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327, 2001.

[92] Iain M. Johnstone and Arthur Yu Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486):682–693, 2009.

[93] Sungkyu Jung and J. Stephen Marron. PCA consistency in high dimension, low sample size context. *The Annals of Statistics*, 37(6B):4104–4130, 2009.

[94] Olav Kallenberg. *Foundations of Modern Probability*. Springer Science & Business Media, 2006.

[95] Jan Kallsen and Murad S. Taqqu. Option Pricing in ARCH-type Models. *Mathematical Finance*, 8(1):13–26, 1998.

[96] Jan Kallsen and Bernhard Vesenmayer. COGARCH as a continuous-time limit of GARCH(1,1). *Stochastic Processes and their Applications*, 119(1):74–98, 2009.

[97] Noureddine El Karoui. On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity. *arXiv preprint math/0309355*, 2003.

[98] Achim Klenke. *Probability Theory: A Comprehensive Course*. Springer Science & Business Media, 2013. ISBN 978-1-4471-5361-0.

[99] Claudia Klüppelberg, Andreas E. Kyprianou, and Ross A. Maller. Ruin probabilities and overshoots for general Lévy insurance risk processes. *Ann. Appl. Probab.*, 14(4):1766–1801, November 2004. ISSN 1050-5164, 2168-8737. doi: 10.1214/105051604000000927.

[100] Claudia Klüppelberg, Alexander Lindner, and Ross A. Maller. A continuous-time GARCH process driven by a Lévy process stationarity and second-order behaviour. *Journal of Applied Probability*, 2004.

[101] Claudia Klüppelberg, Alexander Lindner, and Ross Maller. Continuous time volatility modelling: COGARCH versus Ornstein–Uhlenbeck models. In *From Stochastic Calculus to Mathematical Finance*, pages 393–419. Springer, 2006.

[102] Claudia Klüppelberg, Ross Maller, and Alexander Szimayer. The COGARCH: A review, with news on option pricing and statistical inference. 2010.

[103] Uwe Küchler and Eckhard Platen. Weak discrete time approximation of stochastic differential equations with time delay. *Mathematics and Computers in Simulation*, 59(6):497–507, 2002. ISSN 0378-4754. doi: 10.1016/S0378-4754(01)00431-1.

[104] Clifford Lam and Qiwei Yao. Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics*, 40(2):694–726, 2012.

[105] Clifford Lam, Qiwei Yao, and Neil Bathia. Estimation of latent factors for high-dimensional time series. *Biometrika*, 98(4):901–918, 2011.

[106] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes*. Springer Science & Business Media, 2013.

[107] Paul Lescot and Michael Röckner. Perturbations of Generalized Mehler Semigroups and Applications to Stochastic Heat Equations with Levy Noise and Singular Drift. *Potential Analysis*, 20(4):317–344, June 2004. ISSN 0926-2601. doi: 10.1023/B:POTA.0000009814.54278.34.

[108] Daniel Li and Hervé Queffélec. *Introduction to Banach Spaces: Analysis and Probability*, volume 1. Cambridge University Press, 2017.

[109] Jia Li, Viktor Todorov, and George Tauchen. Adaptive estimation of continuous-time regression models using high-frequency data. *Journal of Econometrics*, 200(1):36–47, 2017.

[110] Zeng Li, Guangming Pan, and Jianfeng Yao. On singular value distribution of large-dimensional autocovariance matrices. *Journal of Multivariate Analysis*, 137: 119–140, 2015.

[111] Zeng Li, Qinwen Wang, and Jianfeng Yao. Identifying the number of factors from singular values of a large sample auto-covariance matrix. *The Annals of Statistics*, 45(1):257–288, 2017.

[112] Ming Liao. Lévy processes and Fourier analysis on compact Lie groups. *Annals of probability*, pages 1553–1573, 2004.

[113] Ming Liao. *Lévy Processes in Lie Groups*, volume 162. Cambridge university press, 2004.

[114] Werner Linde. *Probability in Banach Spaces: Stable and Infinitely Divisible Distributions*. Wiley, 1986.

[115] Alexander Lindner and Ross Maller. Lévy integrals and the stationarity of generalised Ornstein–Uhlenbeck processes. *Stochastic Processes and their Applications*, 115(10): 1701–1722, October 2005. ISSN 0304-4149. doi: 10.1016/j.spa.2005.05.004.

[116] Alexander M. Lindner. Stationarity, mixing, distributional properties and moments of GARCH (p, q)–processes. In *Handbook of Financial Time Series*, pages 43–69. Springer, 2009.

[117] Arne Lokka, Bernt Oksendal, and Frank Proske. Stochastic partial differential equations driven by Levy space-time white noise. *The Annals of Applied Probability*, 14(3):1506–1528, 2004. ISSN 1050-5164. doi: 10.1214/105051604000000413.

[118] Robert Lorenz. *Weak Approximation of Stochastic Delay Differential Equations with Bounded Memory by Discrete Time Series*. Phd Thesis, 2006.

[119] Kevin Weichao Lu. Weak Subordination of Multivariate Lévy Processes. 2018.

[120] Elisa Luciano and Wim Schoutens. A multivariate jump-driven financial asset model. *Quantitative finance*, 6(5):385–402, 2006.

[121] Elisa Luciano and Patrizia Semeraro. Multivariate time changes for Lévy asset models: Characterization and calibration. *Journal of Computational and Applied Mathematics*, 233(8):1937–1953, 2010.

[122] Dilip B. Madan and Eugene Seneta. The Variance Gamma (V.G.) Model for Share Market Returns. *The Journal of Business*, 63(4):511–524, 1990. ISSN 0021-9398.

[123] Dilip B. Madan, Peter P. Carr, and Eric C. Chang. The variance gamma process and option pricing. *Review of Finance*, 2(1):79–105, 1998.

[124] Ross A. Maller, Gernot Müller, and Alex Szimayer. GARCH modelling in continuous time for irregularly spaced time series data. *Bernoulli*, 14(2):519–542, 2008.

[125] Carlo Marinelli. Well-posedness and invariant measures for HJM models with deterministic volatility and Lévy noise. *Quantitative Finance*, 10(1):39–47, 2010.

[126] Tina Marquardt. Fractional Lévy processes with an application to long memory moving average processes. *Bernoulli*, 12(6), 2006.

[127] Nour Meddahi, Eric Renault, and Bas Werker. GARCH and irregularly spaced data. *Economics Letters*, 90(2):200–204, 2006.

[128] Markus Michaelsen. Information flow dependence in financial markets. *Available at SSRN 3051180*, 2018.

[129] Markus Michaelsen and Alexander Szimayer. Marginal consistent dependence modelling using weak subordination for Brownian motions. *Quantitative Finance*, 18(11): 1909–1925, 2018.

[130] Thomas Mikosch and Cătălin Stărică. Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects. *Review of Economics and Statistics*, 86(1):378–390, February 2004. ISSN 0034-6535, 1530-9142. doi: 10.1162/003465304323023886.

[131] Gernot Müller. MCMC Estimation of the COGARCH(1,1) Model. *Journal of Financial Econometrics*, 8(4):481–510, October 2010. ISSN 1479-8409. doi: 10.1093/jjfinec/nbq029.

[132] Daniel B. Nelson. ARCH models as diffusion approximations. *Journal of Econometrics*, 45(1):7–38, 1990.

[133] Jiazhu Pan and Qiwei Yao. Modelling multiple time series via common factors. *Biometrika*, 95(2):365–379, 2008.

[134] Kalyanapuram Rangachari Parthasarathy. *Probability Measures on Metric Spaces*, volume 352. American Mathematical Soc., 2005.

[135] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[136] Sandrine Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *arXiv:math/0411487*, July 2005.

[137] Daniel Pena and George EP Box. Identifying a simplifying structure in time series. *Journal of the American statistical Association*, 82(399):836–843, 1987.

[138] Víctor Pérez-Abreu and Alfonso Rocha-Arteaga. Lévy Processes in Cones of Banach Spaces. November 2003.

[139] Victor Perez-Abreu and Alfonso Rocha-Arteaga. Lévy processes in Banach spaces: Distributional properties and subordination. *Contemporary Mathematics*, 336:225–236, 2003.

[140] Víctor Pérez-Abreu and Alfonso Rocha-Arteaga. Covariance-parameter lévy processes in the space of trace-class operators. *Infin. Dimens. Anal. Quantum. Probab. Relat. Top.*, 08(01):33–54, March 2005. ISSN 0219-0257. doi: 10.1142/S0219025705001846.

[141] Víctor Pérez-Abreu and Alfonso Rocha-Arteaga. COVARIANCE-PARAMETER LÉVY PROCESSES IN THE SPACE OF TRACE-CLASS OPERATORS. *Infin. Dimens. Anal. Quantum. Probab. Relat. Top.*, 08(01):33–54, March 2005. ISSN 0219-0257, 1793-6306. doi: 10.1142/S0219025705001846.

[142] S. Peszat and J. Zabczyk. *Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach*. Cambridge University Press, October 2007. ISBN 978-0-521-87989-7.

[143] Mark Podolskij. Ambit fields survey and new challenges. In *XI Symposium on Probability and Stochastic Processes*. Springer, 2015.

[144] Enrico Priola and Jerzy Zabczyk. Structural properties of semilinear SPDEs driven by cylindrical stable processes. *Probability Theory and Related Fields*, 149(1-2): 97–137, February 2011. ISSN 0178-8051, 1432-2064. doi: 10.1007/s00440-009-0243-5.

[145] Philip E. Protter. *Stochastic Integration and Differential Equations*. Number 21 in Applications of Mathematics. Springer, Berlin ; New York, 2nd ed edition, 2004. ISBN 978-3-540-00313-7.

[146] Michael Reed and Barry Simon. *Methods of Mathematical Physics I: Functional Analysis*. Academic Press New York, 1972.

[147] M. Reiß, M. Riedle, and O. van Gaans. Delay differential equations driven by Lévy processes: Stationarity and Feller properties. *Stochastic Processes and their Applications*, 116(10):1409–1432, 2006. ISSN 03044149. doi: 10.1016/j.spa.2006.03.002.

[148] Markus Riedle. Ornstein-Uhlenbeck Processes Driven by Cylindrical Lévy Processes. *Potential Analysis*, 42(4):809–838, May 2015. ISSN 0926-2601, 1572-929X. doi: 10.1007/s11118-014-9458-x.

[149] John Riordan. *Introduction to Combinatorial Analysis*. Courier Corporation, 2012.

[150] Alfonso Rocha-Arteaga. Subordinators in a Class of Banach Spaces. *Random Operators and Stochastic Equations rose*, 14(3):245–258, 2006. doi: 10.1515/156939706778239855.

[151] Gennady Samoradnitsky. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Routledge, 2017.

[152] Ken-iti Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge university press, 1999.

[153] Patrizia Semeraro. A multivariate variance gamma model for financial applications. *International journal of theoretical and applied finance*, 11(01):1–18, 2008.

[154] Robert H. Shumway and David S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer, 2017.

[155] Robert Stelzer. Multivariate COGARCH(1, 1) processes. *Bernoulli*, 16(1):80–115, 2010. ISSN 1350-7265. doi: 10.3150/09-BEJ196.

[156] James H. Stock and Mark Watson. Dynamic factor models. *Oxford Handbooks Online*, 2011.

[157] James H. Stock and Mark W. Watson. Diffusion indexes. *NBER working paper*, (w6702), 1998.

[158] Alex Szimayer and Ross A. Maller. Finite approximation schemes for Lévy processes, and their application to optimal stopping problems. *Stochastic Processes and their Applications*, 117(10):1422–1447, October 2007. ISSN 03044149. doi: 10.1016/j.spa.2007.01.012.

[159] Viktor Todorov. Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *Journal of Econometrics*, 148(2):131–148, 2009.

[160] Cuong Viet Tran. *Convergence of Time Series Processes to Continuous Time Limits*. Phd Thesis, University of New South Wales, 2013.

[161] Qinwen Wang and Jianfeng Yao. On singular values distribution of a large auto-covariance matrix in the ultra-dimensional regime. *Random Matrices: Theory Appl.*, 04(04):1550015, October 2015. ISSN 2010-3263, 2010-3271. doi: 10.1142/S201032631550015X.

[162] Qinwen Wang and Jianfeng Yao. Moment approach for singular values distribution of a large auto-covariance matrix. *Ann. Inst. H. Poincaré Probab. Statist.*, 52(4):1641–1666, November 2016. ISSN 0246-0203. doi: 10.1214/15-AIHP693.

[163] Qinwen Wang and Jianfeng Yao. Extreme eigenvalues of large-dimensional spiked Fisher matrices with application. *The Annals of Statistics*, 45(1):415–460, 2017.

[164] Ran Wang and Lihu Xu. Asymptotics for stochastic reaction–diffusion equation driven by subordinate Brownian motion. *Stochastic Processes and their Applications*, 128(5):1772–1796, May 2018. ISSN 03044149. doi: 10.1016/j.spa.2017.08.010.

[165] Weichen Wang and Jianqing Fan. Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Ann Stat*, 45(3):1342–1374, June 2017. ISSN 0090-5364. doi: 10.1214/16-AOS1487.

[166] Yazhen Wang. Asymptotic nonequivalence of GARCH models and diffusions. *Ann. Statist.*, 30(3):754–783, June 2002. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1028674841.