Econometric Theory, **0**, 2026, 1–45.
doi:10.1017/S0266466625100261

STABILITY OF LINEAR MODELS UNDER TEMPORAL AGGREGATION

STOYAN V. STOYANOV
Charles Schwab Investment Management

Linear factor models dominate the field of empirical asset pricing and are largely considered a successful framework even though testing for model adequacy is uncommon in empirical research. This article develops a probabilistic argument under the assumption that return data are generated by a multivariate Lévy process and shows that linear factor models are asymptotically adequate as the return frequency declines. Rate-of-convergence results are provided assuming finite or infinite second moments and the moment conditions that determine the convergence rate are analyzed. Separately, we consider the combination of cross-sectional and temporal aggregation, which concerns portfolio applications, and we find that the model error of the linear framework can be smaller for portfolios than individual stocks. Although the intended application is finance, the results in the article are valid in other areas where data aggregation is natural.

1. INTRODUCTION

The vast majority of the empirical asset pricing literature works with standard time-series linear factor models in which the dependent variable represents the excess return of a stock or a portfolio and the independent variables represent spreads of certain factor portfolios (Fama and French, 1992, 2015, 2020; Carhart, 1997). There is a large strand of literature on identification of factors (Harvey, Liu, and Zhu, 2016; Hou, Xue, and Zhang, 2020) and the econometric problem of estimation and testing of linear factor models (Jagannathan and Wang, 1996; Feng, Giglio, and Xiu, 2020). Even though the linear time-series regression model is typically not tested for adequacy in this strand of literature, it has become a de facto standard and is widely regarded as a useful framework.

In this article, an asymptotic probabilistic argument is developed which can explain why linear models can be adequate in time-series regressions of variables that represent financial returns. The argument is based on the idea that lower frequency logarithmic returns can be viewed as temporally aggregated higher frequency returns. In particular, we prove that if the data generating process belongs to the class of Lévy processes—a fundamental building block in the field of finance (Rachev and Mittnik, 2000; Cont and Tankov, 2003)—the corresponding

I would like to thank the Editor (Peter C. B. Phillips), the Co-Editor (Matias D. Cattaneo), and two anonymous referees for the constructive comments and suggestions, which improved the article. All opinions and analyses are those of the author and do not reflect those of Charles Schwab Investment Management, Inc. (CSIM) or its affiliates. Address correspondence to Stoyan V. Stoyanov, Quantitative Innovation, Charles Schwab Investment Management, United States, e-mail: stoyan.stoyanov@schwab.com.

© The Author(s), 2026. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

conditional expectation converges to a linear function as the return frequency declines. Furthermore, the rate of convergence is proved to be bounded by a power law with an exponent determined by certain joint moment conditions with econometric content.

The main result is presented first assuming bounded second moments justified by the theoretical and practical significance of this assumption. We demonstrate that the moment conditions determining the exponent of the power law are related to the first-order conditions of a specific least-squares problem. In addition, we show that when the dependent variable is a portfolio return, the rate of convergence can benefit from the cross-sectional aggregation of the individual stock returns. That is, linear models can be more adequate on a relative basis for portfolios rather than individual stocks for a given return frequency under the assumptions in the article. This is an important insight of both theoretical and practical significance for applications in finance.

Relaxing the assumptions of finite second moments leads to a number of difficulties mainly because the conditional expectation of the aggregated process may converge to a nonlinear function. We develop two rate-of-convergence results in this setting. First, the finite variance result is extended assuming conditions that guarantee linearity of the conditional expectation of the corresponding limiting stable law. In this formulation, the stable law appears explicitly in the new moment conditions and its tail index affects the rate of convergence in the limit theorem. Second, these assumptions are relaxed and the rate of convergence to the corresponding nonlinear function is derived. In this general situation, the associated moment conditions are very similar to the ones behind modern rate-of-convergence results in the generalized central limit theorem (GCLT).

The approach developed in the article connects two unrelated fields. On the one hand, it is related to model adequacy testing in the econometric literature because we take advantage of some aspects of the fundamental work by Bierens (1982, 2016). Even though constructing an adequacy test is not a focus of this work, effectively new functionals measuring neglected nonlinearity are developed, which are also suitable for studying a rate-of-convergence problem. Separately, the results in this article can be viewed as an analytic confirmation of the observation made by Granger and Lee (1999) in a Monte Carlo study that linear models can become more adequate on a relative basis for aggregated data.

On the other hand, our approach is related to the probability metrics literature. From the point of view of probability theory, determining the rate of convergence in a limit theorem is best handled by functionals with certain metric properties. The metric approach has been used to address such questions in the central limit theorem (CLT) and the GCLT (Zolotarev, 2011; Rachev et al., 2013). The functionals constructed in this article borrow from that theory to establish the general stability of linear models and their domain of attraction in the context of Lévy processes. More generally, the main results can be restated without modification for the classical GCLT scheme of sums of independent and identically distributed (i.i.d.) random vectors.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

An unexpected result in the article is that the moment conditions determining the rate of convergence do not include the marginals of the joint multivariate process. This is in contrast to similar moment conditions in the GCLT setting. An explanation for the weaker form of these conditions is that the research problem involves distances between conditional expectations rather than probability laws. As a consequence of the weaker conditions, the rate of convergence to a linear conditional mean can be faster than the rate of convergence in the associated GCLT indicating that the results in this article can be viewed as a separate class of limit theorems specialized for conditional means.

Even though the intended application is finance, the results in the article can be relevant for other domains where temporal and cross-sectional data aggregation are natural. For instance, important variables in macroeconomics, such as production, profits, employment, or consumption, are measured as sums of quantities observed for firms or families, which is an example of cross-sectional aggregation (Granger, 1987). Some variables in economics are sums of flows over time and are sampled systematically at lower time frequencies leading to temporal aggregation. Examples include stock variables such as wealth and flow variables such as sales or production (Granger and Lee, 1999). Apart from the field of finance, linear models play a fundamental role in economics as well (Stock and Watson, 2011) and their adequacy may be affected by the degree of data aggregation.

The article is organized in the following way. Section 2 defines the problem, clarifies the connection with model adequacy testing in the econometric literature, and develops a motivating example. Section 3 establishes a rate-of-convergence result assuming the existence of second moments and discusses the combination of cross-sectional and temporal aggregation. Section 4 develops the infinite variance case and comments on the connection with the classical rate-of-convergence results in the GCLT setting. Section 5 provides numerical examples and discusses further extensions of the theory. Finally, Section 6 concludes. The proofs of all lemmas and theorems are provided in the Appendix.

2. MEASURING NEGLECTED NONLINEARITY

In the empirical asset pricing literature, it is common to estimate and test asset pricing models using a linear time-series regression in which the dependent variable $Y_t \in \mathbb{R}$ is represented by the return of a portfolio of risky assets and the independent variables $X_t \in \mathbb{R}^d$ represent returns of factor replicating portfolios. It is common to use monthly returns for model estimation.

Logarithmic returns have a temporal aggregation property. Monthly log-returns can be computed as a sum of daily log-returns and, in this sense, they can be viewed as temporally aggregated daily returns. More generally, log-returns at any frequency can be represented as temporally aggregated higher frequency returns at some fixed frequency of observation. Even though not formally analyzed in this way, a decision to work with lower frequencies can be regarded as a decision to take advantage of temporal aggregation. In this article, we argue that the functional

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

form of the conditional expectation $\mathbb{E}[Y_t|X_t]$ can become asymptotically linear as the degree of temporal aggregation increases and we study the rate of convergence in the limit problem.

In a broader non-financial context, Granger and Lee (1999) argue that many series in economics may be generated at one time interval but observed at another, greater time interval. For example, many variables are systematically sampled quarterly and each quarterly data point is a sum of the observations generated within the quarter. This process leads to temporally aggregated data at the quarterly frequency. In a Monte Carlo simulation, the authors conclude that temporal aggregation can simplify nonlinearity.

To address questions about the asymptotic behavior of $\mathbb{E}[Y_t|X_t]$, we need a functional that meets the next two general conditions:

(a) The functional has the capacity to measure model misspecification.
(b) The functional is suitable to study rate-of-convergence problems.

Concerning the first condition, measuring neglected nonlinearity is a special case of the more general problem of measuring the degree of misspecification in nonlinear models. Fundamental results due to Bierens (1982) establish necessary and sufficient conditions for a model to be correctly specified. Consider the following specification:

$$Y_t = m(X_t, \theta) + \epsilon_t, \tag{1}$$

where θ is a vector of parameters, $m(x, \theta)$ is a nonlinear function, and ϵ_t is a residual. The model in (1) is said to be correctly specified if $\mathbb{E}[\epsilon_t|X_t] = 0$ almost surely, which guarantees that $\mathbb{E}[Y_t|X_t] = m(X_t, \theta)$ in the almost sure sense. The necessary and sufficient conditions are given by

$$\mathbb{E}\epsilon_t e^{iu^{\mathsf{T}}X_t} = 0, \quad \forall u \in \mathbb{R}^d, \tag{2}$$

where $i = \sqrt{-1}$ denotes the imaginary unit.

In the econometric literature, moment conditions of the type presented in equation (2) are the foundation of integrated conditional moment tests employed to test for model adequacy. Bierens and Ploberger (1997) develop asymptotic theory for tests based on the following functional:

$$\eta(\theta) = \int_K \left|\mathbb{E}\epsilon_t w(u^{\mathsf{T}}\Phi(X_t))\right|^2 d\mu(u), \tag{3}$$

where ϵ_t is the residual in (1), $w(z)\colon \mathbb{C} \to \mathbb{C}$ is a weighting function assumed to be analytic, Φ is a bounded one-to-one mapping, and $\mu(u)$ is a probability measure absolutely continuous with respect to the Lebesgue measure on a compact $K \subset \mathbb{R}^d$. A test for model adequacy is based on the result that if $\mathbb{E}[\epsilon_t w(u^{\mathsf{T}}\Phi(X_t))] = 0$ for some $\theta_0 \in \Theta$ and all $u \in \mathbb{R}^d$, or equivalently if $\eta(\theta_0) = 0$, then $\mathbb{E}[\epsilon_t|X_t] = 0$ almost surely. The conditions in (2) arise with the exponential weights $w(z) = e^{iz}$.

Other weighting functions have also been considered. If we generalize the notation to $w(X_t, u)$, then notable alternatives to the exponential function include

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

the indicator function $w(X_t, u) = 1_{\{X_t \leq u\}}$ proposed by Stute (1997) and other alternatives, such as $w(X_t, u) = \sin(u^\top X_t)$ or $w(X_t, u) = 1/(1 + \exp(c - u^\top X_t))$ suggested by Stinchcombe and White (1998), and $w(X_t, u) = 1_{\{\beta^\top X_t \leq u\}}$ suggested by Escanciano (2006) in which β is assumed to be on the unit sphere in $\mathbb{R}^d$ and the integration in equation (3) is extended appropriately. It is worth noting that different weighting function families lead to a different power of the test, as emphasized in the literature.

Regarding the two properties mentioned earlier, any member of the family in equation (3), possibly extended with a more general weighting function, meets the criterion in (a). Unfortunately, no member of the extended family appears suitable as far as the criterion in (b) is concerned, which we illustrate further below.

Since our research question involves aggregation across time, we need an assumption about the underlying stochastic process. We assume that the vector of the dependent and the independent variables $\{V_t = (Y_t, X_t), t > 0\}$ follows a Lévy process, which represents a common assumption in mathematical finance (Rachev and Mittnik, 2000; Cont and Tankov, 2003). More formally, V_t is a multivariate Lévy process if (Sato, 1999):

1. $V_0 = 0$ almost surely.
2. V_t has independent increments; that is, for any choice of n and $0 \leq t_0 < t_1 < \cdots < t_n$, the random vectors $V_{t_0}, V_{t_1} - V_{t_0}, \ldots, V_{t_n} - V_{t_{n-1}}$ are independent.
3. The distribution of $V_{s+t} - V_s$ does not depend on s.
4. The process is stochastically continuous and its sample paths are right-continuous with left limits.

Consider n intervals of equal fixed length Δt and calculate the increments of the process $\Delta V_k = V_{t_k} - V_{t_{k-1}}$, $k = 1, \ldots, n$, where $t_k - t_{k-1} = \Delta t$. From the standpoint of applications, $n\Delta t$ can be interpreted as the sampling frequency while the fixed Δt stands for the frequency at which the series is generated. The process at time $n\Delta t$ can be represented as a sum of increments

$$V_{n\Delta t} = (Y_{n\Delta t}, X_{n\Delta t}) = \left(\sum_{k=1}^{n} \Delta Y_k, \sum_{k=1}^{n} \Delta X_k \right),$$

which are i.i.d. Even though the conditional mean $\mathbb{E}[Y_{\Delta t}|X_{\Delta t}]$ can be arbitrary, in this article, we show that the conditional mean $\mathbb{E}[Y_{n\Delta t}|X_{n\Delta t}]$ is approximately linear for large n.[1]

Next, we demonstrate that the expectation in equation (2) indeed decays for the aggregated process where, without loss of generality, we assume that $\mathbb{E}(V_t) = 0$.

[1] From a practical perspective, suppose the frequency of data generation Δt equals one day. The conditional mean of the daily series can be arbitrary. Also, suppose that the data are sampled systematically at progressively lower frequencies of $n\Delta t$, for example, monthly, quarterly, etc. In each larger time period, the observations are computed by summing up the daily data observed within it. The article shows that the degree of nonlinearity in the conditional mean of the lower-frequency data declines with n. As far as the stochastic process is concerned, we do not consider the underlying paths defined by the cumulative sum of increments within each larger interval but only the terminal values of the sum of increments.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

We note that the question for the asymptotic behavior of the conditional mean of $V_{n\Delta t}$ makes sense only if the limiting distribution of the random vector is non-degenerate, and to this end, the corresponding sums need to be normalized. For this demonstration, we assume that all second-order moments of V_t are bounded. The needed normalization is then provided by $n^{-1/2}$. Additionally, the least-squares estimator denoted by

$$\theta^* = (\mathbb{E}[X_{\Delta t}X_{\Delta t}^{\mathsf{T}}])^{-1}\mathbb{E}[Y_{\Delta t}X_{\Delta t}] \tag{4}$$

is assumed to exist, and the expectation in equation (2) corresponding to a linear specification $m(X_{n\Delta t}, \theta) = X_{n\Delta t}^{\mathsf{T}}\theta$ as a function of n is denoted by

$$\varpi_n(u) = \mathbb{E}\left(n^{-1/2}(Y_{n\Delta t} - X_{n\Delta t}^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}n^{-1/2}X_{n\Delta t}}\right)$$

after normalizing $V_{n\Delta t}$ by $n^{-1/2}$. If the linear specification is asymptotically correct, then $\varpi_n(u)$ with $\theta = \theta^*$ must decay for all u.

By the independence of the increments, θ^* is the same for each individual increment and also for their sum. The expectation in $\varpi_n(u)$ can be decomposed in the following manner:

$$\begin{aligned}
\varpi_n(u) &= \mathbb{E}\left(n^{-1/2}(Y_{n\Delta t} - X_{n\Delta t}^{\mathsf{T}}\theta^*)e^{iu^{\mathsf{T}}n^{-1/2}X_{n\Delta t}}\right) \\
&= n^{-1/2}\sum_{k=1}^{n}\mathbb{E}\left[(\Delta Y_k - \Delta X_k^{\mathsf{T}}\theta^*)e^{i(n^{-1/2}u)^{\mathsf{T}}\sum_{j=1}^{n}\Delta X_j}\right] \\
&= n^{-1/2}\sum_{k=1}^{n}\mathbb{E}\left[(\Delta Y_k - \Delta X_k^{\mathsf{T}}\theta^*)e^{i(n^{-1/2}u)^{\mathsf{T}}\Delta X_k}\right]\mathbb{E}\left[e^{i(n^{-1/2}u)^{\mathsf{T}}\sum_{j\neq k}\Delta X_j}\right] \\
&= n^{1/2}\varpi_1(n^{-1/2}u)\left[\phi_{X_{\Delta t}}(n^{-1/2}u)\right]^{n-1},
\end{aligned} \tag{5}$$

where $\phi_{X_{\Delta t}}(u)$ is the characteristic function (ch.f.) of $X_{\Delta t}$. Equation (5) indicates that the rate of decay of the left side for a fixed u is essentially determined by the rate of decay of $\sqrt{n}\varpi_1(u/\sqrt{n})$ because under the assumption of finite variance

$$\left[\phi_{X_{\Delta t}}(n^{-1/2}u)\right]^{n-1} \to e^{-\frac{1}{2}u^{\mathsf{T}}\mathbf{C}u},$$

where $\mathbf{C} = \mathbb{E}[X_{\Delta t}X_{\Delta t}^{\mathsf{T}}]$. Thus, to understand the behavior of the right side of (5), we employ the Taylor series expansion at $u = 0$,

$$\begin{aligned}
n^{1/2}\varpi_1(n^{-1/2}u) &= n^{1/2}\varpi_1(0) + \sum_{j=1}^{d}\frac{\partial\varpi_1(0)}{\partial u_j}u_j + \mathrm{o}(n^{-1/2}\left\|u\right\|_2) \\
&= n^{1/2}\mathbb{E}[\epsilon_{\Delta t}] + iu^{\mathsf{T}}\mathbb{E}[\epsilon_{\Delta t}X_{\Delta t}] + \mathrm{o}(n^{-1/2}\left\|u\right\|_2) \\
&= \mathrm{o}(n^{-1/2}\left\|u\right\|_2),
\end{aligned} \tag{6}$$

where the last equality holds because $\mathbb{E}[\epsilon_{\Delta t}] = 0$ and $\mathbb{E}[\epsilon_{\Delta t}X_{\Delta t}] = 0$ from the first-order conditions of the least-squares problem. The required smoothness of $\varpi_1(u)$ follows from the assumed existence of the second moment of the process V_t.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

Indeed, $\varpi_n(u)$ converges to zero asymptotically for any fixed $u \in \mathbb{R}^d$ by a classical CLT argument and, as a result, $\mathbb{E}[Y_{n\Delta t}|X_{n\Delta t}]$ converges almost surely to a linear function. That is, the specification

$$Y_t = X_t^{\mathsf{T}}\theta + \epsilon_t$$

with $\theta = \theta^*$ becomes asymptotically correct for large t. Concerning the choice of functional mentioned earlier, consider again $\eta(\theta)$ in equation (3). In view of the steps needed to derive (5), clearly the only possible choice of weighting function is the exponential function because the factorization property used in the derivation identifies it uniquely.

The approach developed in this article is based on functionals incorporating the moment conditions in (2) while at the same time staying close to the structure of ideal probability metrics developed exclusively for the study of rate-of-convergence problems in CLT-type theorems (Zolotarev, 2011; Rachev et al., 2013). The rate of decay is established for the following functional:

$$\boldsymbol{\varsigma}(V_t) = \inf_{\theta} \sup_{u\in\mathbb{R}^d} \psi(u) \left|\mathbb{E}(Y_t - X_t^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}X_t}\right|, \tag{7}$$

in which $\psi(u)$ is a ch.f. of a symmetric random vector with a sufficiently smooth density. It is straightforward to show that $\boldsymbol{\varsigma}(V_t) = 0$ implies $\mathbb{E}(Y_t|X_t)$ is linear almost surely and vice versa.[2] The role of the function ψ is to emphasize the behavior of the expectation at the origin and help establish upper bounds instrumental in the proofs. The structure of $\boldsymbol{\varsigma}$ bears similarity to the so-called smoothed metrics in probability theory.

The finite-variance case, which is consistent with the setup behind (5) and (6), is presented in Section 3. Effectively, our strategy is to establish an upper bound of $\boldsymbol{\varsigma}$ in terms of another functional which meets the condition in (b). A suitable choice turns out to be

$$\boldsymbol{b}_r(V_t) = \sup_{u\in\mathbb{R}^d} \|u\|_1^{-r} \left|\mathbb{E}(Y_t - X_t^{\mathsf{T}}\theta^*)e^{iu^{\mathsf{T}}X_t}\right|, \tag{8}$$

where $r \geq 1$. Similar to $\boldsymbol{\varsigma}$ in equation (7), $\boldsymbol{b}_r$ measures neglected nonlinearity.[3] The same technique is applied to portfolios in Section 3.2, where the rate of convergence can benefit from the cross-sectional aggregation behind portfolio construction.

The additional properties of the functional $\boldsymbol{b}_r$, which make it suitable for the study of a rate-of-convergence problem in the current setup, can be summarized in the following way. If we take two increments of the process V_t, ΔV_1 and ΔV_2, over two non-overlapping time intervals, then

$$\boldsymbol{b}_r(c\Delta V_1) = |c|^{r+1}\boldsymbol{b}_r(\Delta V_1) \tag{9}$$

$$\boldsymbol{b}_r(\Delta V_1 + \Delta V_2) \leq \boldsymbol{b}_r(\Delta V_1) + \boldsymbol{b}_r(\Delta V_2), \tag{10}$$

[2] This statement is proved in Lemma A.1 in the Appendix.

[3] Proofs of these statements, as well as other properties, are provided in Lemmas A.2–A.5 in the Appendix.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

where c is a nonzero constant. These two conditions represent a partial answer to what properties one may look for in order to meet condition (b).[4]

Even though of limited practical importance for financial returns, relaxing the assumption of finite variance creates a number of challenges. A discussion is provided in Section 4. The most significant ones are: (i) the least-squares estimator in (4) is no longer meaningful and (ii) the functional in (8) is no longer bounded. The first challenge is overcome by considering directly the structure of the limit law which, in contrast to the finite-variance case, may not have a linear conditional mean. The second challenge is resolved by an appropriate modification of the functional in (8) drawing inspiration from the probability metrics literature. Even though Section 3 arises as a special case of Section 4, a separate treatment is justified first by its practical significance and, second, by the transparency of the arguments for finite variance.

Some of the results in the article are presented in a more compact form with the help of multi-indices. A multi-index ν is a vector of nonnegative integers $\nu = (\nu_1, \dots, \nu_d) \in \mathbb{N}_0^d$, where $\mathbb{N}_0 = \{0, 1, \dots\}$. Traditional notation involving multi-indices includes

$$|\nu| = \sum_{i=1}^{d} \nu_i, \qquad \nu! = \nu_1! \dots \nu_d!,$$

$$x^\nu = x_1^{\nu_1} x_2^{\nu_2} \dots x_d^{\nu_d}, \qquad \partial^\nu f(x) = \frac{\partial^{\nu_1 + \dots + \nu_d}}{\partial x_1^{\nu_1} \dots \partial x_d^{\nu_d}} f(x_1, \dots, x_d).$$

Finally, we take advantage of the standard norm notation $\|x\|_1 = \sum_{j=1}^d |x_j|$, $\|x\|_2^2 = \sum_{j=1}^d x_j^2$, and $\|x\|_\infty = \max_{1 \le j \le d} |x_j|$.

3. LINEAR MODEL STABILITY ASSUMING BOUNDED SECOND MOMENTS

In this section, we assume that the returns process $V_t = (Y_t, X_t) \in \mathbb{R} \times \mathbb{R}^d$ is a Lévy process whose increments have a finite second moment, $\mathbb{E}\,\|V_{\Delta t}\|_2^2 < \infty$, and that the least-squares estimator θ^* is well-defined. The main assumptions are stated explicitly then followed by interpretations. The main result is provided next with a number of corollaries summarizing special cases. From an application point of view, these results reflect a single risky asset driven by a set of d factors.

In the second part of this section, we consider $V_t = (Y_t, X_t) \in \mathbb{R}^m \times \mathbb{R}^d$ which concerns m risky assets driven by a set of d common factors. In this setup, we consider how the rate of convergence results may change for portfolios constructed

[4]Establishing an inequality of the type $\boldsymbol{\varsigma}(V_t) \le M \boldsymbol{b}_r(V_t)$, where $M > 0$ is a constant, should be accompanied by a demonstration that $\boldsymbol{b}_r$ is bounded. To this end, we consider a third functional $\boldsymbol{\zeta}_r$ introduced in Lemma A.2 in the Appendix. This construction is inspired by the structure of the zeta metric, see Zolotarev (2011, Sect. 1.4) and Rachev et al. (2013, Sect. 15.3). Lemma A.5 proves that $\boldsymbol{\zeta}_r$ measures neglected nonlinearity. Also, it is possible to show that $\boldsymbol{\zeta}_r$ has the properties in equations (9) and (10).

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

from the basic risky assets. From an econometric viewpoint, this analysis covers the combination of cross-sectional and temporal aggregation.

3.1. Aggregation across Time

Before we begin with the formal discussion, we show that linear conditional means have a simple stability property. In essence, if a Lévy process has an increment with a linear conditional mean then the conditional mean of any other increment is linear as well. This property does not require bounded second moments. It suffices to assume existence of the mean only.

THEOREM 1. *Suppose that $V_t = (Y_t, X_t)$ is a zero-mean Lévy process. If for some $t = \tau$, the conditional mean $\mathbb{E}(Y_\tau | X_\tau)$ is linear with a vector of parameters θ, then the conditional mean $\mathbb{E}(Y_t | X_t)$ is linear with the same vector of parameters for any $t \geq 0$.*

In the remaining part of this section, we show that, under certain assumptions, the conditional mean $\mathbb{E}(Y_t | X_t)$ of the process V_t becomes asymptotically linear if t/τ is sufficiently large regardless of the functional form of $\mathbb{E}(Y_\tau | X_\tau)$ and we estimate the rate of convergence. The main assumptions are listed below.

Assumption 1. The random vector $V_{\Delta t} = (Y_{\Delta t}, X_{\Delta t}) \in \mathbb{R} \times \mathbb{R}^d$ is an increment of a Lévy process with $\mathbb{E}(V_{\Delta t}) = 0$ and $\mathbb{E}\,\|V_{\Delta t}\|_2^2 < \infty$. The matrix $\mathbb{E}[X_{\Delta t} X_{\Delta t}^\intercal]$ is assumed invertible.

Assumption 2. The following moment conditions hold:

$$\mathbb{E}\left((Y_{\Delta t} - X_{\Delta t}^\intercal \theta^*) X_{\Delta t}^\nu\right) = 0, \quad \forall \nu : |\nu| \leq k \tag{11}$$

$$\mathbb{E}\left(|Y_{\Delta t} - X_{\Delta t}^\intercal \theta^*| \, \|X_{\Delta t}\|_1^r\right) < \infty, \tag{12}$$

where $k \geq 1$ is an integer, θ^* is provided in equation (4), ν denotes a multi-index, and $r = k + \delta$ with $\delta \in (0, 1]$.

The role of the assumptions can be explained in the following way. Assumption 1 defines the process and guarantees existence of θ^* while Assumption 2 ensures boundedness of the functional $\boldsymbol{b}_r$ defined in equation (8).[5]

Concerning the moment conditions in equation (11), note that $V_{\Delta t}$ is assumed to have a mean of zero in Assumption 1, which guarantees equation (11) holds for $k = 0$. From the definition of θ^* in equation (4), it follows directly that $\mathbb{E}[(Y_{\Delta t} - X_{\Delta t}^\intercal \theta^*) X_{\Delta t}] = 0$ and, therefore, equation (11) holds automatically for $k = 1$ as well. Thus, the moment conditions are needed only for $k \geq 2$.

The condition in equation (12) looks complicated but a simpler upper bound is easy to obtain. Indeed, the next inequality holds

$$\mathbb{E}\left(|Y_{\Delta t} - X_{\Delta t}^\intercal \theta^*| \, \|X_{\Delta t}\|_1^r\right) \leq \max(1, \left\|\theta^*\right\|_\infty) \mathbb{E}\,\|V_{\Delta t}\|_1^{r+1}. \tag{13}$$

[5] A proof is provided in Lemma A.3 in the Appendix.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

An advantage is that we can isolate a simple moment condition involving the vector $V_{\Delta t}$ separately from θ^*.

The main result is provided in the following theorem.

THEOREM 2. *Under Assumptions 1 and 2,*

$$\varsigma(n^{-1/2}V_{n\Delta t}) \leq C_r n^{1-\frac{r+1}{2}} \boldsymbol{b}_r(V_{\Delta t}), \tag{14}$$

where $r \geq 1$, $C_r = \sup_u \|u\|_1^r \psi(u)$. *Finally,* $\boldsymbol{b}_r(V_{\Delta t})$ *can be bounded in the following way:*

$$\boldsymbol{b}_r(V_{\Delta t}) \leq A_\delta \frac{\Gamma(\delta+1)}{\Gamma(r+1)} \mathbb{E}\left(|Y_{\Delta t} - X_{\Delta t}^\intercal \theta^*| \, \|X_{\Delta t}\|_1^r\right), \tag{15}$$

in which $r = k + \delta$, $k \geq 1$ *is an integer, and* $\delta \in (0,1]$, $A_\delta = \sup_{y\in\mathbb{R}} |y|^{-\delta}|e^{iy} - 1|$ *is an absolute constant, and* $\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du$, $\mathrm{Re}(x) > 0$, *denotes the gamma function.*

The constant C_r depends entirely on the choice of the function ψ and can be computed explicitly in some instances. For example, we prove in Lemma A.4 that if ψ is chosen to be the ch.f. of a d-dimensional Gaussian distribution with standardized independent components, $\psi(u) = \exp(-u^\intercal u/2)$, then $C_r = e^{-\frac{r}{2}}(rd)^{\frac{r}{2}}$. It is important to emphasize that the rate of convergence does not depend on the choice of ψ.

Next, we consider an important special case. If we set $r = 1 + \delta$ with $\delta \in (0,1]$, then, as explained previously, the moment conditions in equation (11) are satisfied automatically. We obtain the following corollary.

COROLLARY 1. *If Assumption 1 holds together with the moment condition in equation (12) with* $r = 1 + \delta$, $\delta \in (0,1]$, *then*

$$\varsigma(n^{-1/2}V_{n\Delta t}) \leq n^{-\delta/2} C_{1+\delta} \boldsymbol{b}_{1+\delta}(V_{\Delta t}).$$

The corollary illustrates that the conditions in Assumption 2 determine the rate of decay in Theorem 2. Clearly, the rate of decay increases with r. At the limit, if $\mathbb{E}\|V_{\Delta t}\|^r < \infty$ for all r, then, by (13), the condition in equation (12) holds for all r. In addition, if the conditions in equation (11) hold for all k, then for any analytic function $h\colon \mathbb{R}^d \to \mathbb{C}$, we have

$$\begin{aligned}
\mathbb{E}[(Y_{\Delta t} - X_{\Delta t}^\intercal \theta^*)h(X_{\Delta t})] &= \mathbb{E}\left[(Y_{\Delta t} - X_{\Delta t}^\intercal \theta^*) \sum_{|\nu|=0}^{\infty} \frac{\partial^\nu h(0)}{\nu!} X_{\Delta t}^\nu\right] \\
&= \sum_{|\nu|=0}^{\infty} \frac{\partial^\nu h(0)}{\nu!} \mathbb{E}\left[(Y_{\Delta t} - X_{\Delta t}^\intercal \theta^*) X_{\Delta t}^\nu\right] \\
&= 0.
\end{aligned}$$

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

The family of functions $\{h_u(x) = e^{iu^\mathsf{T}x}, u \in \mathbb{R}^d\}$ belongs to that class and, therefore, $\mathbb{E}[(Y_{\Delta t} - X_{\Delta t}^\mathsf{T}\theta^*)h_u(X_{\Delta t})] = \mathbb{E}[(Y_{\Delta t} - X_{\Delta t}^\mathsf{T}\theta^*)e^{iu^\mathsf{T}X_{\Delta t}}] = 0$ for all $u \in \mathbb{R}^d$, which means that $\mathbb{E}[Y_{\Delta t}|X_{\Delta t}] = X_{\Delta t}^\mathsf{T}\theta^*$ almost surely. By Lemma A.5, $\boldsymbol{b}_r(V_{n\Delta t}) = 0$ and the upper bound in equation (14) equals zero for any n.

The moment conditions in equation (11) permit an econometric interpretation. It is straightforward to notice that they represent first-order conditions of a least-squares problem. Consider, for example, the case $k = 2$. If we add the components of $X_{\Delta t}$ raised to the second power and all mixed products linearly, we obtain the following least-squares problem:

$$(\tilde{\theta}, \tilde{\beta}) = \arg\min_{(\theta,\beta)} \mathbb{E}\left(Y_{\Delta t} - X_{\Delta t}^\mathsf{T}\theta - \sum_{|\nu|=2} X_{\Delta t}^\nu \beta_\nu\right)^2,$$

where β is a vector containing all the β_ν parameters. If the optimal solution is $(\tilde{\theta}, \tilde{\beta}) = (\theta^*, 0)$, then the moment conditions in equation (11) hold with $k = 2$. Clearly, $\tilde{\beta} = 0$ does not necessarily imply a linear conditional mean but according to Theorem 2 it does imply a faster rate of convergence if the corresponding condition in equation (12) holds as well. The same analysis can be repeated when $k > 2$.

The rate of decay in Theorem 2 is also affected by the choice of δ. The conditions in equations (12) and (13) indicate that δ is determined by the tail behavior of $V_{\Delta t}$. For instance, suppose that the conditions in equation (11) hold for some fixed value of k and $\mathbb{E}\|V_{\Delta t}\|_1^{k+\delta+1} < \infty$ for all $0 < \delta < \delta^* \leq 1$ but $\mathbb{E}\|V_{\Delta t}\|_1^{k+\delta^*+1}$ is unbounded. This is possible for distributions with a Pareto-type tail, see Section 5.1.2 for an example. In this case, δ can be chosen to be close but below δ^* bearing in mind a potential trade-off—as δ approaches δ^* the rate of decay improves but the value of the scaling constant $\boldsymbol{b}_r(V_{\Delta t})$ may deteriorate. If the moment $\mathbb{E}\|V_{\Delta t}\|_1^{k+2}$ exists, then we can set $\delta = 1$. For example, if $\mathbb{E}\|V_{\Delta t}\|_1^3 < \infty$, then Corollary 1 holds with $\delta = 1$.

Another special case can be constructed by imposing additional properties on the random vector $V_{\Delta t}$. Suppose that $V_{\Delta t}$ is symmetric, $V_{\Delta t} \stackrel{d}{=} -V_{\Delta t}$. It is easy to see that the conditions in equation (11) hold for any even value of $|\nu|$. As a result, the conditional mean converges to a linear function faster for symmetric distributions.

COROLLARY 2. *If $V_{\Delta t} \stackrel{d}{=} -V_{\Delta t}$ and Assumption 1 holds together with the moment condition in equation (12) with $r = 2 + \delta$, $\delta \in (0, 1]$, then*

$$\zeta(n^{-1/2}V_{n\Delta t}) \leq n^{-\frac{1+\delta}{2}} C_{2+\delta}\boldsymbol{b}_{2+\delta}(V_{\Delta t}).$$

Modern CLT rate-of-convergence results for metrics with uniform structure include moment conditions similar to the ones in Assumption 2, see, for example, Rachev et al. (2013, Thms 15.4.2 and 15.4.3 and Remark 15.4.4) and Section 4.1 for a more detailed discussion. In terms of our notation, these moment conditions

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

take the following form:

$$\mathbb{E}(V_{\Delta t}^{\gamma} - Z^{\gamma}) = 0, \quad \forall \gamma : |\gamma| \leq k \tag{16}$$

$$\boldsymbol{\nu}_{r+1}(V_{\Delta t}, Z) = \int_{\mathbb{R}^{d+1}} \|\xi\|_1^{r+1} |F - \Phi|(d\xi) < \infty, \tag{17}$$

where $r = k + \delta$, $\delta \in (0, 1]$, γ is a multi-index, Z denotes a $(d+1)$-dimensional Gaussian vector, and F and Φ denote the distribution functions of $V_{\Delta t}$ and Z, respectively. The functional $\boldsymbol{\nu}_s$ is known as an absolute pseudomoment of order s and its role is illustrated in Section 4.1. Even though the left side of equation (11) for any fixed ν is a linear combination of some of the moments $\mathbb{E}[V_{\Delta t}^{\nu+1}]$, one key difference from equation (16) is that the conditions in Assumption 2 do not involve moments of the marginals of $V_{\Delta t}$ on a stand-alone basis. For this reason, the conditions in Assumption 2 are weaker than those in equation (16) and the rate of decay in Theorem 2 can be faster than the corresponding rate of convergence in the CLT implied by the conditions in equations (16) and (17).

A somewhat trivial illustration of this observation can be constructed in the following way. If we choose any multivariate law for $V_{\Delta t}$ with linear conditional expectations, then, by Theorem 1 and Lemma A.1, we have $\boldsymbol{\varsigma}(V_{n\Delta t}) = 0$ for all $n \geq 1$ regardless of the marginal distributions of $V_{\Delta t}$. At the same time, the conditions in equation (16) may not hold for any $k \geq 3$ and the rate of convergence in the CLT for the chosen multivariate law may be as slow as the classical rate of $n^{-1/2}$. Section 5.2 provides a non-trivial numerical example of this phenomenon.

3.2. Cross-Sectional Aggregation Combined with Temporal Aggregation

The simulation study by Granger and Lee (1999) concludes that if there are prominent factors with a nonlinear impact at micro-level, the nonlinear effects may persist in cross-sectionally aggregated data. In some applications, however, the aggregation process is applied jointly cross-sectionally and in the time dimension. In this section, we focus on the joint impact.

In finance, for example, portfolios are formed from individual stocks and portfolio returns can be viewed as cross-sectional weighted averages of stock returns while return frequency reflects temporal aggregation. In fact, factor models are commonly applied to portfolios rather than individual basic risky assets. In this section, we change the setting to accommodate linear combinations of Y variables interpreted as portfolios. We demonstrate that under fairly general conditions linear models can become more adequate for portfolios faster with n than for individual stocks when the data are temporally aggregated.

Consider a multivariate process $V_t = (Y_t, X_t) \in \mathbb{R}^m \times \mathbb{R}^d$, where m denotes the number of basic risky assets and d denotes the number of factors. The linear combination $Y_{p,t} = \sum_{i=1}^{m} w_i Y_{i,t}$ denotes a portfolio of the basic risky assets, where $w_i \geq 0$ stand for portfolio weights. Even though we focus on long-only portfolios, a similar analysis holds in the general case. Finally, denote by $\mathbf{T}_j V_t = (Y_{j,t}, X_t)$

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

an operator that selects the j-th component of Y_t together with all factors and let $\mathbf{T}_p V_t = (Y_{p,t}, X_t)$ be an operator that constructs the portfolio and selects all factors.

The two assumptions can be adapted in the following way. Assumption 1 holds for the whole vector V_t while Assumption 2 is modified to hold only for a specific portfolio.

Assumption 2*. The following moment conditions hold for $\mathbf{T}_p V_{\Delta t} = (Y_{p,\Delta t}, X_{\Delta t})$:

$$\mathbb{E}\left((Y_{p,\Delta t} - X_{\Delta t}^{\mathsf{T}}\theta_p^*)X_{\Delta t}^{\nu}\right) = 0, \quad \forall \nu\colon |\nu| \le k_p \tag{18}$$

$$\mathbb{E}\left(|Y_{p,\Delta t} - X_{\Delta t}^{\mathsf{T}}\theta_p^*|\,\|X_{\Delta t}\|_1^{r_p}\right) < \infty, \tag{19}$$

where $Y_{p,t} = \sum_{i=1}^m w_i Y_{i,t}$, $k_p \ge 1$ is an integer, ν denotes a multi-index, $r_p = k_p + \delta$ with $\delta \in (0,1]$, and θ_p^* is provided by

$$\theta_p^* = (\mathbb{E}[X_{\Delta t}X_{\Delta t}^{\mathsf{T}}])^{-1}\mathbb{E}[Y_{p,\Delta t}X_{\Delta t}].$$

The next corollary follows directly from Theorem 2.

COROLLARY 3. *Under Assumptions 1 and 2*,*

$$\boldsymbol{\varsigma}(n^{-1/2}\mathbf{T}_p V_{n\Delta t}) \le C_{r_p} n^{1-\frac{r_p+1}{2}} \boldsymbol{b}_{r_p}(\mathbf{T}_p V_{\Delta t}), \tag{20}$$

where $r_p \ge 1$.

Cross-sectional aggregation can improve the upper bound in (20) relative to the upper bounds of the individual assets when Corollary 3 is applied independently to $\mathbf{T}_j V_t$. For example, let us assume the worst-case scenario. If $\mathbb{E}\|V_{\Delta t}\|_1^{2+\delta} < \infty$, then (19) holds for any portfolio and any individual risky asset, respectively. As illustrated previously, the condition in (18) holds automatically for $k_p = 1$ for any portfolio as well. Therefore, the inequality in (20) takes the form

$$\begin{aligned}\boldsymbol{\varsigma}(n^{-1/2}\mathbf{T}_p V_{n\Delta t}) &\le C_{1+\delta} n^{-\frac{\delta}{2}} \boldsymbol{b}_{1+\delta}(\mathbf{T}_p V_{\Delta t}) \\ &\le C_{1+\delta} n^{-\frac{\delta}{2}} \sum_{j=1}^m w_j \boldsymbol{b}_{1+\delta}(\mathbf{T}_j V_{\Delta t}),\end{aligned}$$

where the second inequality follows easily from $\theta_p = \sum_{i=1}^m w_i \theta_i^*$ and the definition of $\boldsymbol{b}_r$. Even though in this case the rate of convergence is the same regardless of the weights, the corresponding upper bound for the portfolio is smaller than the average of the individual risky assets' upper bounds.

Next, suppose $\mathbb{E}\|V_{\Delta t}\|_1^{3+\delta} < \infty$. Also, assume the condition in equation (18) does not hold for $\nu\colon |\nu| > 1$ for any of the individual assets. The condition in (18) for any portfolio can be expressed through the weights and the individual assets in

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

the following way:

$$\mathbb{E}\left((Y_{p,\Delta t} - X_{\Delta t}^{\mathsf{T}}\theta_p^*)X_{\Delta t}^{\nu}\right) = \mathbb{E}\left[\left(\sum_{j=1}^{m} w_j Y_{j,\Delta t} - X_{\Delta t}^{\mathsf{T}}\left(\sum_{j=1}^{m} w_j \theta_j^*\right)\right) X_{\Delta t}^{\nu}\right]$$
$$= \sum_{j=1}^{m} w_j \mathbb{E}\left((Y_{j,\Delta t} - X_{\Delta t}^{\mathsf{T}}\theta_j^*)X_{\Delta t}^{\nu}\right).$$

Thus, the condition in (18) for $k_p = 2$ can be written as

$$\sum_{j=1}^{m} w_j \mathbb{E}\left((Y_{j,\Delta t} - X_{\Delta t}^{\mathsf{T}}\theta_j^*)X_{\Delta t}^{\nu}\right) = 0, \qquad \forall \nu\colon |\nu| = 2, \tag{21}$$

because it holds for $\nu\colon |\nu| = 1$ by assumption. The number of equations in (21) equals $d + d(d-1)/2$ and this number depends only on the dimension d of the vector X. All of the equations are linear with respect to the portfolio weights. Therefore, if

$$d + \frac{d(d-1)}{2} < m, \tag{22}$$

then there exists a portfolio (possibly a long-short one) for which (21) holds. As a consequence, for that portfolio $r_p = 2 + \delta$ and the inequality in (20) becomes

$$\varsigma(n^{-1/2}\mathbf{T}_p V_{n\Delta t}) \le C_{2+\delta} n^{-\frac{\delta+1}{2}} \boldsymbol{b}_{2+\delta}(\mathbf{T}_p V_{\Delta t}).$$

As a result, the rate of decay $(\delta+1)/2$ is strictly faster than the rate $\delta/2$ of the individual risky assets. Section 5.1.3 includes a simple illustration with the hyperbolic skewed-t distribution.

In the practice of finance, factor models are usually of low dimension while the cross section, depending on which universe is considered, typically contains a large number of stocks. Thus, the left side of the inequality in equation (22) is in practice much smaller than m. For example, $m = 1{,}000$ for the large cap Russell 1,000 universe, while for a ten-factor model, the left side equals 55.[6] In the academic literature, very popular factor models include the three-factor model of Fama and French (1992), the four-factor model of Carhart (1997), and the five-factor model of Fama and French (2015). Even though there is a debate about the dimension of the factor space, it is generally believed to be low (Feng et al., 2020).

4. THE CASE OF INFINITE VARIANCE

Dropping the assumption of finite second moment for the process $\{V_t, t > 0\}$ leads to a number of difficulties with the approach developed in Section 3. First, the normalization by $n^{-1/2}$ is no longer correct and, second, the least-squares solution

[6]Other examples include the commonly used CRSP data set (about 4,000 stocks) and the Russell 3,000 universe (1,000 large-cap and 2,000 small-cap stocks).

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

in equation (4) is irrelevant as it involves diverging quantities. Finally and perhaps most importantly, all moments involved in the conditions in Assumption 2 are unbounded. As a result, the functional $\boldsymbol{b}_r(V_{\Delta t})$ is unbounded as well for any $r \geq 1$ and the upper bound in Theorem 2 becomes impractical.

These observations indicate that the infinite variance case requires a different treatment. We adopt a strategy similar to that behind the rate-of-convergence theorems for the GCLT. In essence, we need new assumptions involving moment conditions defined on the difference of V_t and another process which appears as the corresponding limit under the GCLT. In this section, we take advantage of this approach and we generalize Theorem 2.

Denote by $Z_t = (Z_{y,t}, Z_{x,t}) \in \mathbb{R} \times \mathbb{R}^d$ a strictly α-stable Lévy process independent of V_t. The ch.f. of the increment $Z_1 - Z_0 = Z_1$ takes the form (Sato, 1999, Thm. 14.10)

$$\varphi(u) = \exp\left\{ - \int_{\mathbb{S}} |u^{\mathsf{T}}\xi|^{\alpha} \left(1 - i \operatorname{sign}(u^{\mathsf{T}}\xi) \tan\frac{\pi\alpha}{2}\right) \Lambda_1(d\xi) \right\}, \tag{23}$$

where $\alpha \in (0, 2]$ is the tail parameter, $\mathbb{S} = \{z \in \mathbb{R} \times \mathbb{R}^d \colon \|z\|_2 = 1\}$ denotes the unit sphere in $\mathbb{R}^{d+1}$, and Λ_1 denotes a finite nonzero measure defined on the Borel subsets of $\mathbb{S}$, also known as the spectral measure. We assume that $\mathbb{E}|Z_1| < \infty$ which implies that $1 < \alpha \leq 2$. Finally, strict stability in this case means $\mathbb{E}(Z_1) = 0$.

In the finite-variance case, the vector of parameters θ is determined by the least-squares solution. In the infinite variance case, if $V_{\Delta t}$ is in the domain of attraction of a stable law whose corresponding conditional mean is linear, then, by Theorem 1, the vector θ is determined by the spectral measure Λ_1. In Lemma A.6, we prove that if θ_α^* solves

$$\int_{\mathbb{S}} (y - x^{\mathsf{T}}\theta_\alpha^*) \operatorname{sign}(u^{\mathsf{T}}x) |u^{\mathsf{T}}x|^{\alpha-1} \Lambda_1(d\xi) = 0, \quad \forall u \in \mathbb{R}^d \tag{24}$$

$$\int_{\mathbb{S}} (y - x^{\mathsf{T}}\theta_\alpha^*) |u^{\mathsf{T}}x|^{\alpha-1} \Lambda_1(d\xi) = 0, \quad \forall u \in \mathbb{R}^d, \tag{25}$$

where $\xi = (y, x)$, then $\mathbb{E}(Z_{y,1}|Z_{x,1}) = Z_{x,1}^{\mathsf{T}}\theta_\alpha^*$ almost surely and vice versa. By Theorem 1, if θ_α^* solves the equations above, then $\mathbb{E}(Z_{y,t}|Z_{x,t}) = Z_{x,t}^{\mathsf{T}}\theta_\alpha^*$ almost surely for any t.

These conditions appear complicated and the following special cases provide some clarification. If Z_1 is symmetric, then Λ_1 is symmetric and the condition in equation (25) is satisfied automatically. The necessary and sufficient condition in equation (24) for the symmetric case is derived in Miller (1978). Even in the symmetric case, θ_α^* may not be determined uniquely, which is a clear difference from the finite-variance case. If $\alpha = 2$, then Z_1 follows a Gaussian distribution and we are in the finite-variance case. In this situation, equation (24) reduces to the least-squares case, that is, $\theta_2^* = \theta^*$.

We replace Assumptions 1 and 2 with the following ones.

Assumption 1†. The random vector $V_{\Delta t} = (Y_{\Delta t}, X_{\Delta t}) \in \mathbb{R} \times \mathbb{R}^d$ is an increment of a Lévy process with $\mathbb{E}(V_{\Delta t}) = 0$. The random vector $Z_{\Delta t} = (Z_{y,\Delta t}, Z_{x,\Delta t}) \in \mathbb{R} \times$

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

$\mathbb{R}^d$ is an increment of a strictly α-stable process with $\alpha \in (1,2]$ and there exists a vector of parameters θ_α^* satisfying equations (24) and (25).

Assumption 2†. The following moment conditions hold:

$$\int_{\mathbb{R}^{d+1}} (y - x^\mathsf{T}\theta_\alpha^*) x^\nu d(F-G) = 0, \quad \forall \nu \colon |\nu| \le k \tag{26}$$

$$\int_{\mathbb{R}^{d+1}} |y - x^\mathsf{T}\theta_\alpha^*| \, \|x\|_1^r \, |F-G|(d\xi) < \infty, \tag{27}$$

where $\xi = (y,x) \in \mathbb{R} \times \mathbb{R}^d$, F is the distribution function of $V_{\Delta t}$, G is the distribution function of $Z_{\Delta t}$, $k \ge 1$ is an integer, θ_α^* satisfies equations (24) and (25), ν denotes a multi-index, and $r = k + \delta$ with $\delta \in (0,1]$.

Finally, we need a version of $\boldsymbol{b}_r(V_t)$ defined in equation (8), which is well behaved in the infinite variance case. Consider

$$\boldsymbol{b}_r^\theta(V_t, Z_t) = \sup_u \|u\|_1^{-r} \left| \int_{\mathbb{R}^{d+1}} (y - x^\mathsf{T}\theta) e^{iu^\mathsf{T}x} d(F_{V_t} - G_{Z_t}) \right|, \tag{28}$$

where F_{V_t} and G_{Z_t} are the distribution functions of V_t and G_t, respectively. The important difference from $\boldsymbol{b}_r(V_t)$ is that the limit law is subtracted, which regularizes the behavior of the integral at $u = 0$. Lemma A.8 shows that if the conditions in Assumption 2† are satisfied for some θ and arbitrary $V_{\Delta t}$ and $Z_{\Delta t}$, then $\boldsymbol{b}_r^\theta(V_{\Delta t}, Z_{\Delta t}) < \infty$.

Similarly to the finite-variance case, we can find a simpler bound for equation (27). The same approach leads to

$$\begin{aligned} \int_{\mathbb{R}^{d+1}} |y - x^\mathsf{T}\theta_\alpha^*| \, \|x\|_1^r \, |F-G|(d\xi) &\le K \int_{\mathbb{R}^{d+1}} \|\xi\|_1^{r+1} \, |F-G|(d\xi) \\ &= K \boldsymbol{\nu}_{r+1}(V_{\Delta t}, Z_{\Delta t}), \end{aligned} \tag{29}$$

where $K = \max(1, \left\|\theta_\alpha^*\right\|_\infty)$. The inequality in (29) can be combined with Lemma A.8 to conclude that existence of the absolute pseudomoment of order $r+1$ between the increments $V_{\Delta t}$ and $Z_{\Delta t}$, together with the conditions in (26), guarantees boundedness of $\boldsymbol{b}_r^\theta(V_{\Delta t}, Z_{\Delta t})$.

A generalization of Theorem 2 is provided below.

THEOREM 3. *Under Assumptions 1† and 2†,*

$$\varsigma\left(n^{-1/\alpha} V_{n\Delta t}\right) \le C_r n^{1 - \frac{r+1}{\alpha}} \boldsymbol{b}_r^{\theta_\alpha^*}(V_{\Delta t}, Z_{\Delta t}), \tag{30}$$

where $r = k + \delta$, $k \ge 1$ is an integer, $\delta \in (0,1]$, and $C_r = \sup_u \|u\|_1^r \, \psi(u)$. Finally, the next bound holds

$$\boldsymbol{b}_r^{\theta_\alpha^*}(V_{\Delta t}, Z_{\Delta t}) \le A_\delta \frac{\Gamma(\delta+1)}{\Gamma(r+1)} \int_{\mathbb{R}^{d+1}} |y - x^\mathsf{T}\theta_\alpha^*| \, \|x\|_1^r \, |F-G|(d\xi),$$

in which $\xi = (y,x)$, $A_\delta = \sup_{y \in \mathbb{R}} |y|^{-\delta} |e^{iy} - 1|$ is an absolute constant, and $\Gamma(x)$ denotes the gamma function.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

Theorem 2 arises as a special case when $\alpha = 2$ in the following way. First, as noted earlier, the distribution of $Z_{\Delta t}$ in this case is Gaussian and the ch.f. in (23) turns into the Gaussian ch.f. In fact, Λ_1 is no longer unique and there are many symmetric choices for Λ_1 converting the ch.f. in equation (23) into the Gaussian ch.f. (Samorodnitsky and Taqqu, 1994, Sect. 2.4). As a consequence, the condition in (25) is satisfied due to symmetry considerations and $\theta_2^* = \mathbf{C}_{xx}^{-1}\mathbf{C}_{yx}$ is the unique solution of (24), where $\mathbf{C}$ is the covariance matrix of $Z_{\Delta t}$ and the indices indicate the blocks corresponding to the y and x components of the vector $(Z_{y,\Delta t}, Z_{x,\Delta t})$.

Second, the condition in equation (26) holds with $|\nu| = 1$ if $\mathbb{E}[Z_{\Delta t}] = 0$. It holds with $|\nu| = 2$ only if $\mathbf{C}$ coincides with the covariance matrix of the vector $V_{\Delta t}$. As a consequence, $\theta_2^* = \theta^*$, where θ^* is provided in equation (4). Separately, the vector $Z_{\Delta t}$ has linear conditional means and, therefore, $\mathbb{E}[(Z_{y,\Delta t} - Z_{x,\Delta t}^{\mathsf{T}}\theta^*)f(Z_{x,\Delta t})] = 0$ for any measurable function f for which the moment is defined. This shows that $\mathbb{E}[(Z_{y,\Delta t} - Z_{x,\Delta t}^{\mathsf{T}}\theta^*)Z_{x,\Delta t}^{\nu}] = 0$, $\forall \nu\colon |\nu| \le k$ and the condition in equation (26) turns into that in equation (11) after writing the integral as a difference of two expectations.

In a similar way, for the left side of (27), we obtain

$$|I_V - I_Z| \le \int_{\mathbb{R}^{d+1}} |y - x^{\mathsf{T}}\theta_\alpha^*|\,\|x\|_1^r\,|F - \Phi|(d\xi) \le I_V + I_Z,$$

where $I_V = \mathbb{E}(|Y_{\Delta t} - X_{\Delta t}^{\mathsf{T}}\theta^*|\,\|X_{\Delta t}\|_1^r)$, $I_Z = \mathbb{E}(|Z_{y,\Delta t} - Z_{x,\Delta t}^{\mathsf{T}}\theta^*|\,\|Z_{x,\Delta t}\|_1^r)$, and Φ denotes the distribution function of the increment $Z_{\Delta t}$. As a consequence, the condition in equation (27) becomes effectively the same as that in equation (12) because the term I_Z involving $Z_{\Delta t}$ is always bounded.

As a result of this analysis, Assumptions $1^\dagger$ and $2^\dagger$ become essentially the same as Assumptions 1 and 2 when $\alpha = 2$. It remains to demonstrate how the functional in the upper bound in equation (30) reduces to the one in equation (14). Given that the distribution of $Z_{\Delta t}$ is Gaussian, the conditional means of $Z_{\Delta t}$ are linear and $\boldsymbol{b}_r(Z_{\Delta t}) = 0$. Furthermore, from the conditions in Assumption 2, we conclude that $\boldsymbol{b}_r(V_{\Delta t})$ is bounded. Thus, we can represent the integral in the definition of $\boldsymbol{b}_r^{\theta^*}(V_{\Delta t}, Z_{\Delta t})$ in equation (28) as a difference of two well-defined integrals

$$\begin{aligned}\|u\|_1^{-r}\int_{R^{d+1}} (y - x^{\mathsf{T}}\theta^*)e^{iu^{\mathsf{T}}x}d(F-\Phi) \\ = \|u\|_1^{-r}\int_{R^{d+1}} (y - x^{\mathsf{T}}\theta^*)e^{iu^{\mathsf{T}}x}dF - \|u\|_1^{-r}\int_{R^{d+1}} (y - x^{\mathsf{T}}\theta^*)e^{iu^{\mathsf{T}}x}d\Phi \\ = \|u\|_1^{-r}\int_{R^{d+1}} (y - x^{\mathsf{T}}\theta^*)e^{iu^{\mathsf{T}}x}dF.\end{aligned} \tag{31}$$

As a consequence,

$$\boldsymbol{b}_r^{\theta^*}(V_{\Delta t}, Z_{\Delta t}) = \boldsymbol{b}_r(V_{\Delta t})$$

and the upper bound in Theorem 3 is indeed the same as the upper bound in Theorem 2.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

Finally, we note that under Assumption 1†, when $\alpha < 2$, the calculation in (31) is not possible because the two quantities on the right side are not finite on a standalone basis even though the left side can be, as shown in Lemma A.8.

4.1. A Connection with the Generalized CLT

It is important to note that in the infinite variance case, the limit law may have nonlinear conditional means, see Samorodnitsky and Taqqu (1994, Chap. 5) for a complete characterization of the bivariate case. Therefore, if no vector of parameters solves equations (24) and (25), then $\varsigma(n^{-1/\alpha}V_{n\Delta t})$ does not decay. Next, we discuss this situation in more detail and we clarify a connection with a classical rate-of-convergence result in the GCLT setting. To simplify the expressions, we denote the expectation appearing in $\varsigma(n^{-1/\alpha}V_{n\Delta t})$ and the corresponding expectation of the limit by

$$\varpi^{\theta}_{\alpha,n}(u) = \mathbb{E}\left[\frac{1}{n^{1/\alpha}}(Y_{n\Delta t} - X^{\mathsf{T}}_{n\Delta t}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}X_{n\Delta t}}\right]$$

$$\varpi^{\theta}_{\alpha}(u) = \mathbb{E}(Z_{y,\,\Delta t} - Z^{\mathsf{T}}_{x,\,\Delta t}\theta)e^{iu^{\mathsf{T}}Z_{x,\,\Delta t}},$$

where θ is the same in both expressions.

The result in this section involves the following weighted metric between the ch.f. of two random vectors U and W,

$$\boldsymbol{\chi}_s(U,W) = \sup_u \|u\|_1^{-s}\,|\mathbb{E}e^{iu^{\mathsf{T}}U} - \mathbb{E}e^{iu^{\mathsf{T}}W}|, \tag{32}$$

where $s \geq 0$. A classical result in the rate-of-convergence problem for GCLT restated in our notation shows that if the next moment conditions hold,

$$\int_{\mathbb{R}^{d+1}} z^{\gamma}\,d(F-G) = 0,\ \forall\gamma:\ |\gamma| \leq \ell \quad \text{and} \quad \boldsymbol{\nu}_s(V_{\Delta t}, Z_{\Delta t}) < \infty, \tag{33}$$

where $s = \ell + \delta$, $\ell \geq 1$ is an integer, $\delta \in (0,1]$, the absolute pseudomoment $\boldsymbol{\nu}_s$ is provided in equation (17), $Z_{\Delta t}$ has an α-stable distribution, and γ is a multi-index, then

$$\boldsymbol{\chi}_s\left(n^{-1/\alpha}V_{n\Delta t}, n^{-1/\alpha}Z_{n\Delta t}\right) \leq n^{1-s/\alpha}\boldsymbol{\chi}_s(V_{\Delta t}, Z_{\Delta t}), \tag{34}$$

in which $s > \alpha$ and the right side is finite (Rachev et al., 2013, Chap. 15). The metric $\boldsymbol{\chi}_s$ satisfies a set of properties which define it as an ideal probability metric of order s and the inequality in equation (34) is a direct consequence of them.[7] The conditions in equations (16) and (17) are the same as the ones in equation (33)

[7]The statement in equation (34) follows immediately from the ideality properties of the $\boldsymbol{\chi}_s$ metric (see, e.g., Rachev et al., 2013, Chap. 15). Adapted to the finite-dimensional setting of this article, a probability metric $\boldsymbol{\mu}(X,Y)$ is called *ideal* of order $s \in \mathbb{R}$ if for any random vectors X, Y, Z and a nonzero constant c, the next two properties are satisfied (Rachev et al., 2013, Def. 15.3.1): (ideality) $\boldsymbol{\mu}(X+Z, Y+Z) \leq \boldsymbol{\mu}(X,Y)$ and (homogeneity of order s) $\boldsymbol{\mu}(cX, cY) = |c|^s\boldsymbol{\mu}(X,Y)$.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

when $Z_{\Delta t}$ has a multivariate Gaussian distribution and the upper bound in equation (34) holds for the CLT with $\alpha = 2$.

The connection between the GCLT and Theorem 3 is provided below. Basically, the moment conditions in equation (33) replace Assumption 2† and the limiting distribution is not assumed to have a linear conditional mean; that is, the part of Assumption 1† involving the existence of θ_α^* is ignored.

THEOREM 4. *Suppose the random vector $V_{\Delta t} = (Y_{\Delta t}, X_{\Delta t}) \in \mathbb{R} \times \mathbb{R}^d$ is an increment of a Lévy process with $\mathbb{E}(V_{\Delta t}) = 0$ with distribution function F and $Z_{\Delta t} = (Z_{y,\Delta t}, Z_{x,\Delta t}) \in \mathbb{R} \times \mathbb{R}^d$ is an increment of a strictly α-stable process, $1 < \alpha \leq 2$, with distribution function G. Also, suppose that the moment conditions in equation (33) hold with $s = k + \delta$, $k \geq 2$, $\delta \in (0, 1]$. Then, the next inequality holds for any bounded θ,*

$$\sup_u \left\{ \psi(u) \left| \varpi_{\alpha,n}^\theta(u) - \varpi_\alpha^\theta(u) \right| \right\} \leq C_{s-1} n^{1-\frac{s}{\alpha}} \boldsymbol{b}_{s-1}^\theta(V_{\Delta t}, Z_{\Delta t}) + \alpha C_{s-1+\alpha} D_\alpha n^{1-\frac{s}{\alpha}} \boldsymbol{\chi}_s(X_{\Delta t}, Z_{x,\Delta t}), \tag{35}$$

where $\psi(u) > 0$ is the ch.f. of a symmetric random vector with a sufficiently smooth density, $C_p = \sup_u \|u\|_1^p \psi(u)$, $D_\alpha = \sup_{t \in B} \left| \omega_{\alpha,\Lambda}(t) \right|$, in which $B = \{z \in \mathbb{R}^d : \|z\|_1 = 1\}$ and $\omega_{\alpha,\Lambda}(t)$ is provided in Lemma A.9.

To illustrate the connection with Theorem 3, represent $z = (y, x_1, \ldots, x_d)$ and suppose that the conditions in equation (33) hold for some ℓ. Then, we have

$$\int_{\mathbb{R}^{d+1}} y x^\nu d(F - G) = 0 \quad \text{and} \quad \int_{\mathbb{R}^{d+1}} x_j x^\nu d(F - G) = 0 \tag{36}$$

for all multi-indices $\nu : |\nu| = \ell - 1$ and $j = 1, \ldots, d$. It is easy to see that any linear combination of the integrals on the left side of (36) also vanishes and, therefore, the condition in equation (26) in Assumption 2† holds for any choice of θ when $k = \ell - 1$. Finally, the boundedness of the absolute pseudomoment $\boldsymbol{\nu}_s$ assumed in (33) combined with the inequality in (29) ensures that the moment condition in (27) in Assumption 2† holds with $r = s - 1$ as well. As a consequence, Lemma A.8 shows that under these assumptions the functional $\boldsymbol{b}_r^\theta(V_{\Delta t}, Z_{\Delta t})$ exists for any bounded θ.[8]

Since all quantities in the upper bound in (35) are finite, its right side provides a rate of convergence of $\varpi_{\alpha,n}^\theta$ to ϖ_α^θ under the weighted uniform distance and, clearly, ϖ_α^θ may not vanish for all values of u. However, if the corresponding conditional mean of the limit law is linear, then $\varpi_\alpha^\theta(u) = 0$ for all $u \in \mathbb{R}^d$ for some $\theta = \theta_\alpha^*$. In this situation, the constant D_α vanishes as well because in this case there exists θ_α^* satisfying equations (24) and (25) which, in turn, ensure $\omega_{\alpha,\Lambda}(t) = 0$ for

[8]The function $\varpi_\alpha^\theta(u)$ at $\theta = 0$ is the Fourier transform of $\mathbb{E}(Z_{y,\Delta t} | Z_{x,\Delta t} = z) dH(z)$, where H is the distribution function of $Z_{x,\Delta t}$ (Bierens, 1982). Setting $\theta = 0$ in Theorem 4 effectively reveals the rate of convergence of the Fourier transform of the probability weighted conditional mean of the normalized sum to its limit.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

all $t \in \mathbb{R}^d$.[9] As a result, the left side of (35) turns into $\varsigma(n^{-1/\alpha} V_{n\Delta t})$ and the second term on the right side vanishes, which makes the right side of (35) the same as the upper bound in equation (30) in Theorem 3. Even though (30) can be viewed as a special case of equation (35), in practice, Theorem 3 is proved under weaker conditions.

5. NUMERICAL EXAMPLES AND FURTHER COMMENTS

This section includes numerical examples illustrating the behavior of the conditional mean of the normalized sum as a function of n, where we arbitrarily set $\Delta t = 1$. Also, we provide additional comments concerning extensions of the theoretical setup. To simplify the notation further, we denote the properly normalized sum of increments by

$$\widetilde{V}_n = (\widetilde{Y}_n, \widetilde{X}_n) = \frac{1}{n^{1/\alpha}} \left(\sum_{k=1}^{n} \Delta Y_k, \sum_{k=1}^{n} \Delta X_k \right), \tag{37}$$

where $\alpha \in (1, 2]$.

The examples are constructed with flexible parametric classes illustrating both the finite and the infinite variance cases discussed in the article. It is convenient if the ch.f. of the increment enjoys a closed-form expression because of the next connection with the conditional mean (Samorodnitsky and Taqqu, 1994, Sect. 5.1). Suppose that the random vector $(Y, X) \in \mathbb{R} \times \mathbb{R}^d$ has a finite expectation and denote its ch.f. by $\phi(v, u) = \mathbb{E}e^{i(vY + u^{\mathsf{T}}X)}$. The following expression holds:

$$\mathbb{E}[Y|X = x] = \frac{1}{2\pi f_X(x)} \int_{\mathbb{R}^d} \frac{\phi'_v(0, u)}{i} e^{-ix^{\mathsf{T}}u} du, \tag{38}$$

where $\phi'_v(0, u) = \frac{\partial}{\partial v}\phi(v, u)|_{v=0} = i\mathbb{E}(Ye^{iu^{\mathsf{T}}X})$.

As a consequence, the conditional mean can be computed numerically with the fast Fourier transform (FFT) applied to the function $\frac{1}{i}\phi'_v(0, u)$. In a similar vein, if the density $f_X(x)$ is not available in closed form, it can also be approximated with the FFT method applied to the marginal ch.f. $\phi_X(u) = \phi(0, u)$.

5.1. Hyperbolic Skewed-t Examples

The hyperbolic skewed-t process is a special case of the generalized hyperbolic process which has received a lot of attention in the finance literature (Eberlein, 2001; McNeil, Frey, and Embrechts, 2005; Bianchi et al., 2019). Suppose that the vector $V_1 \in \mathbb{R}^{m+d}$ has a hyperbolic skewed-t distribution denoted by

[9] See Lemmas A.6 and A.9.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

$V_1 \in \mathrm{HSt}_\lambda(\mu, \beta, \Sigma, \chi)$ and defined by the following ch.f.:

$$\phi_1(z) = \frac{2^{1+\lambda} e^{iz^\mathsf{T}\mu}}{\Gamma(-\lambda)} \left(-2\chi\left(iz^\mathsf{T}\beta - \frac{1}{2}z^\mathsf{T}\Sigma z\right)\right)^{-\lambda/2} K_\lambda\left(\sqrt{-2\chi\left(iz^\mathsf{T}\beta - \frac{1}{2}z^\mathsf{T}\Sigma z\right)}\right),$$

in which we set $z = (v, u) \in \mathbb{R}^{m+d}$ and $K_\lambda(x) = \frac{1}{2}\int_0^\infty t^{\lambda-1} e^{-x(t+t^{-1})/2} dt$, $x > 0$, denotes the Bessel function of the second kind. The parameters have the following interpretations: $\mu \in \mathbb{R}^{m+d}$ is a location parameter, $\beta \in \mathbb{R}^{m+d}$ is a skewness parameter, Σ is an $(m+d) \times (m+d)$ symmetric matrix, $\chi > 0$ acts like a scaling parameter, and $\lambda < 0$ is a shape parameter.

Most of the intuition behind the interpretation of the parameters follows from the normal location-scale mixture representation:

$$V_1 = \mu + \beta W + \sqrt{W} Z, \tag{39}$$

where W has an inverse gamma distribution $\mathrm{IG}(-\lambda, \chi/2)$ and $Z \in \mathcal{N}(0, \Sigma)$ has a multivariate Gaussian distribution with covariance matrix Σ. From this representation, it is easy to see that linear transformations belong to the same class. Suppose that $\mathbf{A}$ is a $q \times (m+d)$ matrix. Then, $\mathbf{A}V_1 \in \mathrm{HSt}_\lambda(\mathbf{A}\mu, \mathbf{A}\beta, \mathbf{A}\Sigma\mathbf{A}^\mathsf{T}, \chi)$.

Following Section 3.2, partition the random vector V_1 in the following manner $V_1 = (Y_1, X_1) \in \mathbb{R}^m \times \mathbb{R}^d$. The parameters of the distribution are partitioned accordingly

$$\mu = \begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_Y \\ \beta_X \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}.$$

Jamalizadeh and Balakrishnan (2019) provide an expression for the conditional mean of a general normal location-scale mixture which takes the following form:

$$\mathbb{E}(Y_1 | X_1 = x) = \mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(x - \mu_X) + (\beta_Y - \Sigma_{YX}\Sigma_{XX}^{-1}\beta_X)\mathbb{E}W_x, \tag{40}$$

where W_x denotes the conditional distribution $W|(X_1 = x)$. As a consequence, the conditional mean is linear if and only if $\beta_Y = \Sigma_{YX}\Sigma_{XX}^{-1}\beta_X$. Even though the expectation $\mathbb{E}W_x$ can be computed explicitly for the hyperbolic skewed-t distribution, we do not work with these expressions because the class is not closed with respect to the convolution operation. However, the ch.f. of the normalized sum $\widetilde{V}_n$ is easy to write down and the conditional mean can be computed numerically using equation (38).

The hyperbolic skewed-t has both finite and infinite variance representatives. The tail behavior of the distribution can be summarized in the following way. If we take a linear combination $w^\mathsf{T}V_1$ with $w^\mathsf{T}\beta > 0$, then the asymptotic decay of the density is given by

$$f_{w^\mathsf{T}V_1}(y) \sim |y|^{\lambda-1} e^{-2w^\mathsf{T}\beta y}, \qquad y \to \infty$$
$$f_{w^\mathsf{T}V_1}(y) \sim |y|^{\lambda-1}, \qquad y \to -\infty.$$

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

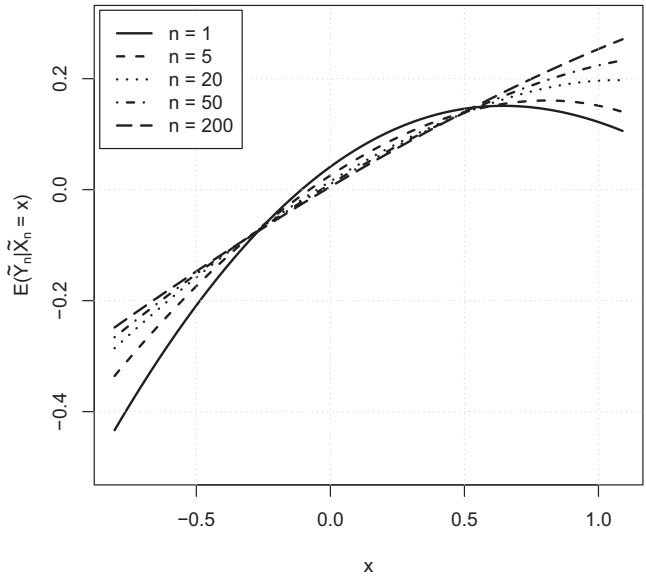


FIGURE 1. The conditional mean of the normalized sum as a function of the number of summands n when V_1 has a finite variance and a nonlinear conditional mean. $V_1 = (Y_1, X_1)$ follows a hyperbolic skewed-t with parameters $\lambda = -5$ and $\beta = (-2, 2)$. The conditional mean converges to a linear function.

As a consequence, V_1 has a finite expectation if $\lambda < -1$ and a finite variance if $\lambda < -2$. The asymptotic property of the tail also suggests the correct exponent α in the normalization in (37) is $\alpha = \min(-\lambda, 2)$.

5.1.1. *Convergence to a Linear Function.* Our first example is an illustration of the convergence result in Section 3. We assume $m = 1$, $d = 1$ and that V_1 has a finite variance and a nonlinear conditional mean. In terms of parameters, we pick $\lambda = -5$, $\chi = 1$,

$$\beta = \begin{pmatrix} -2 \\ 2 \end{pmatrix}, \qquad \mu = \frac{\chi}{2(\lambda+1)}\beta, \qquad \text{and} \qquad \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

The value for μ ensures that the distribution has a mean of zero, the remaining values are chosen for convenience. The shape of the conditional mean as function of the number of summands n is provided in Figure 1. The x-axis covers the range from the 1%-quantile to the 99%-quantile of the distribution of X_1. In line with the developed theory, the conditional mean converges to a linear function.

5.1.2. *Rate of Convergence for Mixture Models.* An interesting question is if there exist parameter values such that the conditions in equation (11) hold with $k \geq 2$. Consider again the case $m = 1$, $d = 1$ and assume that all parameters are

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

fixed except for β_Y. Given that $\mathbb{E}(V_1) = 0$, the stochastic representation in (39) can be restated equivalently as

$$\begin{aligned} Y_1 &= \beta_Y(W - \mathbb{E}W) + \sqrt{W}Z_Y \\ X_1 &= \beta_X(W - \mathbb{E}W) + \sqrt{W}Z_X. \end{aligned} \tag{41}$$

The third-order condition in equation (11) corresponding to $k = 2$ takes the form $\mathbb{E}(Y_1X_1^2) - \theta^*\mathbb{E}X_1^3 = 0$. This is a linear equation in β_Y because $\theta^* = \mathbb{E}(Y_1X_1)/\mathbb{E}(X_1^2)$ is also linear in β_Y and the only solution is $\beta_Y = \sigma_{YX}\beta_X/\sigma_{XX}$ found by direct verification. According to equation (40), however, this choice of β_Y turns the conditional mean into a linear function. As a consequence, in this case, the third-order moment condition in equation (11) guarantees linearity of the conditional mean.

The argument presented above is general and the same conclusion holds for any normal location-scale mixture for which the corresponding moments exist. Corollary 1, therefore, captures the rate of convergence for location-scale mixtures with finite variance. As far as δ is concerned, if the third moments of the marginals are finite, $\lambda < -3$, then $\delta = 1$ can be used. If the third moments are unbounded, then the upper limit on δ is determined by the tail behavior of the marginals. For instance, in the case of the hyperbolic skewed-t, if we require $-3 \leq \lambda < -2$ to guarantee existence of second moments when the third moments are unbounded, the corresponding rate of decay is provided in Corollary 1 with $0 < \delta < \min(-\lambda - 2, 1)$.

A way to engineer a model more flexible than the classical normal location-scale mixture is to change the tail behavior of the components in equation (41). Consider the next modification,

$$\begin{aligned} Y_1 &= \beta_Y(h(W) - \mathbb{E}h(W)) + \sqrt{h(W)}Z_Y \\ X_1 &= \beta_X(W - \mathbb{E}W) + \sqrt{W}Z_X, \end{aligned} \tag{42}$$

where $h(x) = G^{-1}(F(x))$ in which F is the c.d.f. of W and G is the c.d.f. of a positive random variable. This construction is similar to the meta-Student-t distribution featuring different degrees of freedom for the marginals and the meta-elliptical distribution (Fang, Fang, and Kotz, 2002; Demarta and McNeil, 2005). It is easy to see from the definition in (42) that the conditional mean is nonlinear even when $\beta_X = \beta_Y = 0$,

$$\mathbb{E}(Y_1|X_1 = x) = x\frac{\sigma_{YX}}{\sigma_{XX}}\mathbb{E}\left(\frac{h(W_x)}{W_x}\right)^{1/2}, \tag{43}$$

where W_x denotes the conditional distribution $W|(X_1 = x)$. At the same time, the multivariate law is symmetric when $\beta_X = \beta_Y = 0$ and the faster rate of convergence in Corollary 2 holds.

In a similar fashion, the equation $\mathbb{E}(Y_1X_1^2) - \theta^*\mathbb{E}X_1^3 = 0$ is linear in β_Y and can be used to determine β_Y as a function of β_X. The different tail behavior of the components provides more flexibility and, unlike the classical normal

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

location-scale mixture, the moment condition for $k = 2$ can hold when the distribution is asymmetric and the conditional mean $\mathbb{E}(Y_1|X_1 = x)$ is nonlinear.

A similar argument is used in the construction in Section 5.2 in which the conditional mean is specified explicitly. A difficulty with the meta-mixture model in (43) is that the expectation on the right hand-side is not available in closed form.

5.1.3. *Cross-Sectional Aggregation.* The argument developed in Section 3.2 concerning portfolios can be illustrated in the following manner. Suppose $V_1 = (Y_1, X_1) \in \mathbb{R}^m \times \mathbb{R}^d$ in which the first m components are interpreted as basic assets and the last d components are interpreted as factors. The operator $\mathbf{T}_p$ takes the form

$$\mathbf{T}_p = \begin{pmatrix} \mathbf{w}^\mathsf{T} & \mathbf{0}_{1\times d} \\ \mathbf{0}_{d\times m} & \mathbf{I}_{d\times d} \end{pmatrix},$$

where $\mathbf{w} = (w_1, \ldots, w_m)$, $\sum_{j=1}^m w_j = 1$, denote portfolio weights. From the properties of the hyperbolic skewed-t distribution, it follows that the random vector $\mathbf{T}_p V_1$ has the same distribution with parameters $\mathbf{T}_p V_1 \in \mathrm{HSt}_\lambda(\mathbf{T}_p\mu, \mathbf{T}_p\beta, \mathbf{T}_p\Sigma\mathbf{T}_p^\mathsf{T}, \chi)$

$$\mathbf{T}_p\mu = \begin{pmatrix} \mathbf{w}^\mathsf{T}\mu_Y \\ \mu_X \end{pmatrix}, \qquad \beta = \begin{pmatrix} \mathbf{w}^\mathsf{T}\beta_Y \\ \beta_X \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \mathbf{w}^\mathsf{T}\Sigma_{YY}\mathbf{w} & \mathbf{w}^\mathsf{T}\Sigma_{YX} \\ \Sigma_{XY}\mathbf{w} & \Sigma_{XX} \end{pmatrix}.$$

The relationship between the skewness parameters that results in a linear conditional mean can be readily expressed in terms of the weights and the parameters of the distribution of V_1

$$\mathbf{w}^\mathsf{T}(\beta_Y - \Sigma_{YX}\Sigma_{XX}^{-1}\beta_X) = 0.$$

Since this is a linear equation, clearly there are many portfolios with a linear conditional mean to the set of factors in X_1 even though all of the individual assets may have a nonlinear conditional mean to the same factor set.

5.1.4. *Convergence to a Nonlinear Function.* The next example concerns the case of a nonlinear limiting conditional mean in the case of infinite variance. Consider $(Y_1, X_1) \in \mathbb{R} \times \mathbb{R}$ and choose $\lambda = -1.75$ and $\beta = (0.001, 0.5)$, which implies $\alpha = 1.75$ in equation (37). Both marginals have an infinite second moment and their limiting distribution is in the α-stable class. As noted in Theorem 4, in this situation, the limiting conditional mean may not be linear. Indeed, as illustrated in Figure 2, the conditional mean does not converge to a linear function.

5.2. Examples of Different Rates of Convergence

Equation (11) provides the moment conditions that guarantee a given order of the convergence rate when the second moment is finite. In this section, our objective is to develop a bivariate example featuring a degree of analytic tractability in which the distribution of X_1 is asymmetric and remains fixed, the conditional mean is specified in closed form and its parameters can be chosen so that the moment conditions in equation (11) hold with $k \leq 2$. Such an example is interesting also

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

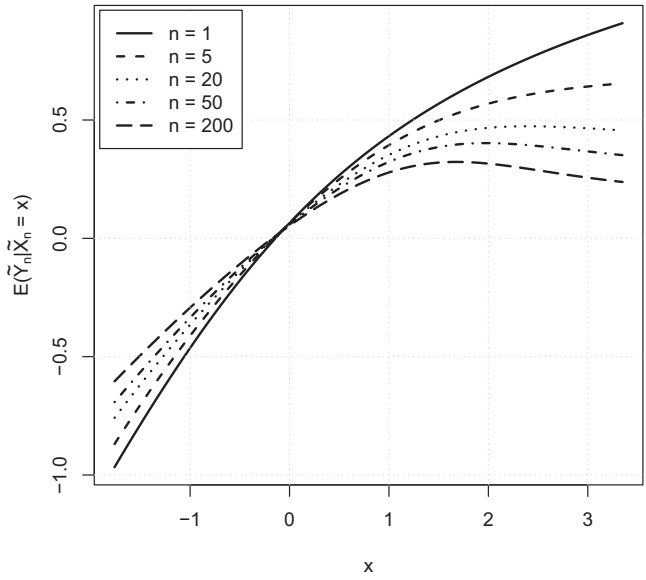


FIGURE 2. The conditional mean of the normalized sum as a function of the number of summands n when V_1 has a nonlinear conditional mean and infinite variance. $V_1 = (Y_1, X_1)$ follows a hyperbolic skewed-t with parameters $\lambda = -1.75$, $\beta = (0.001, 0.5)$. The conditional mean does not converge to a linear function.

to illustrate that the moment conditions in (11) are not a trivial consequence of the CLT and the rate of convergence can be faster than the corresponding CLT rate for the same law.

Let $(Y, X) \in \mathbb{R} \times \mathbb{R}$ and assume X follows an asymmetric Laplace distribution with density

$$f_X(x) = \begin{cases} \dfrac{\kappa_1\kappa_2}{\kappa_1+\kappa_2} \exp\{-\kappa_1(x-\mu)\}, & x \geq \mu \\ \dfrac{\kappa_1\kappa_2}{\kappa_1+\kappa_2} \exp\{-\kappa_2(\mu-x)\}, & x < \mu, \end{cases}$$

where $\kappa_1, \kappa_2 > 0$. Define Y by means of the stochastic representation

$$Y = g(X) + \epsilon, \tag{44}$$

where ϵ and X are independent, ϵ is assumed to have finite variance, and the conditional mean is defined explicitly in the following way:

$$g(x) = \begin{cases} (x-\mu)^{\gamma_1}, & x \geq \mu \\ -(\mu-x)^{\gamma_2}, & x < \mu, \end{cases} \tag{45}$$

where $\gamma_1, \gamma_2 > 0$.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

The utility of this construction is that the expectations behind ς and the corresponding moment conditions can be calculated analytically in closed form. The disadvantage is that although the distribution of X is infinitely divisible, that of Y may not be and in such a case (Y,X) cannot be used to construct a Lévy process. Nevertheless, all results in the article can be restated without modification for the classical CLT scheme of sums of i.i.d. random vectors and the example is relevant (see Section 5.3).

Equation (45) and the relationship $\mathbb{E}(Ye^{iuX}) = \mathbb{E}(\mathbb{E}(Y|X)e^{iuX})$ are used to arrive at the next expressions for the ch.f. of X and all relevant moments

$$\phi_X(u) = \mathbb{E}(e^{iuX}) = \frac{\kappa_1\kappa_2 e^{iu\mu}}{(\kappa_1 - iu)(\kappa_2 + iu)} \tag{46}$$

$$\mathbb{E}(Ye^{iuX}) = \frac{\kappa_1\kappa_2 e^{iu\mu}}{\kappa_1 + \kappa_2}\left[\frac{\Gamma(\gamma_1+1)}{(\kappa_1 - iu)^{\gamma_1+1}} - \frac{\Gamma(\gamma_2+1)}{(\kappa_2 + iu)^{\gamma_2+1}}\right] \tag{47}$$

$$\mathbb{E}(Xe^{iuX}) = \frac{\kappa_1\kappa_2 e^{iu\mu}}{\kappa_1 + \kappa_2}\left[\frac{1}{(\kappa_1 - iu)^2} - \frac{1}{(\kappa_2 + iu)^2}\right] + \mu\phi_X(u). \tag{48}$$

The remaining moments together with θ^* defined in equation (4) needed to compute ς and the conditions in (11) are calculated by evaluating derivatives of (46)–(48) at $u = 0$. Both X and Y need to be centered and, therefore, the location parameter is set to $\mu = -\mathbb{E}X = -(\kappa_2 - \kappa_1)/\kappa_1\kappa_2$. The mean of Y needs to be subtracted as well inside the expectation in (47), which will result in a new term on the right side.

We note that the actual distribution of the residual in (44) is irrelevant in these calculations. The skewed Laplace distribution has exponential tails and the variance of $g(X)$ exists. The requirement for finite variance of ϵ is needed only to make sure the variance of Y is finite.

The conditional mean $\mathbb{E}(\widetilde{Y}_n|\widetilde{X}_n = x)$ is computed in two steps. First, repeating the steps behind the derivation of (5) we find that

$$\mathbb{E}(\widetilde{Y}_n e^{iu\widetilde{X}_n}) = n^{1/2}\mathbb{E}(Ye^{in^{-1/2}uX})\left[\phi_X(n^{-1/2}u)\right]^{n-1}.$$

Second, we use equation (38) and apply the FFT to the function $\mathbb{E}(\widetilde{Y}_n e^{iu\widetilde{X}_n})$ to evaluate numerically the conditional mean. The left plot in Figure 3 shows how the conditional mean changes as a function of n with parameter values $\kappa_1 = 2.5$, $\kappa_2 = 1$, and $\gamma_1 = \gamma_2 = 2$. The distribution of X is skewed to the left and the grid for x ranges from the 1% quantile to the 99% quantile of X. With these parameter values, the third-order moment condition in (11) does not hold. The rate of convergence in Corollary 1 is expected to hold for this situation.

Next, we find a new value for γ_1 by numerically solving the third-order condition in (11), $\mathbb{E}[(Y - \theta^*X)X^2] = 0$. Conditional on the other parameters, the solution is $\gamma_1^* = 5.1 > \gamma_1 = 2$ which is intuitive given that g needs to steepen to compensate for the lighter right tail of X in order for the moment condition to hold. The right plot in Figure 3 illustrates this case. Even though the conditional mean g is now

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

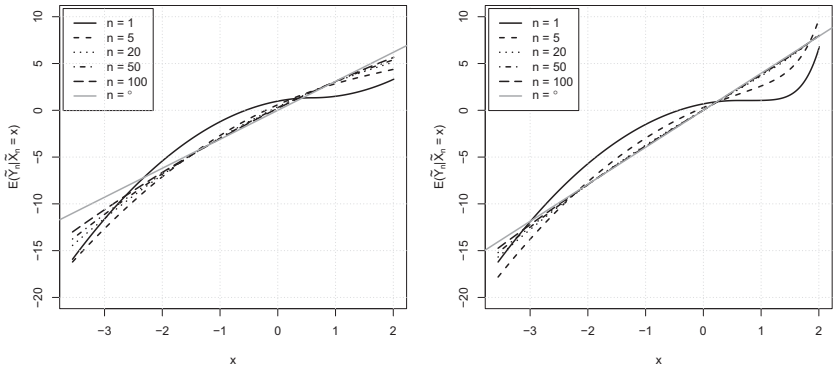


FIGURE 3. The conditional mean of the normalized sum as a function of the number of summands n with parameters $\kappa_1 = 2.5$, $\kappa_2 = 1$, and $\gamma_1 = \gamma_2 = 2$ (left) and $\kappa_1 = 2.5$, $\kappa_2 = 1$, $\gamma_1 = 5.1$, and $\gamma_2 = 2$ (right). The plot on the right illustrates a faster rate of convergence because the choice of parameters ensures that the third-order moment condition in (11) is satisfied. The grid on the horizontal axis starts at the 1% quantile and ends at the 99% quantile of the distribution of X.

more curved, the rate of convergence to the asymptotic linear case is visually faster as expected based on Theorem 2.

The actual rate of convergence in these two cases is estimated formally in the following way. The log-transform applied to both sides of (14) leads to the next asymptotic relationship for large n

$$\log\left(\varsigma(\widetilde{V}_n)\right) = a + \frac{1-r}{2}\log(n) + \varepsilon, \tag{49}$$

where $r = k + \delta$ in which k denotes the corresponding moment order. The left side is calculated numerically from the definition in (7) with $\psi(u) = \exp(-u^2/2)$ in which we set $\theta = \theta^*$. The quantity in the absolute value is calculated using (5) and the explicit representations in (47) and (48). Finally, the maximization of the absolute value is conducted numerically over a grid whose size is dictated by the fast decay of $\psi(u)$.

Both sides of (49) are evaluated for $n = 1, 2, \ldots, 200$ but the coefficients are estimated from the data with $n \geq 100$ because the relationship is asymptotic in theory. For the first set of parameters, we find that $\hat{r}_{\text{slow}} = 1.99$ with $\hat{k}_{\text{slow}} = 1$ and for the second set of parameters, $\hat{r}_{\text{fast}} = 2.98$ with $\hat{k}_{\text{fast}} = 2$, which is exactly as expected. The estimated relationships together with the data are presented in Figure 4.

Finally, we compare the rate of convergence in the CLT to the rates above. It is important to note that the rate of convergence in the classical CLT depends on some additional properties of the distribution of ϵ in (44), while the rates computed above do not depend on ϵ. For example, in line with the conditions in (33), to reach

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

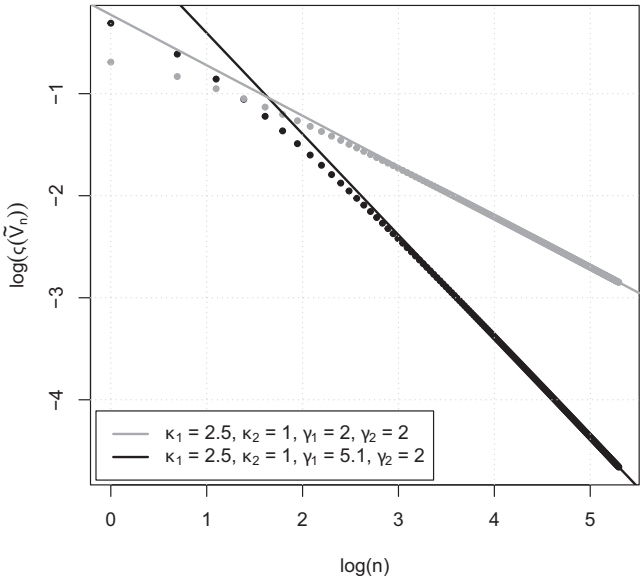


FIGURE 4. Estimated rates of convergence for a distribution that meets the third-order moment condition in (11) (in black) and for one that does not (in gray). In line with theory, the estimated rates of convergence are $\hat{r}_{\text{fast}} = 2.98$ with $\hat{k}_{\text{fast}} = 2$ for the data in black and $\hat{r}_{\text{slow}} = 1.99$ with $\hat{k}_{\text{slow}} = 1$ for the data in gray.

the faster rate $\hat{r}_{\text{fast}}$ one would need to require,

$$\mathbb{E}(Y^{\ell}X^{m}) = \mathbb{E}(Z_y^{\ell}Z_x^{m}),$$

where ℓ, m are positive integers or zero such that $\ell + m = 1, 2, 3$ and (Z_y, Z_x) has a bivariate Gaussian distribution.[10] Clearly, these conditions imply symmetry for the marginals $\mathbb{E}Y^3 = \mathbb{E}X^3 = 0$ and X is asymmetric by construction while the symmetry of Y depends on the properties of ϵ. In this particular example, the rate behind $\hat{r}_{\text{fast}}$ is not attainable for the vector (Y, X) in the CLT, for example, in the Berry–Esseen setup, but even if $X \stackrel{d}{=} -X$, then the CLT rate would depend on the properties of ϵ. As a result, the limit theorems for the conditional mean developed in this article are not a simple consequence of the CLT. Rather, they can be viewed as a separate class of results specialized for conditional means.

5.3. Beyond Lévy Processes

The focus of the article is the assumption that the data generating process follows a multivariate Lévy process. We take advantage of the property that its increments

[10]This statement holds for different types of metrics including χ_s defined in (32), the uniform distance between distribution functions, and other metrics with uniform structure (Rachev et al., 2013, Chap. 15).

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

over time intervals of the equal length are i.i.d. random vectors. The proofs do not rely on the infinite divisibility of the increments and the rates of decay in Theorems 2–4 hold for sequences of i.i.d. random vectors as well.

From the standpoint of applications of discrete time processes, a more appealing setup is that of independent non-identically distributed random vectors. The framework developed in this article can handle this extension. Next, we provide a pair of assumptions similar to Assumptions 1 and 2.

Assumption 1‡. The sequence $V_1, V_2, \dots$ consists of independent non-identically distributed random vectors, $V_j = (Y_j, X_j) \in \mathbb{R} \times \mathbb{R}^d$ with $\mathbb{E}(V_j) = 0$, $\mathbb{E}\|V_j\|_2^2 < \infty$, and $\mathbb{E}[V_j V_j^\top] = \mathbf{C}$ for all j.

Assumption 2‡. The following moment conditions hold for all j :

$$\mathbb{E}\left((Y_j - X_j^\top \theta^*) X_j^\nu\right) = 0, \quad \forall \nu \colon |\nu| \le k \tag{50}$$

$$\mathbb{E}\left(|Y_j - X_j^\top \theta^*| \|X_j\|_1^r\right) \le D < \infty, \tag{51}$$

where $k \ge 1$ is an integer, ν denotes a multi-index, $r = k + \delta$ with $\delta \in (0, 1]$, and $\theta^* = \mathbf{C}_{XX}^{-1} \mathbf{C}_{YX}$ exists in which the index notation identifies the corresponding blocks in $\mathbf{C}$.

Both assumptions include moment conditions holding across all members of the sequence in the same way. Such a degree of uniformity is not required in the proof but simplifies the notation and makes the connection with the i.i.d. case easier to understand. The rate of decay is provided in the following theorem.

THEOREM 5. *Under Assumptions 1‡ and 2‡,*

$$\varsigma\left(n^{-1/2} \sum_{j=1}^{n} V_j\right) \le C_r n^{-\frac{r+1}{2}} \sum_{j=1}^{n} \boldsymbol{b}_r(V_j), \tag{52}$$

where $r \ge 1$, $C_r = \sup_u \|u\|_1^r \psi(u)$. *Furthermore, the next bound holds*

$$\sum_{j=1}^{n} \boldsymbol{b}_r(V_j) \le n A_\delta \frac{\Gamma(\delta + 1)}{\Gamma(r + 1)} D, \tag{53}$$

where the notation is as in Theorem 2.

If the elements of the sequence are i.i.d., then $\boldsymbol{b}_r(V_1) = \boldsymbol{b}_r(V_j)$ for all j and it is easy to see how equations (52) and (53) turn into equations (14) and (15). Combining both inequalities in Theorem 5 indicates the rate of decay is similar to the one in Theorem 2. The main difference is the increase in the scaling constant. The similarity in the rate of decay in the two theorems follows from the conditions in equation (50) holding uniformly for all j.

Relaxing the assumption of independence is a lot more challenging and is beyond the scope of this article but it is important for applications in economics

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

and finance (Rachev and Mittnik, 2000; Cont and Tankov, 2003; Stock and Watson, 2011). From the viewpoint of both theory and practice, one important general relaxation includes $V_1, V_2, \ldots$ forming a multivariate martingale difference sequence. The CLT holds in this case subject to certain regularity conditions and rate-of-convergence results in terms of the uniform metric in discrete and continuous time can be found in Butzer, Hahn, and Roeckerath (1983), Haeusler (1988), Račkauskas (1991), and Mourrat (2013).

Another important relaxation is weak dependence which implies the correlation between the elements of the sequence decreases as the time lag increases. Rate-of-convergence results in the CLT setting for weakly dependent random variables are available, for example, in Tikhomirov (1991, 1994).

Given that subject to certain regularity conditions there is a CLT result in both settings, the conditional expectations of a properly normalized sum converge to a linear function. It is a future topic of research to study the rate-of-convergence problem for conditional means in these settings and to understand how the rates compare to the corresponding CLT rates available in the academic literature.

6. CONCLUSION

Linear models have a central place in empirical asset pricing in the field of finance and are widely considered a successful framework. From an econometric perspective, however, the assumption of linearity is not commonly tested for adequacy. In this article, we propose an asymptotic probabilistic argument which indicates that the model error may decline for lower frequency return data due to temporal aggregation. Apart from the field of finance, the results in the article may be useful in other domains in which temporal aggregation plays a role.

If the return process of the joint vector of the dependent and the independent variables follows a multivariate Lévy process with finite second moments, we prove that the conditional expectation of the dependent variable converges to a linear function as the return frequency declines and the rate of convergence is bounded by a power law. The rate of convergence is determined by a set of moment conditions with an econometric interpretation. If the dependent variable is a return of a portfolio, the additional cross-sectional aggregation can improve the rate of convergence, which indicates that the model error of the linear framework can be smaller for portfolios than individual stocks.

Relaxing the assumptions of finite second moments leads to a non-trivial extension of the asymptotic results. Even though this case is of limited practical significance for the field of finance, it is of theoretical interest. As expected from general limit theory, the index of stability of the increments of the Lévy process appears in the exponent of the power law. Also, the limiting stable law appears explicitly in the scaling constant and in the moment conditions that play a role in the rate of convergence.

Finally, from a probabilistic perspective, the developed theory and the numerical examples indicate that the results presented in the article are not a simple

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

consequence of the CLT and the GCLT. Rather, they can be viewed as a separate class of limit theorems specialized for conditional expectations.

A. APPENDIX

Proof of Theorem 1. Set $\Delta t = \tau$ and assume $\mathbb{E}(Y_\tau|X_\tau) \overset{a.s.}{=} X_\tau^\mathsf{T}\theta$ for some fixed θ. We repeat the arguments behind equation (5)

$$\begin{aligned}\left|\mathbb{E}(Y_{n\tau} - X_{n\tau}^\mathsf{T}\theta)e^{iu^\mathsf{T}X_{n\tau}}\right| &= \left|\sum_{k=1}^{n}\mathbb{E}\left[(\Delta Y_k - \Delta X_k^\mathsf{T}\theta)e^{iu^\mathsf{T}\Delta X_k}\right]\mathbb{E}e^{iu^\mathsf{T}\sum_{j\neq k}\Delta X_j}\right| \\ &\leq \sum_{k=1}^{n}\left|\mathbb{E}(\Delta Y_k - \Delta X_k^\mathsf{T}\theta)e^{iu^\mathsf{T}\Delta X_k}\right|\left|\mathbb{E}e^{iu^\mathsf{T}\sum_{j\neq k}\Delta X_j}\right| \\ &\leq \sum_{k=1}^{n}\left|\mathbb{E}(\Delta Y_k - \Delta X_k^\mathsf{T}\theta)e^{iu^\mathsf{T}\Delta X_k}\right| \\ &= n\left|\mathbb{E}(Y_\tau - X_\tau^\mathsf{T}\theta)e^{iu^\mathsf{T}X_\tau}\right|.\end{aligned}$$

If $\mathbb{E}(Y_\tau - X_\tau^\mathsf{T}\theta)e^{iu^\mathsf{T}X_\tau} = 0$ for all $u \in \mathbb{R}^d$, then the left side vanishes for all u. As a result, $\mathbb{E}(Y_{n\tau}|X_{n\tau}) \overset{a.s.}{=} X_{n\tau}^\mathsf{T}\theta$ for any integer $n \geq 1$.

By Miller (1978, Thm. 3.1), the conditional mean $\mathbb{E}[Y_\tau|X_\tau]$ is linear if and only if the following relationship holds for the derivatives of the ch.f. of V_τ denoted by $\phi_{V_\tau}(z,u) = \mathbb{E}e^{izY_\tau + iu^\mathsf{T}X_\tau}$:

$$\partial_z\phi_{V_\tau}(0,u) = \sum_{i=1}^{d}\theta_i\partial_{u_i}\phi_{V_\tau}(0,u), \tag{A.1}$$

in which ∂_x denotes the first derivative with respect to the variable x. By assumption, equation (A.1) holds for some fixed θ.

Next, from the definition of a Lévy process, the ch.f. equals $\phi_{V_\tau}(z,u) = (\phi_{V_{\tau/k}}(z,u))^k$, where $k \geq 1$ is an integer. Computing the derivatives on both sides of (A.1) results in

$$k(\phi_{V_{\tau/k}}(0,u))^{k-1}\partial_z\phi_{V_{\tau/k}}(0,u) = \sum_{i=1}^{d}\theta_i k(\phi_{V_{\tau/k}}(0,u))^{k-1}\partial_{u_i}\phi_{V_{\tau/k}}(0,u).$$

The term $k(\phi_{V_{\tau/k}}(0,u))^{k-1}$ is common on both sides and, therefore, (A.1) holds for $V_{\tau/k}$ with the same θ. From here, we conclude that $\mathbb{E}(Y_{\tau/k}|X_{\tau/k}) \overset{a.s.}{=} X_{\tau/k}^\mathsf{T}\theta$ for any integer $k \geq 1$.

Combining both statements, we conclude that $\mathbb{E}(Y_{n\tau/k}|X_{n\tau/k}) \overset{a.s.}{=} X_{n\tau/k}^\mathsf{T}\theta$, where n and k are positive integers. As a final step, choose an arbitrary $t \geq 0$. Then, t/τ can be approximated arbitrarily well by a rational number of the form n/k, where n and k are positive integers. Thus, t can be represented in the form $t = n\tau/k$ and the result follows from the convergence $V_{n\tau/k} \to V_t$ in distribution. □

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

LEMMA A.1. *Suppose that* $V = (Y, X) \in \mathbb{R} \times \mathbb{R}^d$ *is a zero-mean random vector with* $\mathbb{E}\|V\|_1 < \infty$. *Define*

$$\varsigma(V) = \inf_{\theta} \varsigma^{\theta}(V)$$

$$\varsigma^{\theta}(V) = \sup_{u \in \mathbb{R}^d} \psi(u) \left| \mathbb{E}(Y - X^{\mathsf{T}}\theta) e^{iu^{\mathsf{T}}X} \right|,$$

where $\psi(u) > 0$ *is a ch.f. of a symmetric random vector. The functional* $\varsigma(V)$ *is nonnegative and equals zero if and only if* $\mathbb{E}(Y|X) = X^{\mathsf{T}}\theta$ *almost surely for some* θ.

Proof. For any fixed θ and any $u \in \mathbb{R}^d$,

$$\psi(u) \left| \mathbb{E}(Y - X^{\mathsf{T}}\theta) e^{iu^{\mathsf{T}}X} \right| \le \max(1, \|\theta\|_{\infty}) \mathbb{E}\|V\|_1 < \infty.$$

Therefore, under the assumptions in the lemma, $\varsigma^{\theta}(V) < \infty$, and, as a consequence, $\varsigma(V)$ is bounded as well.

If $\mathbb{E}(Y|X) = X^{\mathsf{T}}\theta$ almost surely for some θ, then by Bierens (1982, Thm. 1), it follows that $\mathbb{E}(Y - X^{\mathsf{T}}\theta) e^{iu^{\mathsf{T}}X} = 0$ for all $u \in \mathbb{R}^d$. As a result, $\varsigma(V) = 0$.

In the opposite direction, if $\varsigma^{\theta}(V) = 0$ for some θ, then $\mathbb{E}(Y - X^{\mathsf{T}}\theta) e^{iu^{\mathsf{T}}X} = 0$ for all $u \in \mathbb{R}^d$ because $\psi(u) > 0$. This is necessary and sufficient for $\mathbb{E}(Y|X) = X^{\mathsf{T}}\theta$ almost surely by Bierens (1982, Thm. 1). □

LEMMA A.2. *Suppose that* $V = (Y, X) \in \mathbb{R} \times \mathbb{R}^d$ *is a zero-mean random vector such that* $\mathbb{E}|Y - X^{\mathsf{T}}\theta^*| \, \|X\|_1^r < \infty$, *where* $r = k + \delta$, $k \ge 1$ *is an integer,* $0 < \delta \le 1$, *and* $\theta^* = (\mathbb{E}[XX^{\mathsf{T}}])^{-1}\mathbb{E}[XY]$. *Suppose also that the moment conditions* $\mathbb{E}(Y - X^{\mathsf{T}}\theta^*)X^{\nu} = 0, \forall \nu : |\nu| \le k$ *hold, where* ν *is a multi-index. Define*

$$\zeta_r(V) = \sup_{g \in \mathcal{G}} \left| \mathbb{E}(Y - X^{\mathsf{T}}\theta^*) g(X) \right|,$$

where

$$\mathcal{G} = \left\{ g : \mathbb{R}^d \to \mathbb{C}, \left| \partial^{\nu} g(u) - \partial^{\nu} g(v) \right| \le \|u - v\|_1^{\delta}, \delta \in (0, 1], \forall \nu : |\nu| = k \right\}. \tag{A.2}$$

The following bound holds:

$$\zeta_r(V) \le \frac{\Gamma(\delta + 1)}{\Gamma(r + 1)} \mathbb{E}\left(|Y - X^{\mathsf{T}}\theta^*| \, \|X\|_1^r \right).$$

Proof. The construction is similar to the construction of zeta distances in the theory of probability metrics (see Rachev et al., 2013, Chap. 4.4, Lemma 15.3.6, and Remark 15.4.2). The functions from the class $\mathcal{G}$ can be written in terms of their Taylor expansion at $x = 0$ (see, e.g., Cartan, 1971)

$$g(x) = \sum_{|\nu| \le k-1} \frac{\partial^{\nu} g(0)}{\nu!} x^{\nu} + k \sum_{|\nu| = k} \frac{x^{\nu}}{\nu!} \int_0^1 (1 - \tau)^{k-1} \partial^{\nu} g(\tau x) d\tau$$

$$= \sum_{|\nu| \le k} \frac{\partial^{\nu} g(0)}{\nu!} x^{\nu} + k \sum_{|\nu| = k} \frac{x^{\nu}}{\nu!} \int_0^1 (1 - \tau)^{k-1} (\partial^{\nu} g(\tau x) - \partial^{\nu} g(0)) d\tau,$$

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

where $\nu! = \nu_1!\,\nu_2!\ldots\nu_d!$. For the expectation, we find

$$\mathbb{E}(Y - X^{\mathsf{T}}\theta)g(X) = \mathbb{E}\left[(Y - X^{\mathsf{T}}\theta)\left(\sum_{|\nu|\le k}\frac{\partial^{\nu}g(0)}{\nu!}X^{\nu} + R_k(X)\right)\right].$$

The moment conditions $\mathbb{E}(Y - X^{\mathsf{T}}\theta)X^{\nu} = 0, \forall\nu\colon |\nu| \le k$ for some fixed θ are necessary for the existence of $\sup_{g\in\mathcal{G}}\left|\mathbb{E}(Y - X^{\mathsf{T}}\theta)g(X)\right|$. To see why, assume the contrary. If $\mathbb{E}(Y - X^{\mathsf{T}}\theta)X^{\nu} \neq 0$ for some $\nu\colon |\nu| \le k$, then it is possible to find a function in the class $\mathcal{G}$ with a sufficiently large derivative $\partial^{\nu}g(0)$ in absolute value so that the corresponding term in the expansion becomes arbitrarily large in absolute value.

Consider the first two terms. If $k = 0$, then $\mathbb{E}(Y - X^{\mathsf{T}}\theta) = 0$ is satisfied for any θ because by assumption $V = (Y, X)$ is zero-mean. If $k = 1$, the moment condition becomes $\mathbb{E}(Y - X^{\mathsf{T}}\theta)X = 0$ and it is satisfied at $\theta = \theta^*$ by design. For $k > 1$, the corresponding moment condition must hold at $\theta = \theta^*$. Thus, the necessary conditions for boundedness become $\mathbb{E}(Y - X^{\mathsf{T}}\theta^*)X^{\nu} = 0, \forall\nu\colon |\nu| \le k$.

Denote the term which does not vanish by I,

$$I = \mathbb{E}\left(\left(Y - X^{\mathsf{T}}\theta^*\right)k\sum_{|\nu|=k}\frac{X^{\nu}}{\nu!}\int_0^1(1-\tau)^{k-1}(\partial^{\nu}g(\tau X) - \partial^{\nu}g(0))d\tau\right).$$

The following inequalities hold:

$$\begin{aligned}
|I| &\le \mathbb{E}\left(\left|Y - X^{\mathsf{T}}\theta^*\right|\sum_{|\nu|=k}k\frac{|X|^{\nu}}{\nu!}\int_0^1(1-\tau)^{k-1}\left|\partial^{\nu}g(\tau X) - \partial^{\nu}g(0)\right|d\tau\right)\\
&\le \mathbb{E}\left(\left|Y - X^{\mathsf{T}}\theta^*\right|\sum_{|\nu|=k}k\frac{|X|^{\nu}}{\nu!}\int_0^1(1-\tau)^{k-1}\|\tau X\|_1^{\delta}\,d\tau\right)\\
&= \mathbb{E}\left(\left|Y - X^{\mathsf{T}}\theta^*\right|\sum_{|\nu|=k}k\frac{|X|^{\nu}}{\nu!}\|X\|_1^{\delta}\frac{\Gamma(k)\Gamma(\delta+1)}{\Gamma(k+\delta+1)}\right)\\
&= \mathbb{E}\left(\left|Y - X^{\mathsf{T}}\theta^*\right|\sum_{|\nu|=k}k!\frac{|X|^{\nu}}{\nu!}\|X\|_1^{\delta}\frac{\Gamma(\delta+1)}{\Gamma(k+\delta+1)}\right)\\
&= \frac{\Gamma(\delta+1)}{\Gamma(k+\delta+1)}\mathbb{E}\left(\left|Y - X^{\mathsf{T}}\theta^*\right|\left(\sum_{j=1}^{d}|X_j|\right)^{k}\|X\|_1^{\delta}\right)\\
&= \frac{\Gamma(\delta+1)}{\Gamma(k+\delta+1)}\mathbb{E}\left(|Y - X^{\mathsf{T}}\theta^*|\,\|X\|_1^{k+\delta}\right),
\end{aligned}$$

in which we take advantage of the multinomial theorem, $\sum_{|\nu|=k}\frac{k!}{\nu!}|X|^{\nu} = (\sum_{j=1}^{d}|X_j|)^k$.

As a result, for any $g \in \mathcal{G}$,

$$\left|\mathbb{E}(Y - X^{\mathsf{T}}\theta^*)g(X_t)\right| = |I| \le \frac{\Gamma(\delta+1)}{\Gamma(k+\delta+1)}\mathbb{E}\left(|Y - X^{\mathsf{T}}\theta^*|\,\|X\|_1^{k+\delta}\right),$$

and the proof is completed after taking the supremum over all $g \in \mathcal{G}$. □

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

LEMMA A.3. *Suppose that $V=(Y,X)\in\mathbb{R}\times\mathbb{R}^d$ is a zero-mean random vector and that the moment conditions in Lemma A.2 hold. Define*

$$\boldsymbol{b}_r(V)=\sup_{u\in\mathbb{R}^d}\|u\|_1^{-r}\left|\mathbb{E}(Y-X^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}X}\right|,\quad r\geq 1.$$

The functional is unbounded if $\theta\neq\theta^*$. The next upper bound holds

$$\boldsymbol{b}_r(V)\leq A_\delta\boldsymbol{\zeta}_r(V), \tag{A.3}$$

where $r=k+\delta$, $k\geq 1$ is an integer, $0<\delta\leq 1$, and $A_\delta=\sup_{y\in\mathbb{R}}|y|^{-\delta}|e^{iy}-1|$. Finally, the following scaling property holds:

$$\boldsymbol{b}_r(aV)=|a|^{r+1}\boldsymbol{b}_r(V),$$

where $a\in\mathbb{R}$ is a constant.

Proof. We show that $\boldsymbol{b}_r(V)$ can be represented as a special case of the $\boldsymbol{\zeta}_r(V)$ functional in which we consider a sub-class of $\mathcal{G}$. Choose $h_u(x)=\exp(iu^{\mathsf{T}}x)/\|u\|_1^{k+\delta}$, where $u\in\mathbb{R}^d$, and a multi-index ν such that $|\nu|=k$. A direct calculation shows that

$$\begin{aligned}\left|\partial^\nu h_u(x)-\partial^\nu h_u(y)\right|&\leq\frac{|u|^\nu|e^{iu^{\mathsf{T}}x}-e^{iu^{\mathsf{T}}y}|}{\|u\|_1^{k+\delta}}\\&\leq\frac{|u|^\nu|e^{iu^{\mathsf{T}}(x-y)}-1|}{\|u\|_1^{k+\delta}},\end{aligned}$$

where $|u|^\nu=\prod_{j=1}^d|u_j|^{\nu_j}$. We divide both sides by $\|x-y\|_1^\delta$ and we obtain

$$\begin{aligned}\frac{1}{\|x-y\|_1^\delta}\left|\partial^\nu h_u(x)-\partial^\nu h_u(y)\right|&\leq\frac{|u|^\nu|e^{iu^{\mathsf{T}}(x-y)}-1|}{\|u\|_1^{k+\delta}\|x-y\|_1^\delta}\\&\leq\frac{\|u\|_1^k|e^{iu^{\mathsf{T}}(x-y)}-1|}{\|u\|_1^{k+\delta}\|x-y\|_1^\delta}\\&=\frac{|e^{iu^{\mathsf{T}}(x-y)}-1|}{\|u\|_1^\delta\|x-y\|_1^\delta},\end{aligned}$$

which uses the fact that $\prod_j|u_j|^{\nu_j}\leq(\sum_j|u_j|)^k$ following from the multinomial theorem. The right hand-side can be increased further using

$$\left|\sum_{i=1}^d u_i(x_i-y_i)\right|\leq\sum_{i=1}^d|u_i||x_i-y_i|\leq\left(\sum_{i=1}^d|u_i|\right)\left(\sum_{j=1}^d|x_j-y_j|\right).$$

We obtain

$$\frac{1}{\|x-y\|_1^\delta}\left|\partial^\nu h_u(x)-\partial^\nu h_u(y)\right|\leq\frac{|e^{iu^{\mathsf{T}}(x-y)}-1|}{|u^{\mathsf{T}}(x-y)|^\delta}\leq\sup_{z\in\mathbb{R}}\frac{|e^{iz}-1|}{|z|^\delta}=A_\delta.$$

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

It is easy to see that $A_\delta < \infty$. As a result, we have proved the following inclusion:

$$\mathcal{H} = \left\{ h \colon \mathbb{R}^d \to \mathbb{C}, h(x) = \frac{1}{A_\delta} \frac{e^{iu^\mathsf{T} x}}{\|u\|_1^{k+\delta}} \right\} \subset \mathcal{G}$$

and, as a consequence, the next inequality holds with $\theta = \theta^*$

$$\begin{aligned} \frac{1}{A_\delta} \sup_{u \in \mathbb{R}^d} \|u\|_1^{-k-\delta} \left| \mathbb{E}(Y - X^\mathsf{T}\theta^*) e^{iu^\mathsf{T} X} \right| &= \sup_{h \in \mathcal{H}} \left| \mathbb{E}(Y - X^\mathsf{T}\theta^*) h(X) \right| \\ &\leq \sup_{g \in \mathcal{G}} \left| \mathbb{E}(Y - X^\mathsf{T}\theta^*) g(X) \right|, \end{aligned}$$

which completes the proof of the inequality in equation (A.3).

If $\theta \neq \theta^*$, then $\boldsymbol{b}_r(V)$ is unbounded. The argument is the same as the one in the proof of Lemma A.2. For any $\theta \neq \theta^*$, the moment $\mathbb{E}(Y - X^\mathsf{T}\theta)X \neq 0$. The corresponding term in the Taylor expansion of $h_u(x)$ then becomes unbounded when $\|u\|_1 \to 0$.

The scaling property follows easily from the definition. Let $\theta = \theta^*$, then

$$\begin{aligned} \sup_{u \in \mathbb{R}^d} \|u\|_1^{-r} \left| \mathbb{E}(aY - aX^\mathsf{T}\theta) e^{iu^\mathsf{T} aX} \right| &= |a| \sup_{u \in \mathbb{R}^d} \|u\|_1^{-r} \left| \mathbb{E}(Y - X^\mathsf{T}\theta) e^{iau^\mathsf{T} X} \right| \\ &= |a| \sup_{au \in \mathbb{R}^d} |a|^r \|au\|_1^{-r} \left| \mathbb{E}(Y - X^\mathsf{T}\theta) e^{iau^\mathsf{T} X} \right| \\ &= |a|^{r+1} \sup_{v \in \mathbb{R}^d} \|v\|_1^{-r} \left| \mathbb{E}(Y - X^\mathsf{T}\theta) e^{iv^\mathsf{T} X} \right|. \end{aligned}$$

□

LEMMA A.4. *Suppose $V_1, V_2, \ldots, V_n$ are independent copies of $V = (Y, X)$. Then,*

$$\varsigma(V_1 + \cdots + V_n) \leq \varsigma^\theta(V_1 + \cdots + V_n) \leq nC_r \boldsymbol{b}_r(V),$$

where $C_r = \sup_{u \in \mathbb{R}^d} \|u\|_1^r \psi(u)$. If $\psi(u) = e^{-\frac{1}{2}u^\mathsf{T} u}$, then $C_r = e^{-\frac{r}{2}} (rd)^{\frac{r}{2}}$.

Proof. We start from the definition,

$$\begin{aligned} \sup_u \left\{ \psi(u) \left| \mathbb{E}\left(\sum_{l=1}^n \left(Y_l - X_l^\mathsf{T}\theta \right) \right) e^{iu^\mathsf{T}(\sum_{m=1}^n X_m)} \right| \right\} &\leq \sum_{l=1}^n \sup_u \left\{ \psi(u) \left| \mathbb{E}\left(Y_l - X_l^\mathsf{T}\theta \right) e^{iu^\mathsf{T} X_l} \right| \right\} \\ &\leq \sum_{l=1}^n \sup_u \left\{ \|u\|_1^{-r} \left| \mathbb{E}\left(Y_l - X_l^\mathsf{T}\theta \right) e^{iu^\mathsf{T} X_l} \right| \right\} \sup_u \left\{ \psi(u) \|u\|_1^r \right\} \\ &= nC_r \sup_u \left\{ \|u\|_1^{-r} \left| \mathbb{E}\left(Y - X^\mathsf{T}\theta \right) e^{iu^\mathsf{T} X} \right| \right\}, \end{aligned}$$

where

$$C_r = \sup_{u \in \mathbb{R}^d} \|u\|_1^r \psi(u).$$

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

It remains to calculate the value of the constant with $\psi(u) = e^{-\frac{1}{2}u^{\mathsf{T}}u}$. The first-order conditions are given by

$$u_j\left(\sum_{l=1}^{d} u_l\right) = r, \qquad j = 1, \ldots, d,$$

and we find that at the optimal solution $u_1 = \cdots = u_d = \sqrt{r/d}$. Thus, $\|u\|_1 = \sqrt{rd}$ and $u^{\mathsf{T}}u = r$, which completes the proof. □

LEMMA A.5. *Suppose that $V = (Y, X) \in \mathbb{R} \times \mathbb{R}^d$ is a zero-mean random vector and that the moment conditions in Lemma A.2 hold. The functionals $\boldsymbol{b}_r(V)$ and $\boldsymbol{\zeta}_r(V)$, $r \geq 1$, are nonnegative and equal zero if and only if $\mathbb{E}(Y|X) = X^{\mathsf{T}}\theta^*$ almost surely.*

Proof. Lemma A.3 shows that $\theta = \theta^*$ is necessary for boundedness of both $\boldsymbol{b}_r(V)$ and $\boldsymbol{\zeta}_r(V)$, $r \geq 1$. The next inequalities follow from Lemmas A.3 and A.4

$$\boldsymbol{\varsigma}(V) \leq C_r \boldsymbol{b}_r(V) \leq C_r A_\delta \boldsymbol{\zeta}_r(V), \tag{A.4}$$

where C_r and A_δ are positive constants.

Using the inequalities in equation (A.4) and Lemma A.1, we conclude that $\boldsymbol{\zeta}_r(V) = 0$ leads to $\mathbb{E}(Y|X) = X^{\mathsf{T}}\theta^*$ almost surely. In the opposite direction, if $\mathbb{E}(Y|X) = X^{\mathsf{T}}\theta^*$ almost surely, then $\mathbb{E}[(Y - X^{\mathsf{T}}\theta^*)f(X)] = 0$ for any measurable function f for which the moment is defined and, therefore, $\boldsymbol{\zeta}_r(V) = 0$. The same statements hold for $\boldsymbol{b}_r(V)$. □

Proof of Theorem 2. From Lemmas A.3 and A.4,

$$\begin{aligned}\boldsymbol{\varsigma}(n^{-1/2}(\Delta V_1 + \cdots + \Delta V_n)) &\leq \boldsymbol{\varsigma}^{\theta^*}(n^{-1/2}(\Delta V_1 + \cdots + \Delta V_n)) \\ &\leq n C_r \boldsymbol{b}_r(n^{-1/2} V_{\Delta t}) \\ &= n^{1-\frac{r+1}{2}} C_r \boldsymbol{b}_r(V_{\Delta t}),\end{aligned}$$

where $C_r = \sup_{u \in \mathbb{R}^d} \|u\|_1^r \psi(u)$. The upper bound in equation (15) follows from Lemmas A.2 and A.3. □

LEMMA A.6. *Suppose $Z = (Z_y, Z_x) \in \mathbb{R} \times \mathbb{R}^d$ is a strictly α-stable random vector with tail parameter $1 < \alpha \leq 2$ and a spectral measure Λ. Denote by $\xi = (y, x) \in \mathbb{R} \times \mathbb{R}^d$ and let $\mathbb{S} = \{z \in \mathbb{R}^{d+1} : \|z\|_2 = 1\}$ be the unit sphere in $\mathbb{R}^{d+1}$. If θ is a fixed vector of parameters, then the conditions*

$$\int_{\mathbb{S}} (y - x^{\mathsf{T}}\theta)\,\mathrm{sign}(u^{\mathsf{T}}x)|u^{\mathsf{T}}x|^{\alpha-1}\Lambda(d\xi) = 0, \quad \forall u \in \mathbb{R}^d$$

$$\int_{\mathbb{S}} (y - x^{\mathsf{T}}\theta)|u^{\mathsf{T}}x|^{\alpha-1}\Lambda(d\xi) = 0, \quad \forall u \in \mathbb{R}^d,$$

are identical with $\mathbb{E}[Z_y|Z_x] \overset{a.s.}{=} Z_x^{\mathsf{T}}\theta$.

Proof. The plan is to calculate the expectation $\mathbb{E}(Z_y - Z_x^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}Z_x}$ and then show that the conditions in equations (24) and (25) are equivalent to requiring that the expectation vanishes for all $u \in \mathbb{R}^d$.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

Denote the ch.f. of $Z = (Z_y, Z_x)$ by $\varphi_Z(v,u) = \mathbb{E}e^{i(vZ_y + u^\mathsf{T} Z_x)}$. The expectation can be expressed in terms of certain derivatives of the ch.f. and the vector of parameters θ

$$\mathbb{E}(Z_y - Z_x^\mathsf{T}\theta)e^{iu^\mathsf{T} Z_x} = \frac{\partial}{\partial v}\varphi_Z(0,u) - \sum_{k=1}^{d}\theta_k \frac{\partial}{\partial u_k}\varphi_Z(0,u).$$

We express the right side as a function of the spectral measure using the explicit form of the ch.f. of stable laws (see Samorodnitsky and Taqqu, 1994, Chap. 2).

Adopt the notation $t = (v,u)$. It is straightforward to calculate the derivative

$$\begin{aligned}\frac{\partial}{\partial t_k}\varphi_Z(t) &= \frac{\partial}{\partial t_k}\exp\left\{-\int_{\mathbb{S}} |t^\mathsf{T}\xi|^\alpha \left(1 - i\,\mathrm{sign}(\xi^\mathsf{T} s)\tan\frac{\pi\alpha}{2}\right)\Lambda(d\xi)\right\} \\ &= \alpha\varphi_Z(t)\left(-\int_{\mathbb{S}} \xi_k\,\mathrm{sign}(t^\mathsf{T}\xi)|t^\mathsf{T}\xi|^{\alpha-1}\Lambda(d\xi) + i\int_{\mathbb{S}}\xi_k|t^\mathsf{T}\xi|^{\alpha-1}\tan\frac{\pi\alpha}{2}\Lambda(d\xi)\right).\end{aligned}$$

Using this expression and $\varphi_Z(0,u) = \varphi_{Z_x}(u)$, for the expectation, we obtain

$$\begin{aligned}\mathbb{E}(Z_y - Z_x^\mathsf{T}\theta)e^{iu^\mathsf{T} Z_x} &= \alpha\varphi_{Z_x}(u)\left(-\int_{\mathbb{S}}(y - x^\mathsf{T}\theta)\,\mathrm{sign}(u^\mathsf{T} x)|u^\mathsf{T} x|^{\alpha-1}\Lambda(d\xi)\right. \\ &\qquad \left. + i\int_{\mathbb{S}}(y - x^\mathsf{T}\theta)|u^\mathsf{T} x|^{\alpha-1}\tan\frac{\pi\alpha}{2}\Lambda(d\xi)\right) \\ &= -\alpha\varphi_{Z_x}(u)(\gamma(u) - i\eta(u)), \qquad \textbf{(A.5)}\end{aligned}$$

where $\xi = (y,x) \in \mathbb{R}\times\mathbb{R}^d$. As a result, $\mathbb{E}(Z_y - Z_x^\mathsf{T}\theta)e^{iu^\mathsf{T} Z_x} = 0$ for all $u \in \mathbb{R}^d$ if and only if $\gamma(u) = \eta(u) = 0$ for all $u \in \mathbb{R}^d$ because $\varphi_{Z_x}(u)$ is a complex exponential and does not vanish. The proof is now complete. □

LEMMA A.7. *Denote by $V = (Y,X) \in \mathbb{R}\times\mathbb{R}^d$ and $Z = (Z_y, Z_x) \in \mathbb{R}\times\mathbb{R}^d$ two arbitrary random vectors and assume that for some θ, the next conditions hold*

$$\mathbb{E}\left[(Y - X^\mathsf{T}\theta)X^\nu - (Z_y - Z_x^\mathsf{T}\theta)Z_x^\nu\right] = 0, \quad \forall \nu\colon |\nu| \le k. \qquad \textbf{(A.6)}$$

Define

$$\zeta_r^\theta(V,Z) = \sup_{g\in\mathcal{G}}\left|\int_{\mathbb{R}^{d+1}}(y - x^\mathsf{T}\theta)g(x)d(F - G)\right|, \qquad \textbf{(A.7)}$$

where $\mathcal{G}$ is the class defined in equation (A.2). The following bound holds:

$$\zeta_r^\theta(V,Z) \le \frac{\Gamma(\delta+1)}{\Gamma(r+1)}\int_{\mathbb{R}^{d+1}}|y - x^\mathsf{T}\theta|\,\|x\|_1^r\,|F - G|\,(d\xi),$$

where $\xi = (y,x)$, $r \ge 1$, $r = k + \delta$, where k is an integer, $\delta \in (0,1]$, and F and G are the distribution functions of V and Z, respectively.

Proof. The proof follows the same steps as that of Lemma A.2. The functions from the class $\mathcal{G}$ can be written in terms of their Taylor expansion at $x = 0$

$$g(x) = \sum_{|\nu|\le k-1}\frac{\partial^\nu g(0)}{\nu!}x^\nu + k\sum_{|\nu|=k}\frac{x^\nu}{\nu!}\int_0^1(1-u)^{k-1}\partial^\nu g(ux)du.$$

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

To simplify the notation, denote the two residuals by $\varepsilon = Y - X^{\mathsf{T}}\theta$ and $\eta = Z_y - Z_x^{\mathsf{T}}\theta$. For the expectation in the definition of $\boldsymbol{\zeta}_r$, we can write

$$
\begin{aligned}
I &= \mathbb{E}\left(\varepsilon g(X) - \eta g(Z_x)\right) \\
&= \sum_{|\nu|\le k-1} \frac{\partial^\nu g(0)}{\nu!}\mathbb{E}\left[\varepsilon X^\nu - \eta Z_x^\nu\right] \\
&\quad + k\sum_{|\nu|=k}\int_0^1 \frac{(1-u)^{k-1}}{\nu!}\mathbb{E}\left(\varepsilon X^\nu \partial^\nu g(uX) - \eta Z_x^\nu \partial^\nu g(uZ_x)\right) du.
\end{aligned}
$$

The first sum vanishes due to the condition in equation (A.6). The expectations in the integral can be written in the following way:

$$
\mathbb{E}\left(\varepsilon X^\nu \partial^\nu g(uX) - \eta Z_x^\nu \partial^\nu g(uZ_x)\right) = \int_{\mathbb{R}^{d+1}} (y - x^{\mathsf{T}}\theta)x^\nu \left[\partial^\nu g(ux) - \partial^\nu g(0)\right] d(F-G),
$$

in which F and G are the distribution functions of V and Z and

$$
\partial^\nu g(0)\int_{\mathbb{R}^{d+1}} (y - x^{\mathsf{T}}\theta)x^\nu d(F-G) = \partial^\nu g(0)\mathbb{E}\left(\varepsilon X^\nu - \eta Z_x^\nu\right) = 0
$$

follows from equation (A.6) with $|\nu| = k$. As a result, we obtain

$$
I = k\sum_{|\nu|=k}\int_0^1 \frac{(1-u)^{k-1}}{\nu!}\left[\int_{\mathbb{R}^{d+1}} (y - x^{\mathsf{T}}\theta)x^\nu \left(\partial^\nu g(ux) - \partial^\nu g(0)\right) d(F-G)\right] du.
$$

The absolute value $|I|$ is bounded in the same way as in the proof of Lemma A.2. As a result, for any $g \in \mathcal{G}$, we obtain

$$
|\mathbb{E}\varepsilon g(X) - \mathbb{E}\eta g(Z_x)| = |I| \le \frac{\Gamma(\delta+1)}{\Gamma(r+1)}\int_{\mathbb{R}^{d+1}} |y - x^{\mathsf{T}}\theta|\,\|x\|_1^{k+\delta}\,|F-G|\,(d\xi),
$$

where $\xi = (y,x)$ and the proof is completed after taking the supremum over all $g \in \mathcal{G}$. □

LEMMA A.8. *Denote by $V = (Y,X) \in \mathbb{R}\times\mathbb{R}^d$ and $Z = (Z_y, Z_x) \in \mathbb{R}\times\mathbb{R}^d$ two arbitrary random vectors. The following inequality holds:*

$$
\boldsymbol{b}_r^\theta(V,Z) \le A_\delta \boldsymbol{\zeta}_r^\theta(V,Z),
$$

where $r \ge 1$, $r = k+\delta$, in which k is an integer, $\delta \in (0,1]$, and θ is fixed. The next scaling property holds

$$
\boldsymbol{b}_r^\theta(aV, aZ) = |a|^{r+1}\boldsymbol{b}_r^\theta(V,Z),
$$

where $a \in \mathbb{R}$ is a constant.

Proof. The proof is the same as that of Lemma A.3 and relies on the properties of the functional $\boldsymbol{\zeta}_r^\theta(V,Z)$ defined in equation (A.7), which are established in Lemma A.7. □

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

LEMMA A.9. *Suppose $Z_j = (Z_{y,j}, Z_{x,j}) \in \mathbb{R} \times \mathbb{R}^d$ are independent copies of a strictly α-stable random vector $Z = (Z_y, Z_x)$ with $1 < \alpha \leq 2$. Then,*

$$\mathbb{E}\left(\frac{1}{n^{1/\alpha}}\sum_{j=1}^{n}\left(Z_{y,j} - Z_{x,j}^{\mathsf{T}}\theta\right)e^{iu^{\mathsf{T}}\frac{Z_{x,j}}{n^{1/\alpha}}}\right) = -\alpha\varphi_{Z_x}(n^{-1/\alpha}u)\omega_{\alpha,\Lambda}(u),$$

where $\varphi_{Z_x}(u) = \mathbb{E}e^{iu^{\mathsf{T}}Z_x}$ is the ch.f. of Z_x,

$$\omega_{\alpha,\Lambda}(u) = \int_{\mathbb{S}}(y - x^{\mathsf{T}}\theta)(u^{\mathsf{T}}x)^{\alpha-1}\Lambda(d\xi) - i\int_{\mathbb{S}}(y - x^{\mathsf{T}}\theta)|u^{\mathsf{T}}x|^{\alpha-1}\tan\frac{\pi\alpha}{2}\Lambda(d\xi),$$

in which $\xi = (y, x) \in \mathbb{R} \times \mathbb{R}^d$, $(x)^\alpha = \text{sign}(x)|x|^\alpha$, and Λ is the spectral measure of Z.

Proof. Using independence, it is easy to see that

$$\mathbb{E}\left(\frac{1}{n^{1/\alpha}}\sum_{j=1}^{n}\left(Z_{y,j} - Z_{x,j}^{\mathsf{T}}\theta\right)e^{iu^{\mathsf{T}}\frac{Z_{x,j}}{n^{1/\alpha}}}\right) = n^{\frac{\alpha-1}{\alpha}}\mathbb{E}(Z_y - Z_x^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{Z_x}{n^{1/\alpha}}}.$$

The expectation on the right side is treated similarly as in the proof of Lemma A.6. Using equation (A.5), we obtain

$$\begin{aligned} n^{\frac{\alpha-1}{\alpha}}\mathbb{E}(Z_y - Z_x^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{Z_x}{n^{1/\alpha}}} &= -n^{\frac{\alpha-1}{\alpha}}\alpha\varphi_{Z_x}\left(\frac{u}{n^{1/\alpha}}\right)\left[\gamma\left(\frac{u}{n^{1/\alpha}}\right) - i\eta\left(\frac{u}{n^{1/\alpha}}\right)\right] \\ &= -\alpha\varphi_{Z_x}\left(\frac{u}{n^{1/\alpha}}\right)(\gamma(u) - i\eta(u)), \end{aligned}$$

which completes the proof. □

Proof of Theorem 3. To simplify the expressions in the proof, we modify the notation in the following way. The two Lévy processes are denoted by $V(t) = (Y(t), X(t))$ and $Z(t) = (Z_y(t), Z_x(t))$. We construct n intervals of equal fixed length $t_i - t_{i-1} = \Delta t$, where $t_0 = 0$ and $i = 1, \ldots, n$. From the definition of Lévy processes,

$$V(n\Delta t) = \sum_{i=1}^{n}\Delta V(t_i - t_{i-1}) = \sum_{i=1}^{n}V_i,$$

in which $V_1, \ldots, V_n$ are independent copies of a random vector V_0 with a distribution function F because the increments in the sum are considered over intervals of the same length. The notation for the process $Z(t)$ is analogous; that is, $Z_1, \ldots, Z_n$ is a sequence of independent copies of a random vector Z_0 with a distribution function G representing increments over intervals of the same length Δt.

By assumption, the vector Z_0 has a strictly α-stable distribution and satisfies the stability property $Z_0 \stackrel{d}{=} n^{-1/\alpha}\sum_{i=1}^{n}Z_i$. The stability property implies that $\mathbb{E}[g(Z_0)] = \mathbb{E}[g(n^{-1/\alpha}\sum_{i=1}^{n}Z_i)]$, where g is a measurable function and the moment is assumed to exist.

Using Assumption 1[†], there exists θ_α^* which satisfies equations (24) and (25) and we set $\theta = \theta_\alpha^*$ everywhere in the proof. With that choice of θ, by Theorem 1 and Lemma A.6, the conditional mean of Z_0 is linear and $\mathbb{E}(Z_{y,0} - Z_{x,0}^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}Z_{x,0}} = 0$ for all u. Consider the

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

following difference:

$$\Delta_n := \mathbb{E}\left(\frac{1}{n^{1/\alpha}}\left(Y(n\Delta t) - X(n\Delta t)^{\mathsf{T}}\theta\right)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}X(n\Delta t)}\right) - \mathbb{E}(Z_{y,0} - Z_{x,0}^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}Z_{x,0}}$$
$$= \mathbb{E}\left(\frac{1}{n^{1/\alpha}}\sum_{j=1}^{n}(Y_j - X_j^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k=1}^{n}X_k}\right)$$
$$- \mathbb{E}\left(\frac{1}{n^{1/\alpha}}\sum_{j=1}^{n}(Z_{y,j} - Z_{x,j}^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k=1}^{n}Z_{x,k}}\right),$$

where we also applied the stability property. In each summand, we isolate the j-th term in the sum in the exponent and use the independence property

$$\Delta_n = \mathbb{E}\left(\frac{1}{n^{1/\alpha}}\sum_{j=1}^{n}(Y_j - X_j^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}X_j}\right)\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}X_k}$$
$$- \mathbb{E}\left(\frac{1}{n^{1/\alpha}}\sum_{j=1}^{n}(Z_{y,j} - Z_{x,j}^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}Z_{x,j}}\right)\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}Z_{x,k}}$$
$$= I_n\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}X_k} - J_n\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}Z_{x,k}}.$$

Next, we add and subtract J_n to the term I_n

$$\Delta_n = (I_n - J_n)\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}X_k} - J_n\left(\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}Z_{x,k}} - \mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}X_k}\right).$$

The second term is related to the distance between the ch.f. of the normalized sums, while the first term is handled separately. The absolute value of the left side is bounded by

$$|\Delta_n| \leq |I_n - J_n|\left|\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}X_k}\right| + |J_n|\left|\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}Z_{x,k}} - \mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}X_k}\right|.$$

Next, multiply both sides by $\psi(u)$,

$$\psi(u)|\Delta_n| \leq \psi(u)|I_n - J_n|\left|\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}X_k}\right|$$
$$+ \psi(u)|J_n|\left|\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}^{n}Z_{x,k}} - \mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}X_k}\right| \tag{A.8}$$
$$= \psi(u)\,\|u\|_1^r\,\|u\|_1^{-r}\,|I_n - J_n|\left|\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}X_k}\right|$$
$$+ \psi(u)\,\|u\|_1^{\alpha-1}\,\|u\|_1^{1-\alpha}\,|J_n|$$
$$\times\left|\mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}Z_{x,k}} - \mathbb{E}e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}\sum_{k\neq j}X_k}\right|$$
$$\leq C_r\sup_u\|u\|_1^{-r}\,|I_n - J_n| + 2C_{\alpha-1}\left(\sup_u\|u\|_1^{1-\alpha}\,|J_n|\right), \tag{A.9}$$

where $C_p = \sup_u\|u\|_1^p\,\psi(u)$ and we used the inequality $|e^{ix}| \leq 1$.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

It remains to show the rate of decay of the right side of (A.9). We begin with the last term. From the result in Lemma A.9, we obtain

$$\|u\|_1^{1-\alpha}\,|J_n| \le \alpha\,\|u\|_1^{1-\alpha}\,|\omega_{\alpha,\Lambda}(u)|,$$

in which Λ denotes the spectral measure. From Assumption $1^{\dagger}$, θ satisfies equations (24) and (25). The two equations guarantee $|\omega_{\alpha,\Lambda}(u)| = 0$ and, as a consequence, the second term in equation (A.9) vanishes.

The first term in (A.9) contains the expression $I_n - J_n$ which we represent in the following manner:

$$\begin{aligned}
I_n - J_n &= \mathbb{E}\left(\frac{1}{n^{1/\alpha}}\sum_{j=1}^{n}(Y_j - X_j^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}X_j}\right) - \mathbb{E}\left(\frac{1}{n^{1/\alpha}}\sum_{j=1}^{n}(Z_{y_j} - Z_{x,j}^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}Z_{x,j}}\right)\\
&= \frac{1}{n^{1/\alpha}}\sum_{j=1}^{n}\left(\mathbb{E}(Y_j - X_j^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}X_j} - \mathbb{E}(Z_{y_j} - Z_{x,j}^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}Z_{x,j}}\right)\\
&= n^{1-\frac{1}{\alpha}}\left(\mathbb{E}(Y_0 - X_0^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}X_0} - \mathbb{E}(Z_{y,0} - Z_{x,0}^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}Z_{x,0}}\right)\\
&= n^{1-\frac{1}{\alpha}}\int_{\mathbb{R}^{d+1}}(y - x^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}x}d(F-G).
\end{aligned}$$

Thus, for the first term in (A.9), we obtain

$$\begin{aligned}
\|u\|_1^{-r}\,|I_n - J_n| &= n^{1-\frac{1}{\alpha}}\,\|u\|_1^{-r}\left|\int_{\mathbb{R}^{d+1}}(y - x^{\mathsf{T}}\theta)e^{iu^{\mathsf{T}}\frac{1}{n^{1/\alpha}}x}d(F-G)\right|\\
&= n^{1-\frac{r+1}{\alpha}}\,\|v\|_1^{-r}\left|\int_{\mathbb{R}^{d+1}}(y - x^{\mathsf{T}}\theta)e^{iv^{\mathsf{T}}x}d(F-G)\right|,
\end{aligned}$$

where $v = u/n^{1/\alpha}$. As a result,

$$\begin{aligned}
\sup_u \|u\|_1^{-r}\,|I_n - J_n| &= n^{1-\frac{r+1}{\alpha}}\sup_v \|v\|_1^{-r}\left|\int_{\mathbb{R}^{d+1}}(y - x^{\mathsf{T}}\theta)e^{iv^{\mathsf{T}}x}d(F-G)\right|\\
&= n^{1-\frac{r+1}{\alpha}}\,\boldsymbol{b}_r^{\theta}(V_0, Z_0). \qquad \textbf{(A.10)}
\end{aligned}$$

We need to provide conditions under which the right side in the equation above is finite. Applying Lemmas A.7 and A.8, we obtain

$$\boldsymbol{b}_r^{\theta}(V_0, Z_0) \le A_\delta\,\frac{\Gamma(\delta+1)}{\Gamma(r+1)}\int_{\mathbb{R}^{d+1}}|y - x^{\mathsf{T}}\theta|\,\|x\|_1^r\,|F-G|(d(y,x)),$$

where $r = k + \delta$, $k \ge 1$ is an integer and $\delta \in (0, 1]$.

The final form of the upper bound in (35) is obtained by plugging the expressions in equation (A.10) into (A.9)

$$\psi(u)|\Delta_n| \le C_r n^{1-\frac{r+1}{\alpha}}\,\boldsymbol{b}_r^{\theta}(V_0, Z_0).$$

□

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

Proof of Theorem 4. The proof uses the notation adopted in the proof of Theorem 3. Starting from the expression for $\psi(u)|\Delta_n|$ provided in equation (A.8), we obtain

$$\begin{aligned}
\psi(u)|\Delta_n| &\leq \psi(u)|I_n - J_n| \left| \mathbb{E} e^{iu^{\mathsf{T}} \frac{1}{n^{1/\alpha}} \sum_{k \neq j} X_k} \right| \\
&\quad + \psi(u)|J_n| \left| \mathbb{E} e^{iu^{\mathsf{T}} \frac{1}{n^{1/\alpha}} \sum_{k \neq j}^{n} Z_{x,k}} - \mathbb{E} e^{iu^{\mathsf{T}} \frac{1}{n^{1/\alpha}} \sum_{k \neq j} X_k} \right| \\
&= \psi(u) \|u\|_1^r \|u\|_1^{-r} |I_n - J_n| \left| \mathbb{E} e^{iu^{\mathsf{T}} \frac{1}{n^{1/\alpha}} \sum_{k \neq j} X_k} \right| \\
&\quad + \psi(u) \|u\|_1^{s+\alpha-1} \|u\|_1^{1-\alpha} |J_n| \\
&\quad \times \|u\|_1^{-s} \left| \mathbb{E} e^{iu^{\mathsf{T}} \frac{1}{n^{1/\alpha}} \sum_{k \neq j} Z_{x,k}} - \mathbb{E} e^{iu^{\mathsf{T}} \frac{1}{n^{1/\alpha}} \sum_{k \neq j} X_k} \right| \\
&\leq C_r \sup_u \|u\|_1^{-r} |I_n - J_n| \\
&\quad + C_{s+\alpha-1} \left(\sup_u \|u\|_1^{1-\alpha} |J_n| \right) \boldsymbol{\chi}_s \left(\frac{1}{n^{1/\alpha}} \sum_{k=1}^{n-1} X_k, \frac{1}{n^{1/\alpha}} \sum_{k=1}^{n-1} Z_{x,k} \right), \qquad \textbf{(A.11)}
\end{aligned}$$

where $C_p = \sup_u \|u\|_1^p \psi(u)$ and $\boldsymbol{\chi}_s$ is defined in equation (32).

It remains to show the rate of decay of the right side of (A.11). From the result in Lemma A.9, we obtain

$$\begin{aligned}
\|u\|_1^{1-\alpha} |J_n| \leq \alpha \|u\|_1^{1-\alpha} |\omega_{\alpha,\Lambda}(u)| &= \alpha \left| \omega_{\alpha,\Lambda} \left(\frac{u}{\|u\|_1} \right) \right| \\
&\leq \alpha \sup_{v \in B} |\omega_{\alpha,\Lambda}(v)| | \\
&= \alpha D_\alpha, \qquad \textbf{(A.12)}
\end{aligned}$$

in which $B = \{z \in \mathbb{R}^d \colon \|z\|_1 = 1\}$ and Λ denotes the spectral measure. For the first term in (A.11), we use the expression derived in equation (A.10). The final form of the upper bound in (35) is obtained by plugging in the expressions in equations (A.12) and (A.10) into (A.11). From the ideality properties of the $\boldsymbol{\chi}_s$ metric (Rachev et al., 2013, Chap. 15),

$$\begin{aligned}
\boldsymbol{\chi}_s \left(\frac{1}{n^{1/\alpha}} \sum_{k=1}^{n-1} X_k, \frac{1}{n^{1/\alpha}} \sum_{k=1}^{n-1} Z_{x,k} \right) &\leq (n-1) n^{-s/\alpha} \boldsymbol{\chi}_s(X_0, Z_{x,0}) \\
&\leq n^{1-s/\alpha} \boldsymbol{\chi}_s(X_0, Z_{x,0}),
\end{aligned}$$

which completes the proof.

The conditions in equation (33) guarantee boundedness of $\boldsymbol{\chi}_s$. We now demonstrate that they also guarantee boundedness of $\boldsymbol{b}_{s-1}$. If we represent $\xi = (y, x_1, \ldots, x_d)$, then from the first condition in (33), we obtain

$$\int_{\mathbb{R}^{d+1}} y x^{\nu} d(F - G) = 0 \quad \text{and} \quad \int_{\mathbb{R}^{d+1}} x_j x^{\nu} d(F - G) = 0, \quad \forall \nu \colon |\nu| \leq k - 1,$$

where $j = 1, \ldots, d$. Then, any linear combination of these integrals equals zero and, thus, the condition in equation (A.6) holds for all multi-indices ν such that $|\nu| \leq k - 1$. The absolute

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

pseudomoment $\boldsymbol{v}_s(V_0, Z_0)$ is assumed bounded in (33). Combining this with the inequality in equation (29) and Lemma A.8 shows that $\boldsymbol{b}_{s-1}$ is bounded as well. □

Proof of Theorem 5. Assumption 1‡ ensures that the least-squares estimator for the vector $\sum_{l=1}^{n} V_l$ is θ^* provided in Assumption 2‡. Repeating the steps in Lemma A.4 while bearing in mind $V_1, V_2, \dots$ are non-identically distributed yields

$$\begin{aligned}
\boldsymbol{\zeta}^{\theta^*}\left(\sum_{l=1}^{n} V_l\right) &= \sup_u \psi(u) \left| \mathbb{E}\left(\sum_{l=1}^{n}\left(Y_l - X_l^{\mathsf{T}}\theta^*\right)\right) e^{iu^{\mathsf{T}}\left(\sum_{m=1}^{n} X_m\right)} \right| \\
&\leq \sum_{l=1}^{n} \sup_u \psi(u) \left| \mathbb{E}\left(Y_l - X_l^{\mathsf{T}}\theta^*\right) e^{iu^{\mathsf{T}}X_l} \right| \\
&\leq \left(\sup_u \psi(u) \|u\|_1^r\right) \sum_{l=1}^{n} \sup_u \|u\|_1^{-r} \left| \mathbb{E}\left(Y_l - X_l^{\mathsf{T}}\theta^*\right) e^{iu^{\mathsf{T}}X_l} \right| \\
&= C_r \sum_{l=1}^{n} \boldsymbol{b}_r(V_l).
\end{aligned}$$

Assumption 2‡ ensures that the quantities $\boldsymbol{b}_r(V_l)$ are finite. Combining this estimate with Lemma A.3, we obtain

$$\boldsymbol{\zeta}\left(n^{-1/2}\sum_{l=1}^{n} V_l\right) \leq \boldsymbol{\zeta}^{\theta^*}\left(n^{-1/2}\sum_{l=1}^{n} V_l\right) \leq C_r n^{-\frac{r+1}{2}} \sum_{l=1}^{n} \boldsymbol{b}_r(V_l).$$

Applying Lemmas A.2 and A.3 to each term $\boldsymbol{b}_r(V_l)$ leads to an upper bound analogous to the one in equation (15). Summing up all terms and applying the inequality in equation (51) from Assumption 2‡ yields,

$$\sum_{l=1}^{n} \boldsymbol{b}_r(V_l) \leq A_\delta \frac{\Gamma(\delta+1)}{\Gamma(r+1)} \sum_{l=1}^{n} \mathbb{E}\left(|Y_l - X_l^{\mathsf{T}}\theta^*| \, \|X_l\|_1^r\right) \leq n A_\delta \frac{\Gamma(\delta+1)}{\Gamma(r+1)} D,$$

which completes the proof. □

FUNDING STATEMENT

The author declares that no specific funding has been received for this article.

COMPETING INTEREST STATEMENT

The author declares that no competing interests exist.

REFERENCES

Bianchi, M. L., Stoyanov, S. V., Tassinari, G. L., Fabozzi, F. J., & Focardi, S. M. (2019). *Handbook of heavy-tailed distributions in asset management and risk management*. World Scientific.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, 20(1), 105–134.

Bierens, H. J. (2016). *Econometric model specification*. World Scientific.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

Bierens, H. J., & Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica*, 65, 1129–1151.
Butzer, P., Hahn, L., & Roeckerath, M. T. (1983). Central limit theorem and weak law of large numbers with rates for martingales in Banach spaces. *Journal of Multivariate Analysis*, 13(2), 287–301.
Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57–82.
Cartan, H. (1971). *Differential calculus*. Houghton Mifflin.
Cont, R., & Tankov, P. (2003). *Financial modelling with jump processes*. Chapman and Hall/CRC.
Demarta, S., & McNeil, A. J. (2005). The t copula and related copulas. *International Statistical Review*, 73(1), 111–129.
Eberlein, E. (2001). Application of generalized hyperbolic Lévy motions to finance. In O. E. Barndorff-Nielsen, T. Mikosch, & S. Resnick (Eds.), *Lévy processes: Theory and applications* (pp. 319–336). Birkhäuser.
Escanciano, J. C. (2006). Goodness-of-fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association*, 101(474), 531–541.
Fama, E., & French, K. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2), 783–803.
Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22.
Fama, E. F., & French, K. R. (2020). Comparing cross-section and time-series factor models. *The Review of Financial Studies*, 33(5), 1891–1926.
Fang, H.-B., Fang, K.-T., & Kotz, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82(1), 1–16.
Feng, G., Giglio, S., & Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3), 1327–1370.
Granger, C. W. (1987). Implications of aggregation with common factors. *Econometric Theory*, 3(2), 208–222.
Granger, C. W., & Lee, T.-H. (1999). The effect of aggregation on nonlinearity. *Econometric Reviews*, 18(3), 259–269.
Haeusler, E. (1988). On the rate of convergence in the central limit theorem for martingales with discrete and continuous time. *The Annals of Probability*, 16(1), 275–299.
Harvey, C. R., Liu, Y., & Zhu, H. (2016). … and the cross-section of expected returns. *The Review of Financial Studies*, 29(1), 5–68.
Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *The Review of Financial Studies*, 33(5), 2019–2133.
Jagannathan, R., & Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *The Journal of Finance*, 51(1), 3–53.
Jamalizadeh, A., & Balakrishnan, N. (2019). Conditional distributions of multivariate normal mean–variance mixtures. *Statistics & Probability Letters*, 145, 312–316.
McNeil, A., Frey, R., & Embrechts, P. (2005). *Quantitative risk management: Concepts Techniques and Tools*. Princeton University Press.
Miller, G. (1978). Properties of certain symmetric stable distributions. *Journal of Multivariate Analysis*, 8(3), 346–360.
Mourrat, J.-C. (2013). On the rate of convergence in the martingale central limit theorem. *Bernoulli*, 19(2), 633–645.
Rachev, S., Klebanov, L., Stoyanov, S., & Fabozzi, F. (2013). *The methods of distances in the theory of probability and statistics*. Springer.
Rachev, S. T., & Mittnik, S. (2000). *Stable Paretian models in finance*. John Wiley & Sons.
Račkauskas, A. (1991). On the convergence rate in martingale CLT in Hilbert spaces. *Lithuanian Mathematical Journal*, 31(3), 345–355.
Samorodnitsky, G., & Taqqu, M. (1994). *Stable non-Gaussian random processes*. Chapman & Hall.
Sato, K.-I. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge University Press.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press

Stinchcombe, M. B., & White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory*, 14(3), 295–325.

Stock, J. H., & Watson, M. W. (2011). Dynamic factor models. In M. P. Clements, & D. F. Hendry (Eds.), *The Oxford handbook of economic forecasting* (pp. 35–60). Oxford University Press.

Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics*, 25(2), 613–641.

Tikhomirov, A. N. (1991). On the accuracy of normal approximation of the probability of hitting a ball of sums of weakly dependent Hilbert space valued random variables I. *Theory of Probability & Its Applications*, 36, 738–751.

Tikhomirov, A. N. (1994). On the accuracy of normal approximation of the probability of hitting a ball of sums of weakly dependent Hilbert space valued random variables II. *Theory of Probability & Its Applications*, 38(1), 80–94.

Zolotarev, V. M. (2011). *Modern theory of summation of random variables*. De Gruyter.

https://doi.org/10.1017/S0266466625100261 Published online by Cambridge University Press