The Annals of Statistics
2011, Vol. 39, No. 4, 2205–2242
DOI: 10.1214/11-AOS901

STATISTICAL INFERENCE FOR TIME-CHANGED LÉVY PROCESSES VIA COMPOSITE CHARACTERISTIC FUNCTION ESTIMATION

BY DENIS BELOMESTNY

Duisburg-Essen University

In this article, the problem of semi-parametric inference on the parameters of a multidimensional Lévy process L_t with independent components based on the low-frequency observations of the corresponding time-changed Lévy process $L_{\mathcal{T}(t)}$, where $\mathcal{T}$ is a nonnegative, nondecreasing real-valued process independent of L_t, is studied. We show that this problem is closely related to the problem of composite function estimation that has recently gotten much attention in statistical literature. Under suitable identifiability conditions, we propose a consistent estimate for the Lévy density of L_t and derive the uniform as well as the pointwise convergence rates of the estimate proposed. Moreover, we prove that the rates obtained are optimal in a minimax sense over suitable classes of time-changed Lévy models. Finally, we present a simulation study showing the performance of our estimation algorithm in the case of time-changed Normal Inverse Gaussian (NIG) Lévy processes.

1. Introduction.

The problem of nonparametric statistical inference for jump processes or more generally for semimartingale models has long history and goes back to the works of Rubin and Tucker (1959) and Basawa and Brockwell (1982). In the past decade, one has witnessed the revival of interest in this topic which is mainly related to a wide availability of financial and economical time series data and new types of statistical issues that have not been addressed before. There are two major strands of recent literature dealing with statistical inference for semimartingale models. The first type of literature considers the so-called high-frequency setup, where the asymptotic properties of the corresponding estimates are studied under the assumption that the frequency of observations tends to infinity. In the second strand of literature, the frequency of observations is assumed to be fixed (the so-called low-frequency setup) and the asymptotic analysis is done under the premiss that the observational horizon tends to infinity. It is clear that none of the above asymptotic hypothesis can be perfectly realized on real data and they can only serve as a convenient approximation, as in practice the frequency of

Received December 2010; revised April 2011.
[1] Supported in part by SFB 649 "Economic Risk."
MSC2010 subject classifications. Primary 62F10; secondary 62J12, 62F25, 62H12.
Key words and phrases. Time-changed Lévy processes, dependence, pointwise and uniform rates of convergence, composite function estimation.

observations and the horizon are always finite. The present paper studies the problem of statistical inference for a class of semimartingale models in low-frequency setup.

Let $X = (X_t)_{t \geq 0}$ be a stochastic process valued in $\mathbb{R}^d$ and let $\mathcal{T} = (\mathcal{T}(s))_{s \geq 0}$ be a nonnegative, nondecreasing stochastic process not necessarily independent of X with $\mathcal{T}(0) = 0$. A time-changed process $Y = (Y_s)_{s \geq 0}$ is then defined as $Y_s = X_{\mathcal{T}(s)}$. The process $\mathcal{T}$ is usually referred to as time change. Even in the case of the one-dimensional Brownian motion X, the class of time-changed processes $X_{\mathcal{T}}$ is very large and basically coincides with the class of all semimartingales [see, e.g., Monroe (1978)]. In fact, the construction in Monroe (1978) is not direct, meaning that the problem of specification of different models with the specific properties remains an important issue. For example, the base process X can be assumed to possess some independence property (e.g., X may have independent components), whereas a nonlinear time change can induce deviations from the independence. Along this line, the time change can be used to model dependence for stochastic processes. In this work, we restrict our attention to the case of time-changed Lévy processes, that is, the case where $X = L$ is a multivariate Lévy process and $\mathcal{T}$ is an independent of L time change. Time-changed Lévy processes are one step further in increasing the complexity of models in order to incorporate the so-called stylized features of the financial time series, like volatility clustering [for more details, see Carr et al. (2003)]. This type of processes in the case of the one-dimensional Brownian motion was first studied by Bochner (1949). Clark (1973) introduced Bochner's time-changed Brownian motion into financial economics: he used it to relate future price returns of cotton to the variations in volume during different trading periods. Recently, a number of parametric time-changed Lévy processes have been introduced by Carr et al. (2003), who model the stock price S_t by a geometric time-changed Lévy model

$$S_t = S_0 \exp(L_{\mathcal{T}(t)}),$$

where L is a Lévy process and $\mathcal{T}(t)$ is a time change of the form

$$\mathcal{T}(t) = \int_0^t \rho(u)\, du \tag{1.1}$$

with $\{\rho(u)\}_{u \geq 0}$ being a positive mean-reverting process. Carr et al. (2003) proposed to model $\rho(u)$ via the Cox–Ingersoll–Ross (CIR) process. Taking different parametric Lévy models for L (such as the normal inverse Gaussian or the variance Gamma processes) results in a wide range of processes with rather rich volatility structure (depending on the rate process ρ) and various distributional properties (depending on the specification of L). From statistical point of view, any parametric model (especially one using only few parameters) is prone to misspecification problems. One approach to deal with the misspecification issue is to adopt the general nonparametric models for the functional parameters of the underlying process. This may reduce the estimation bias resulting from an inadequate parametric model. In the case of time-changed Lévy models, there are two natural nonparametric parameters: Lévy density ν, which determines the jump dynamics of the process L and the marginal distribution of the process $\mathcal{T}$.

In this paper, we study the problem of statistical inference on the characteristics of a multivariate Lévy process L with independent components based on low-frequency observations of the time-changed process $Y_t = L_{\mathcal{T}(t)}$, where $\mathcal{T}(t)$ is a time change process independent of L with strictly stationary increments. We assume that the distribution of $\mathcal{T}(t)$ is unknown, except of its mean value. This problem is rather challenging and has not been yet given attention in the literature, except for the special case of $\mathcal{T}(t) \equiv t$ [see, e.g., Neumann and Reiß (2009) and Comte and Genon-Catalot (2010)]. In particular, the main difficulty in constructing nonparametric estimates for the Lévy density ν of L lies in the fact that the jumps are unobservable variables, since in practice only discrete observations of the process Y are available. The more frequent the observations, the more relevant information about the jumps of the underlying process, and hence, about the Lévy density ν are contained in the sample. Such high-frequency based statistical approach has played a central role in the recent literature on nonparametric estimation for Lévy type processes. For instance, under discrete observations of a pure Lévy process L_t at times $t_j = j\Delta$, $j = 0, \dots, n$, Woerner (2003) and Figueroa-López (2004) proposed the quantity

$$\widehat{\beta}(f) = \frac{1}{n\Delta} \sum_{k=1}^{n} f(L_{t_k} - L_{t_{k-1}})$$

as a consistent estimator for the functional

$$\beta(f) = \int f(x)\nu(x)\,dx,$$

where f is a given "test function." Turning back to the time-changed Lévy processes, it was shown in Figueroa-López (2009) [see also Rosenbaum and Tankov (2010)] that in the case, where the rate process ρ in (1.1) is a positive ergodic diffusion independent of the Lévy process L, $\widehat{\beta}(f)$ is still a consistent estimator for $\beta(f)$ up to a constant, provided the time horizon $n\Delta$ and the sampling frequency Δ^{-1} converge to infinite at suitable rates. In the case of low-frequency data (Δ is fixed), we cannot be sure to what extent the increment $L_{t_k} - L_{t_{k-1}}$ is due to one or several jumps or just to the diffusion part of the Lévy process so that at first sight it may appear surprising that some kind of inference in this situation is possible at all. The key observation here is that for any bounded "test function" f

$$\frac{1}{n} \sum_{j=1}^{n} f\big(L_{\mathcal{T}(t_j)} - L_{\mathcal{T}(t_{j-1})}\big) \to \mathrm{E}_{\pi}\big[f\big(L_{\mathcal{T}(\Delta)}\big)\big], \qquad n \to \infty, \tag{1.2}$$

provided the sequence $\mathcal{T}(t_j) - \mathcal{T}(t_{j-1})$, $j = 1, \dots, n$, is stationary and ergodic with the invariant stationary distribution π. The limiting expectation in (1.2) is then given by

$$\mathrm{E}_\pi\big[f\big(L_{\mathcal{T}(\Delta)}\big)\big] = \int_0^\infty \mathrm{E}[f(L_s)]\pi(ds).$$

Taking $f(z) = f_u(z) = \exp(\mathrm{i}u^\top z), u \in \mathbb{R}^d$, and using the independence of L and $\mathcal{T}$, we arrive at the following representation for the c.f. of $L_{\mathcal{T}(s)}$:

$$\mathrm{E}\big[\exp\big(\mathrm{i}u^\top L_{\mathcal{T}(\Delta)}\big)\big] = \int_0^\infty \exp(t\psi(u))\pi(dt) = \mathcal{L}_\Delta(-\psi(u)), \tag{1.3}$$

where $\psi(u) = t^{-1}\log[\mathrm{E}\exp(\mathrm{i}uL_t)]$ is the characteristic exponent of the Lévy process L and $\mathcal{L}_\Delta$ is the Laplace transform of π. In fact, the most difficult part of estimation procedure comes only now and consists in reconstructing the characteristics of the underlying Lévy process L from an estimate for $\mathcal{L}_\Delta(-\psi(u))$. As we will see, the latter statistical problem is closely related to the problem of composite function estimation, which is known to be highly nonlinear and ill-posed. The identity (1.3) also reveals the major difference between high-frequency and low-frequency setups. While in the case of high-frequency data one can directly estimate linear functionals of the Lévy measure ν, under low-frequency observations, one has to deal with nonlinear functionals of ν rendering the underlying estimation problem nonlinear and ill-posed. Last but not least, the increments of time-changed Lévy processes are not any longer independent, hence advanced tools from time series analysis have to be used for the estimation of $\mathcal{L}_\Delta(-\psi(u))$.

The paper is organized as follows. In Section 2.1, we introduce the main object of our study, the time-changed Lévy processes. In Section 2.2, our statistical problem is formulated and its connection to the problem of composite function estimation is established. In Section 2.3, we impose some restrictions on the structure of the time-changed Lévy processes in order to ensure the identifiability and avoid the "curse of dimensionality." Section 3 contains the main estimation procedure. In Section 4, asymptotic properties of the estimates defined in Section 3 are studied. In particular, we derive uniform and pointwise rates of convergence (Sections 4.3 and 4.4, resp.) and prove their optimality over suitable classes of time-changed Lévy models (Section 4.5). Section 4.7 contains some discussion. Finally, in Section 5 we present a simulation study. The rest of the paper contains proofs of the main results and some auxiliary lemmas. In particular, in Section 7.3 a useful inequality on the probability of large deviations for empirical processes in uniform metric for the case of weakly dependent random variables can be found.

2. Main setup.

2.1. *Time-changed Lévy processes.* Let L_t be a d-dimensional Lévy process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the characteristic exponent $\psi(u)$, that is,

$$\psi(u) = t^{-1}\log\mathrm{E}[\exp(\mathrm{i}u^\top L_t)].$$

We know by the Lévy–Khintchine formula that

$$\psi(u) = \mathrm{i}\mu^\top u - \frac{1}{2}u^\top \Sigma u + \int_{\mathbb{R}^d} \bigl(e^{\mathrm{i}u^\top y} - 1 - \mathrm{i}u^\top y \cdot \mathbf{1}_{\{|y|\le 1\}}\bigr)\nu(dy), \tag{2.1}$$

where $\mu \in \mathbb{R}^d$, Σ is a positive-semidefinite symmetric $d \times d$ matrix and ν is a Lévy measure on $\mathbb{R}^d \setminus \{0\}$ satisfying

$$\int_{\mathbb{R}^d \setminus \{0\}} (|y|^2 \wedge 1)\nu(dy) < \infty.$$

A triplet (μ, Σ, ν) is usually called a characteristic triplet of the d-dimensional Lévy process L_t.

Let $t \to \mathcal{T}(t), t \ge 0$ be an increasing right-continuous process with left limits such that $\mathcal{T}(0) = 0$ and for each fixed t, the random variable $\mathcal{T}(t)$ is a stopping time with respect to the filtration $\mathcal{F}$. Suppose furthermore that $\mathcal{T}(t)$ is finite $\mathbb{P}$-a.s. for all $t \ge 0$ and that $\mathcal{T}(t) \to \infty$ as $t \to \infty$. Then the family of $(\mathcal{T}(t))_{t\ge 0}$ defines a random time change. Now consider a d-dimensional process $Y_t := L_{\mathcal{T}(t)}$. The process Y_t is called the time-changed Lévy process. Let us look at some examples. If $\mathcal{T}(t)$ is a Lévy process, then Y_t would be another Lévy process. A more general situation is when $\mathcal{T}(t)$ is modeled by a nondecreasing semimartingale

$$\mathcal{T}(t) = b_t + \int_0^t \int_0^\infty y\rho(dy, ds),$$

where b is a drift and ρ is the counting measure of jumps in the time change. As in Carr and Wu (2004), one can take $b_t = 0$ and consider locally deterministic time changes

$$\mathcal{T}(t) = \int_0^t \rho(s_-)\, ds, \tag{2.2}$$

where ρ is the instantaneous activity rate which is assumed to be nonnegative. When L_t is the Brownian motion and ρ is proportional to the instantaneous variance rate of the Brownian motion, then Y_t is a pure jump Lévy process with the Lévy measure proportional to ρ. Let us now compute the characteristic function of Y_t. Since $\mathcal{T}(t)$ and L_t are independent, we get

$$\phi_Y(u|t) = \mathrm{E}\bigl(e^{\mathrm{i}u^\top L_{\mathcal{T}(t)}}\bigr) = \mathcal{L}_t(-\psi(u)), \tag{2.3}$$

where $\mathcal{L}_t$ is the Laplace transform of $\mathcal{T}(t)$:

$$\mathcal{L}_t(\lambda) = \mathrm{E}\bigl(e^{-\lambda \mathcal{T}(t)}\bigr).$$

2.2. *Statistical problem.* In this paper, we are going to study the problem of estimating the characteristics of the Lévy process L from low-frequency observations $Y_0, Y_\Delta, \dots, Y_{n\Delta}$ of the process Y for some fixed $\Delta > 0$. Moving to the spectral domain and taking into account (2.1), we can reformulate our problem as the problem of semi-parametric estimation of the characteristic exponent ψ under structural assumption (2.1) from an estimate of $\phi_Y(u|\Delta)$ based on $Y_0, Y_\Delta, \ldots, Y_{n\Delta}$. The formula (2.3) shows that the function $\phi_Y(u|\Delta)$ can be viewed as a composite function and our statistical problem is hence closely related to the problem of statistical inference on the components of a composite function. The latter type of problems in regression setup has gotten much attention recently [see, e.g., Horowitz and Mammen (2007) and Juditsky, Lepski and Tsybakov (2009)]. Our problem has, however, some features not reflected in the previous literature. First, the unknown link function $\mathcal{L}_\Delta$, being the Laplace transform of the r.v. $\mathcal{T}(\Delta)$, is completely monotone. Second, the complex-valued function ψ is of the form (2.1) implying, for example, a certain asymptotic behavior of $\psi(u)$ as $u \to \infty$. Finally, we are not in regression setup and $\phi_Y(u|\Delta)$ is to be estimated by its empirical counterpart

$$\widehat{\phi}(u) = \frac{1}{n} \sum_{j=1}^{n} e^{\mathrm{i}u^\top (Y_{\Delta j} - Y_{\Delta(j-1)})}.$$

The contribution of this paper to the literature on composite function estimation is twofold. On the one hand, we introduce and study a new type of statistical problems which can be called estimation of a composite function under structural constraints. On the other hand, we propose new and constructive estimation approach which is rather general and can be used to solve other open statistical problems of this type. For example, one can directly adapt our method to the problem of semi-parametric inference in distributional Archimedian copula-based models [see, e.g., McNeil and Nešlehová (2009) for recent results], where one faces the problem of estimating a multidimensional distribution function of the form

$$F(x_1, \ldots, x_d) = G\big(f_1(x_1) + \cdots + f_d(x_d)\big), \qquad (x_1, \ldots, x_d) \in \mathbb{R}^d,$$

with a completely monotone function G and some functions $f_1, \ldots, f_d$. Further discussion on the problem of composite function estimation can be found in Remark 4.14.

2.3. *Specification analysis.* It is clear that without further restrictions on the class of time-changed Lévy processes our problem of estimating ν is not well defined, as even in the case of the perfectly known distribution of the process Y the parameters of the Lévy process L are generally not identifiable. Moreover, the corresponding statistical procedure will suffer from the "curse of dimensionality" as the dimension d increases. In order to avoid these undesirable features, we have to impose some additional restrictions on the structure of the time-changed process Y. In statistical literature, one can basically find two types of restricted composite models: additive models and single-index models. While the latter class of models is too restrictive in our situation, the former one naturally appears if one assumes the independence of the components of L_t. In this paper, we study a class of time-changed Lévy processes satisfying the following two assumptions:

(ALI) The Lévy process L_t has independent components such that at least two of them are nonzero, that is,

$$\phi_Y(u|t) = \mathcal{L}_t\big(-\psi_1(u_1) - \cdots - \psi_d(u_d)\big), \tag{2.4}$$

where $\psi_k, k = 1, \dots, d$, are the characteristic exponents of the components of L_t of the form

$$\begin{aligned}\psi_k(u) &= \mathrm{i}\mu_k u - \sigma_k^2 u^2/2 \\ &\quad + \int_{\mathbb{R}} \big(e^{\mathrm{i}ux} - 1 - \mathrm{i}ux \cdot \mathbf{1}_{\{|x|\le 1\}}\big)\nu_k(dx), \qquad k = 1, \dots, d,\end{aligned} \tag{2.5}$$

and

$$|\mu_l| + \sigma_l^2 + \int_{\mathbb{R}} x^2 \nu_l(dx) \neq 0 \tag{2.6}$$

for at least two different indexes l.

(ATI) The time change process $\mathcal{T}$ is independent of the Lévy process L and satisfies $\mathrm{E}[\mathcal{T}(t)] = t$.

Discussion. The advantage of the modeling framework (2.4) is twofold. On the one hand, models of this type are rather flexible: the distribution of Y_t for a fixed t is in general determined by $d + 1$ nonparametric components and $2 \times d$ parametric ones. On the other hand, these models remain parsimonious and, as we will see later, admit statistical inference not suffering from the "curse of dimensionality" as d becomes large. The latter feature of our model is in accordance with the well documented behavior of the additive models in regression setting and may become particularly important if one is going to use it, for instance, to model large portfolios of assets. The nondegeneracy assumption (2.6) basically excludes one-dimensional models and is not restrictive since it can be always checked prior to estimation by testing that

$$-\partial_{u_l u_l}\widehat{\phi}(u)|_{u=0} = \frac{1}{n}\sum_{j=1}^{n}\big(Y_{\Delta j,l} - Y_{\Delta(j-1),l}\big)^2 > 0$$

for at least two different indexes l. Let us make a few remarks on the one-dimensional case, where

$$\phi_Y(u|t) = \mathcal{L}_t(-\psi_1(u)), \qquad t \ge 0. \tag{2.7}$$

If $\mathcal{L}_\Delta$ is known, that is, the distribution of the r.v. $\mathcal{T}(\Delta)$ is known, we can consistently estimate the Lévy measure ν_1 by inverting $\mathcal{L}_\Delta$ (see Section 4.6 for more details). In the case when the function $\mathcal{L}_\Delta$ is unknown, one needs some additional assumptions (e.g., absolute continuity of the time change) to ensure identifiability.

Indeed, consider a class of the one-dimensional Lévy processes of the so-called compound exponential type with the characteristic exponent of the form

$$\psi(u) = \log\left[\frac{1}{1-\widetilde{\psi}(u)}\right],$$

where $\widetilde{\psi}(u)$ is the characteristic exponent of another one-dimensional Lévy process $\widetilde{L}_t$. It is well known [see, e.g., Section 3 in Chapter 4 of Steutel and van Harn (2004)] that $\exp(\psi(u))$ is the characteristic function of some infinitely divisible distribution if $\exp(\widetilde{\psi}(u))$ does. Introduce

$$\widetilde{\mathcal{L}}_\Delta(z) = \mathcal{L}_\Delta\big(\log(1+z)\big).$$

As can be easily seen, the function $\widetilde{\mathcal{L}}_\Delta$ is completely monotone with $\widetilde{\mathcal{L}}_\Delta(0) = 1$ and $\widetilde{\mathcal{L}}'_\Delta(0) = \mathcal{L}'_\Delta(0)$. Moreover, it is fulfilled $\widetilde{\mathcal{L}}_\Delta(-\widetilde{\psi}(u)) = \mathcal{L}_\Delta(-\psi(u))$ for all $u \in \mathbb{R}$. The existence of the time change (increasing) process $\mathcal{T}$ with a given marginal $\mathcal{T}(\Delta)$ can be derived from the general theory of stochastic partial ordering [see Kamae and Krengel (1978)]. The above construction indicates that the assumption $\mathrm{E}[\mathcal{T}(t)] = t$, $t \geq 0$, is not sufficient to ensure the identifiability in the case of one-dimensional time-changed Lévy models.

3. Estimation.

3.1. *Main ideas.* Assume that the Lévy measures of the component processes $L_t^1, \dots, L_t^d$ are absolutely continuous with integrable densities $\nu_1(x), \dots, \nu_d(x)$ that satisfy

$$\int_{\mathbb{R}} x^2 \nu_k(x)\, dx < \infty, \qquad k = 1, \dots, d.$$

Consider the functions

$$\bar{\nu}_k(x) = x^2 \nu_k(x), \qquad k = 1, \dots, d.$$

By differentiating ψ_k two times, we get

$$\psi_k''(u) = -\sigma_k^2 - \int_{\mathbb{R}} e^{\mathrm{i}ux} \bar{\nu}_k(x)\, dx.$$

For the sake of simplicity, in the sequel we will make the following assumption:

(ALS) The diffusion volatilities $\sigma_k, k = 1, \dots, d$, of the Lévy process L are supposed to be known.

A way how to extend our results to the case of the unknown (σ_k) is outlined in Section 4.6. Introduce the functions $\bar{\psi}_k(u) = \psi_k(u) + \sigma_k^2 u^2/2$ to get

$$\mathbf{F}[\bar{\nu}_k](u) = -\bar{\psi}_k''(u) = -\sigma_k^2 - \psi_k''(u), \tag{3.1}$$

where $\mathbf{F}[\bar{\nu}_k](u)$ stands for the Fourier transform of $\bar{\nu}_k$. Denote $Z = Y_\Delta$, $\phi_k(u) = \partial_{u_k}\phi_Z(u)$, $\phi_{kl}(u) = \partial_{u_k u_l}\phi_Z(u)$ and $\phi_{jkl}(u) = \partial_{u_j u_k u_l}\phi_Z(u)$ for $j, k, l \in \{1, \dots, d\}$ with

$$\phi_Z(u) = \mathrm{E}[\exp(\mathrm{i}u^\top Z)] = \mathcal{L}_\Delta\bigl(-\psi_1(u_1) - \cdots - \psi_d(u_d)\bigr). \tag{3.2}$$

Fix some $k \in \{1, \dots, d\}$ and for any real number u introduce a vector

$$u^{(k)} = (0, \dots, 0, u, 0, \dots, 0) \in \mathbb{R}^d$$

with u being placed at the kth coordinate of the vector $u^{(k)}$. Choose some $l \neq k$, such that the component L_t^l is not degenerated. Then we get from (3.2)

$$\frac{\phi_k(u^{(k)})}{\phi_l(u^{(k)})} = \frac{\psi_k'(u)}{\psi_l'(0)}, \tag{3.3}$$

if $\mu_l \neq 0$ and

$$\frac{\phi_k(u^{(k)})}{\phi_{ll}(u^{(k)})} = \frac{\psi_k'(u)}{\psi_l''(0)} \tag{3.4}$$

in the case $\mu_l = 0$. The identities $\phi_l(\mathbf{0}) = -\psi_l'(0)\mathcal{L}_\Delta'(0)$ and $\phi_{ll}(\mathbf{0}) = [\psi_l'(0)]^2 \times \mathcal{L}_\Delta''(0) - \psi_l''(0)\mathcal{L}_\Delta'(0)$ imply $\psi_l'(0) = -[\mathcal{L}_\Delta'(0)]^{-1}\phi_l(\mathbf{0}) = \Delta^{-1}\phi_l(\mathbf{0})$ and $\psi_l''(0) = -[\mathcal{L}_\Delta'(0)]^{-1}\phi_{ll}(\mathbf{0}) = \Delta^{-1}\phi_{ll}(\mathbf{0})$ if $\psi_l'(0) = 0$, since $\mathcal{L}_\Delta'(0) = -\mathrm{E}[\mathcal{T}(\Delta)] = -\Delta$. Combining this with (3.3) and (3.4), we derive

$$\psi_k''(u) = \Delta^{-1}\phi_l(\mathbf{0})\frac{\phi_{kk}(u^{(k)})\phi_l(u^{(k)}) - \phi_k(u^{(k)})\phi_{lk}(u^{(k)})}{\phi_l^2(u^{(k)})}, \qquad \mu_l \neq 0, \tag{3.5}$$

$$\psi_k''(u) = \Delta^{-1}\phi_{ll}(\mathbf{0})\frac{\phi_{kk}(u^{(k)})\phi_{ll}(u^{(k)}) - \phi_k(u^{(k)})\phi_{llk}(u^{(k)})}{\phi_{ll}^2(u^{(k)})}, \qquad \mu_l = 0. \tag{3.6}$$

Note that in the above derivations we have repeatedly used assumption (ATI), that turns out to be crucial for the identifiability. The basic idea of the algorithm, we shall develop in the Section 3.2, is to estimate $\bar{\nu}_k$ by an application of the regularized Fourier inversion formula to an estimate of $\bar{\psi}_k''(u)$. As indicated by formulas (3.5) and (3.6), one could, for example, estimate $\bar{\psi}_k''(u)$, if some estimates for the functions $\phi_k(u)$, $\phi_{lk}(u)$ and $\phi_{llk}(u)$ are available.

REMARK 3.1. One important issue we would like to comment on is the robustness of the characterizations (3.5) and (3.6) with respect to the independence assumption for the components of the Lévy process L_t. First, note that if the components are dependent, then the key identity (3.1) is not any longer valid for ψ_k'' defined as in (3.5) or (3.6). Let us determine how strong can it be violated. For concreteness, assume that $\mu_l > 0$ and that the dependence in the components of L_t

is due to a correlation between diffusion components. In particular, let $\Sigma(k,l) > 0$. Since in the general case

$$\partial_{u_k}\psi\big(u^{(k)}\big) = \partial_{u_l}\psi\big(u^{(k)}\big)\frac{\phi_k(u^{(k)})}{\phi_l(u^{(k)})}$$

and $\partial_{u_k u_k}\psi(u^{(k)}) = -\sigma_k^2 - \mathbf{F}[\bar{\nu}_k](u)$, we get

$$\mathbf{F}[\bar{\nu}_k](u) + \psi_k''(u) + \sigma_k^2 = \frac{\Sigma(k,l)}{2}\left[u\,\partial_{u_k}\left\{\frac{\phi_k(u^{(k)})}{\phi_l(u^{(k)})}\right\} + \frac{\phi_k(u^{(k)})}{\phi_l(u^{(k)})}\right].$$

Using the fact that both functions $u\partial_{u_k}\{\phi_k(u^{(k)})/\phi_l(u^{(k)})\}$ and $\phi_k(u^{(k)})/\phi_l(u^{(k)})$ are uniformly bounded for $u \in \mathbb{R}$, we get that the model "misspecification bias" is bounded by $C\Sigma(k,l)$ with some constant $C > 0$. Thus, the weaker is the dependence between components L^k and L^l, the smaller is the resulting "misspecification bias."

3.2. *Algorithm.* Set $Z_j = Y_{\Delta j} - Y_{\Delta(j-1)}$, $j = 1,\dots,n$, and denote by Z_j^k the kth coordinate of Z_j. Note that Z_j, $j = 1,\dots,n$, are identically distributed. The estimation procedure consists basically of three steps:

Step 1. First, we are interested in estimating partial derivatives of the function $\phi_Z(u)$ up to the third order. To this end, define

$$\widehat{\phi}_k(u) = \frac{1}{n}\sum_{j=1}^n Z_j^k \exp(\mathrm{i}u^\top Z_j), \tag{3.7}$$

$$\widehat{\phi}_{lk}(u) = \frac{1}{n}\sum_{j=1}^n Z_j^k Z_j^l \exp(\mathrm{i}u^\top Z_j), \tag{3.8}$$

$$\widehat{\phi}_{llk}(u) = \frac{1}{n}\sum_{j=1}^n Z_j^k Z_j^l Z_j^l \exp(\mathrm{i}u^\top Z_j). \tag{3.9}$$

Step 2. In a second step, we estimate the second derivative of the characteristic exponent $\psi_k(u)$. Set

$$\widehat{\psi}_{k,2}(u) = \Delta^{-1}\widehat{\phi}_l(\mathbf{0})\frac{\widehat{\phi}_{kk}(u^{(k)})\widehat{\phi}_l(u^{(k)}) - \widehat{\phi}_k(u^{(k)})\widehat{\phi}_{lk}(u^{(k)})}{[\widehat{\phi}_l(u^{(k)})]^2}, \tag{3.10}$$

if $|\widehat{\phi}_l(\mathbf{0})| > \kappa/\sqrt{n}$ and

$$\widehat{\psi}_{k,2}(u) = \Delta^{-1}\widehat{\phi}_{ll}(\mathbf{0})\frac{\widehat{\phi}_{kk}(u^{(k)})\widehat{\phi}_{ll}(u^{(k)}) - \widehat{\phi}_k(u^{(k)})\widehat{\phi}_{llk}(u^{(k)})}{[\widehat{\phi}_{ll}(u^{(k)})]^2} \tag{3.11}$$

otherwise, where κ is a positive number.

Step 3. Finally, we construct an estimate for $\bar{\nu}_k(x)$ by applying the Fourier inversion formula combined with a regularization to $\widehat{\psi}_{k,2}(u)$:

$$\widehat{\nu}_k(x) = -\frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux}[\widehat{\psi}_{k,2}(u) + \sigma_k^2]\mathcal{K}(uh_n)\,du, \tag{3.12}$$

where $\mathcal{K}(u)$ is a regularizing kernel supported on $[-1, 1]$ and h_n is a sequence of bandwidths which tends to 0 as $n \to \infty$. The choice of the sequence h_n will be discussed later on.

REMARK 3.2. The parameter κ determines the testing error for the hypothesis $H : \mu_l > 0$. Indeed, if $\mu_l = 0$, then $\phi_l(\mathbf{0}) = 0$ and by the central limit theorem

$$\begin{aligned}\mathbb{P}(|\widehat{\phi}_l(\mathbf{0})| > \kappa/\sqrt{n}) &\le \mathbb{P}(\sqrt{n}|\widehat{\phi}_l(\mathbf{0}) - \phi_l(\mathbf{0})| > \kappa)\\ &\to \mathbb{P}(|\xi| > \kappa/\sqrt{\mathrm{Var}[Z^l]}), \qquad n \to \infty,\end{aligned}$$

with $\xi \sim \mathcal{N}(0, 1)$.

4. Asymptotic analysis. In this section, we are going to study the asymptotic properties of the estimates $\widehat{\nu}_k(x)$, $k = 1, \dots, d$. In particular, we prove almost sure uniform as well as pointwise convergence rates for $\widehat{\nu}_k(x)$. Moreover, we will show the optimality of the above rates over suitable classes of time-changed Lévy models.

4.1. *Global vs. local smoothness of Lévy densities.* Let L_t be a one-dimensional Lévy process with a Lévy density ν. Denote $\bar{\nu}(x) = x^2\nu(x)$. For any two nonnegative numbers β and γ such that $\gamma \in [0, 2]$ consider two following classes of Lévy densities ν:

$$\mathfrak{S}_\beta = \left\{\nu : \int_{\mathbb{R}} (1 + |u|^\beta)\mathbf{F}[\bar{\nu}](u)\,du < \infty\right\} \tag{4.1}$$

and

$$\mathfrak{B}_\gamma = \left\{\nu : \int_{|y|>\epsilon} \nu(y)\,dy \asymp \frac{\Pi(\epsilon)}{\epsilon^\gamma}, \epsilon \to +0\right\}, \tag{4.2}$$

where Π is some positive function on $\mathbb{R}_+$ satisfying $0 < \Pi(+0) < \infty$. The parameter γ is usually called the Blumenthal–Geetor index of L_t. This index γ is related to the "degree of activity" of jumps of L_t. All Lévy measures put finite mass on the set $(-\infty, -\epsilon] \cup [\epsilon, \infty)$ for any arbitrary $\epsilon > 0$. If $\nu([-\epsilon, \epsilon]) < \infty$ the process has finite activity and $\gamma = 0$. If $\nu([-\epsilon, \epsilon]) = \infty$, that is, the process has infinite activity and in addition the Lévy measure $\nu((-\infty, -\epsilon] \cup [\epsilon, \infty))$ diverges near 0 at a rate $|\epsilon|^{-\gamma}$ for some $\gamma > 0$, then the Blumenthal–Geetor index of L_t is equal to γ. The higher γ gets, the more frequent the small jumps become.

Let us now investigate the connection between classes $\mathfrak{S}_\beta$ and $\mathfrak{B}_\gamma$. First, consider an example. Let L_t be a tempered stable Lévy process with a Lévy density

$$\nu(x) = \frac{2^\gamma \cdot \gamma}{\Gamma(1-\gamma)} x^{-(\gamma+1)} \exp\left(-\frac{x}{2}\right) \mathbf{1}_{(0,\infty)}(x), \qquad x > 0,$$

where $\gamma \in (0, 1)$. It is clear that $\nu \in \mathfrak{B}_\gamma$ but what is about $\mathfrak{S}_\beta$? Since

$$\bar{\nu}(x) = \frac{2^\gamma \cdot \gamma}{\Gamma(1-\gamma)} x^{1-\gamma} \exp\left(-\frac{x}{2}\right) \mathbf{1}_{(0,\infty)}(x),$$

we derive

$$\mathbf{F}[\bar{\nu}](u) = \int_0^\infty e^{\mathrm{i}ux} \bar{\nu}(x)\, dx \asymp 2^\gamma \gamma (1-\gamma) e^{\mathrm{i}\pi(1-\gamma/2)} u^{-2+\gamma}, \qquad u \to +\infty,$$

by the Erdélyi lemma [see Erdélyi (1956)]. Hence, ν cannot belong to $\mathfrak{S}_\beta$ as long as $\beta > 1-\gamma$. The message of this example is that given the activity index γ, the parameter β determining the smoothness of $\bar{\nu}$, cannot be taken arbitrary large. The above example can be straightforwardly generalized to a class of Lévy densities supported on $\mathbb{R}_+$. It turns out that if the Lévy density ν is supported on $[0, \infty)$, is infinitely smooth in $(0, \infty)$ and $\nu \in \mathfrak{B}_\gamma$ for some $\gamma \in (0, 1)$, then $\nu \in \mathfrak{S}_\beta$ for all β satisfying $0 \le \beta < 1-\gamma$ and $\nu \notin \mathfrak{S}_\beta$ for $\beta > 1-\gamma$. As a matter of fact, in the case $\gamma = 0$ (finite activity case) the situation is different and β can be arbitrary large.

The above discussion indicates that in the case $\nu \in \mathfrak{B}_\gamma$ with some $\gamma > 0$ it is reasonable to look at the local smoothness of $\bar{\nu}_k$ instead of the global one. To this end, fix a point $x_0 \in \mathbb{R}$ and a positive integer number $s \ge 1$. For any $\delta > 0$ and $D > 0$ introduce a class $\mathfrak{H}_s(x_0, \delta, D)$ of Lévy densities ν defined as

$$\begin{aligned}\mathfrak{H}_s(x_0, \delta, D) = \Big\{ &\nu : \bar{\nu}(x) \in C^s(]x_0-\delta, x_0+\delta[), \\ &\sup_{x \in]x_0-\delta, x_0+\delta[} \left|\bar{\nu}^{(l)}(x)\right| \le D \text{ for } l = 1, \dots, s \Big\}.\end{aligned} \tag{4.3}$$

4.2. *Assumptions*. In order to prove the convergence of $\widehat{\nu}_k(x)$, we need the assumptions listed below:

(AL1) The Lévy densities $\nu_1, \dots, \nu_d$ are in the class $\mathfrak{B}_\gamma$ for some $\gamma > 0$.

(AL2) For some $p > 2$, the Lévy densities ν_k, $k = 1, \dots, d$, have finite absolute moments of the order p:

$$\int_{\mathbb{R}} |x|^p \nu_k(x)\, dx < \infty, \qquad k = 1, \dots, d.$$

(AT1) The time change $\mathcal{T}$ is independent of the Lévy process L and the sequence $T_k = \mathcal{T}(\Delta k) - \mathcal{T}(\Delta(k-1))$, $k \in \mathbb{N}$, is strictly stationary, α-mixing with the mixing coefficients $(\alpha_T(j))_{j \in \mathbb{N}}$ satisfying

$$\alpha_T(j) \le \bar{\alpha}_0 \exp(-\bar{\alpha}_1 j), \qquad j \in \mathbb{N},$$

for some positive constants $\bar{\alpha}_0$ and $\bar{\alpha}_1$. Moreover, assume that

$$\mathrm{E}[\mathcal{T}^{-2/\gamma}(\Delta)] < \infty, \qquad \mathrm{E}[\mathcal{T}^{2p}(\Delta)] < \infty$$

with γ and p being from assumptions (AL1) and (AL2), respectively.

(AT2) The Laplace transform $\mathcal{L}_t(z)$ of $\mathcal{T}(t)$ fulfills

$$\mathcal{L}_t'(z) = o(1), \qquad \mathcal{L}_t''(z)/\mathcal{L}_t'(z) = O(1), \qquad |z| \to \infty, \qquad \operatorname{Re} z > 0.$$

(AK) The regularizing kernel $\mathcal{K}$ is uniformly bounded, is supported on $[-1, 1]$ and satisfies

$$\mathcal{K}(u) = 1, \qquad u \in [-a_K, a_K],$$

with some $0 < a_K < 1$.

(AH) The sequence of bandwidths h_n is assumed to satisfy

$$h_n^{-1} = O(n^{1-\delta}), \qquad M_n \sqrt{\frac{\log n}{n}} \sqrt{\frac{1}{h_n} \log \frac{1}{h_n}} = o(1), \qquad n \to \infty,$$

for some positive number δ fulfilling $2/p < \delta \leq 1$, where

$$M_n = \max_{l \neq k} \sup_{\{|u| \leq 1/h_n\}} |\phi_l^{-1}(u^{(k)})|.$$

REMARK 4.1. By requiring $\nu_k \in \mathfrak{B}_\gamma$, $k = 1, \ldots, d$, with some $\gamma > 0$, we exclude from our analysis pure compound Poisson processes and some infinite activity Lévy processes with $\gamma = 0$. This is mainly done for the sake of brevity: we would like to avoid additional technical calculations related to the fact that the distribution of Y_t is not in general absolutely continuous in this case.

REMARK 4.2. Assumption (AT1) is satisfied if, for example, the process $\mathcal{T}(t)$ is of the form (1.1), where the rate process $\rho(u)$ is strictly stationary, geometrically α-mixing and fulfills

$$\mathrm{E}[\rho^{2p}(u)] < \infty, \qquad u \in [0, \Delta], \qquad \mathrm{E}\left(\int_0^\Delta \rho(u)\,du\right)^{-2/\gamma} < \infty. \tag{4.4}$$

In the case of the Cox–Ingersoll–Ross process ρ (see Section 5.2), assumptions (4.4) are satisfied for any $p > 0$ and any $\gamma > 0$.

REMARK 4.3. Let us comment on assumption (AH). Note that in order to determine M_n, we do not need the characteristic function $\phi(u)$ itself, but only a low bound for its tails. Such low bound can be constructed if, for example, a low bound for the tail of $\mathcal{L}_t'(z)$ and an upper bound for the Blumenthal–Geetor index γ are available [see Belomestny (2010b) for further discussion]. In practice, of course, one should prefer adaptive methods for choosing h_n. One such method, based on the so called "quasi-optimality" approach, is proposed and used in Section 5.1. The theoretical analysis of this method is left for future research.

4.3. *Uniform rates of convergence.* Fix some k from the set $\{1, 2, \ldots, d\}$. Define a weighting function $w(x) = \log^{-1/2}(e + |x|)$ and denote

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L_\infty(\mathbb{R},w)} = \sup_{x\in\mathbb{R}}[w(|x|)|\bar{\nu}_k(x) - \widehat{\nu}_k(x)|].$$

Let ξ_n be a sequence of positive r.v. and q_n be a sequence of positive real numbers. We shall write $\xi_n = O_{\text{a.s.}}(q_n)$ if there is a constant $D > 0$ such that $\mathbb{P}(\limsup_{n\to\infty} q_n^{-1}\xi_n \le D) = 1$. In the case $\mathbb{P}(\limsup_{n\to\infty} q_n^{-1}\xi_n = 0) = 1$, we shall write $\xi_n = o_{\text{a.s.}}(q_n)$.

THEOREM 4.4. *Suppose that assumptions* (AL1), (AL2), (AT1), (AT2), (AK) *and* (AH) *are fulfilled. Let $\widehat{\nu}_k(x)$ be the estimate for $\bar{\nu}_k(x)$ defined in Section* 3.2. *If $\nu_k \in \mathfrak{S}_\beta$ for some $\beta > 0$, then*

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L_\infty(\mathbb{R},w)} = O_{\text{a.s.}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}\int_{-1/h_n}^{1/h_n}\mathfrak{R}_k^2(u)\,du} + h_n^\beta\right)$$

for arbitrary small $\varepsilon > 0$, where

$$\mathfrak{R}_k(u) = \frac{(1 + |\psi_k'(u)|)^2}{|\mathcal{L}_\Delta'(-\psi_k(u))|}.$$

COROLLARY 4.5. *Suppose that $\sigma_k = 0$, $\gamma \in (0, 1]$ in assumption* (AL1) *and*

$$|\mathcal{L}_\Delta'(z)| \gtrsim \exp(-a|z|^\eta), \qquad |z| \to \infty, \qquad \operatorname{Re} z \ge 0,$$

for some $a > 0$ and $\eta > 0$. If $\mu_k > 0$, then

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L_\infty(\mathbb{R},w)} = O_{\text{a.s.}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}}\exp(ac \cdot h_n^{-\eta}) + h_n^\beta\right) \tag{4.5}$$

with some constant $c > 0$. In the case $\mu_k = 0$ we have

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L_\infty(\mathbb{R},w)} = O_{\text{a.s.}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}}\exp(ac \cdot h_n^{-\gamma\eta}) + h_n^\beta\right). \tag{4.6}$$

Choosing h_n in such a way that the r.h.s. of (4.5) *and* (4.6) *are minimized, we obtain the rates shown in the Table* 1. *If $\gamma \in (0, 1]$ in assumption* (AL1) *and*

$$|\mathcal{L}_\Delta'(z)| \gtrsim |z|^{-\alpha}, \qquad |z| \to \infty, \qquad \operatorname{Re} z \ge 0,$$

for some $\alpha > 0$, then

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L_\infty(\mathbb{R},w)} = O_{\text{a.s.}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}} h_n^{-1/2-\alpha} + h_n^\beta\right)$$

TABLE 1
Uniform convergence rates for $\widehat{\nu}_k$ in the case $\sigma_k = 0$

$\lvert\mathcal{L}'_\Delta(z)\rvert \gtrsim \lvert z\rvert^{-\alpha}$		$\lvert\mathcal{L}'_\Delta(z)\rvert \gtrsim \exp(-a\lvert z\rvert^{\eta})$	
$\mu_k > 0$	$\mu_k = 0$	$\mu_k > 0$	$\mu_k = 0$
$n^{-\beta/(2\alpha+2\beta+1)} \times \log^{(3+\varepsilon)\beta/(2\alpha+2\beta+1)}(n)$	$n^{-\beta/(2\alpha\gamma+2\beta+1)} \times \log^{(3+\varepsilon)\beta/(2\alpha\gamma+2\beta+1)}(n)$	$\log^{-\beta/\eta} n$	$\log^{-\beta/\gamma\eta} n$

provided $\mu_k > 0$. *In the case* $\mu_k = 0$, *one has*

$$\|\bar{\nu}_k - \widehat{\nu}_k\|_{L_\infty(\mathbb{R},w)} = O_{\text{a.s.}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}}h_n^{-1/2-\alpha\gamma} + h_n^{\beta}\right).$$

The choices $h_n = n^{-1/(2(\alpha+\beta)+1)} \log^{(3+\varepsilon)/(2(\alpha+\beta)+1)}(n)$ *and*

$$h_n = n^{-1/(2(\alpha\gamma+\beta)+1)} \log^{(3+\varepsilon)/(2(\alpha\gamma+\beta)+1)}(n)$$

for the cases $\mu_k > 0$ *and* $\mu_k = 0$, *respectively, lead to the bounds shown in Table* 1. *In the case* $\sigma_k > 0$, *the rates of convergence are given in Table* 2.

REMARK 4.6. As one can see, assumption (AH) is always fulfilled for the optimal choices of h_n given in Corollary 4.5, provided $\alpha\gamma + \beta > 0$ and $p > 2 + 1/(\alpha\gamma + \beta)$.

4.4. *Pointwise rates of convergence*. Since the transformed Lévy density $\bar{\nu}_k$ is usually not smooth at 0 (see Section 4.1), pointwise rates of convergence might be more informative than the uniform ones if $\nu_k \in \mathfrak{B}_\gamma$ for some $\gamma > 0$. It is remarkable that the same estimate $\widehat{\nu}_k$ as before will achieve the optimal pointwise convergence rates in the class $\mathfrak{H}_s(x_0, \delta, D)$, provided the kernel $\mathcal{K}$ satisfies (AK) and is sufficiently smooth.

THEOREM 4.7. *Suppose that assumptions* (AL1), (AL2), (AT1), (AT2), (AK) *and* (AH) *are fulfilled. If* $\nu_k \in \mathfrak{H}_s(x_0, \delta, D)$ *with* $\mathfrak{H}_s(x_0, \delta, D)$ *being defined in*

TABLE 2
Uniform convergence rates for $\widehat{\nu}_k$ in the case $\sigma_k > 0$

$\lvert\mathcal{L}'_\Delta(z)\rvert \gtrsim \lvert z\rvert^{-\alpha}$	$\lvert\mathcal{L}'_\Delta(z)\rvert \gtrsim \exp(-a\lvert z\rvert^{\eta})$
$n^{-\beta/(4\alpha+2\beta+1)} \log^{(3+\varepsilon)\beta/(4\alpha+2\beta+1)}(n)$	$\log^{-\beta/2\eta} n$

(4.3), *for some* $s \geq 1$, $\delta > 0$, $D > 0$, *and* $\mathcal{K} \in C^m(\mathbb{R})$ *for some* $m \geq s$, *then*

$$(4.7)\qquad |\widehat{\nu}_k(x_0) - \bar{\nu}_k(x_0)| = O_{\text{a.s.}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n} \int_{-1/h_n}^{1/h_n} \mathfrak{R}_k^2(u)\, du} + h_n^s\right)$$

with $\mathfrak{R}_k(u)$ *as in Theorem* 4.4. *As a result, the pointwise rates of convergence for different asymptotic behaviors of the Laplace transform* $\mathcal{L}_t$ *coincide with ones given in Tables* 1 *and* 2, *if we replace* β *with* s.

REMARK 4.8. If the kernel $\mathcal{K}$ is infinitely smooth, then it will automatically "adapt" to the pointwise smoothness of $\bar{\nu}_k$, that is, (4.7) will hold for arbitrary large $s \geq 1$, provided $\nu_k \in \mathfrak{H}_s(x_0, \delta, D)$ with some $\delta > 0$ and $D > 0$. An example of infinitely smooth kernels satisfying (AK) is given by the so called flat-top kernels (see Section 5.1 for the definition).

4.5. *Lower bounds*. In this section, we derive a lower bound on the minimax risk of an estimate $\widehat{\nu}(x)$ over a class of one-dimensional time-changed Lévy processes $Y_t = L_{\mathcal{T}(t)}$ with the known distribution of $\mathcal{T}$, such that the Lévy measure ν of the Lévy process L_t belongs to the class $\mathfrak{S}_\beta \cap \mathfrak{B}_\gamma$ with some $\beta > 0$ and $\gamma \in (0, 1]$. The following theorem holds.

THEOREM 4.9. *Let* L_t *be a Lévy process with zero diffusion part, a drift* μ *and a Lévy density* ν. *Consider a time-changed Lévy process* $Y_t = L_{\mathcal{T}(t)}$, *where the Laplace transform of the time change* $\mathcal{T}(t)$ *fulfills*

$$(4.8)\qquad \mathcal{L}_\Delta^{(k+1)}(z)/\mathcal{L}_\Delta^{(k)}(z) = O(1), \qquad |z| \to \infty, \qquad \operatorname{Re} z \geq 0,$$

for $k = 0, 1, 2$, *and uniformly in* $\Delta \in [0, 1]$. *Then*

$$(4.9)\qquad \liminf_{n\to\infty} \inf_{\widehat{\nu}} \sup_{\nu \in \mathfrak{S}_\beta \cap \mathfrak{B}_\gamma} \mathbb{P}_{(\nu,\mathcal{T})}\big(\|\bar{\nu} - \widehat{\nu}\|_{L_\infty(\mathbb{R},w)} > \varepsilon h_n^\beta \log^{-1}(1/h_n)\big) > 0$$

for any $\varepsilon > 0$ *and any sequence* h_n *satisfying*

$$n\Delta^{-1}[\mathcal{L}'_\Delta(c \cdot h_n^{-\gamma})]^2 h_n^{2\beta+1} = O(1), \qquad n \to \infty,$$

in the case $\mu = 0$ *and*

$$n\Delta^{-1}[\mathcal{L}'_\Delta(c \cdot h_n^{-1})]^2 h_n^{2\beta+1} = O(1), \qquad n \to \infty,$$

in the case $\mu > 0$, *with some positive constant* $c > 0$. *Note that the infimum in* (4.9) *is taken over all estimators of* ν *based on* n *observations of the r.v.* Y_Δ *and* $\mathbb{P}_{(\nu,\mathcal{T})}$ *stands for the distribution of* n *copies of* Y_Δ.

COROLLARY 4.10. *Suppose that the underlying Lévy process is driftless, that is, $\mu = 0$ and $\mathcal{L}_t(z) = \exp(-azt)$ for some $a > 0$, corresponding to a deterministic time change process $\mathcal{T}(t) = at$. Then by taking*

$$h_n = \left(\frac{\log n - ((2\beta + 1)/\gamma)\log\log n}{2ac\Delta}\right)^{-1/\gamma},$$

we arrive at

$$\liminf_{n\to\infty} \inf_{\widehat{\nu}} \sup_{\nu \in \mathfrak{S}_\beta \cap \mathfrak{B}_\gamma} \mathbb{P}_{(\nu,\mathcal{T})}\left(\|\bar{\nu} - \widehat{\nu}\|_{L_\infty(\mathbb{R},w)} > \varepsilon \cdot \Delta^{\beta/\gamma} \log^{-\beta/\gamma} n\right) > 0.$$

COROLLARY 4.11. *Again let $\mu = 0$. Take $\mathcal{L}_t(z) = 1/(1+z)^{\alpha_0 t}$, $\operatorname{Re} z > 0$ for some $\alpha_0 > 0$, resulting in a Gamma process $\mathcal{T}(t)$ (see Section 5.1 for the definition). Under the choice*

$$h_n = (n\Delta)^{-1/(2\alpha\gamma + 2\beta + 1)}$$

we get

$$\liminf_{n\to\infty} \inf_{\widehat{\nu}} \sup_{\nu \in \mathfrak{S}_\beta \cap \mathfrak{B}_\gamma} \mathbb{P}_{(\nu,\mathcal{T})}\left(\|\bar{\nu} - \widehat{\nu}\|_{L_{\infty,w}(\mathbb{R})} > \varepsilon \cdot (n\Delta)^{-\beta/(2\alpha\gamma+2\beta+1)} \log^{-1} n\right) > 0,$$

where $\alpha = \alpha_0\Delta + 1$.

REMARK 4.12. Theorem 4.9 continues to hold for $\Delta \to 0$ and therefore can be used to derive minimax lower bounds for the risk of $\widehat{\nu}$ in high-frequency setup. As can be seen from Corollaries 4.10 and 4.11, the rates will strongly depend on the specification of the time change process $\mathcal{T}$.

The pointwise rates of convergence obtained in (4.7) turn out to be optimal over the class $\mathfrak{H}_s(x_0, \delta, D) \cap \mathfrak{B}_\gamma$ with $s \geq 1$, $\delta > 0$, $x_0 \in \mathbb{R}$, $D > 0$ and $\gamma \in (0, 1]$ as the next theorem shows.

THEOREM 4.13. *Let L_t be a Lévy process with zero diffusion part, a drift μ and a Lévy density ν. Consider a time-changed Lévy process $Y_t = L_{\mathcal{T}(t)}$, where the Laplace transform of the time change $\mathcal{T}(t)$ fulfills (4.8). Then*

$$\liminf_{n\to\infty} \inf_{\widehat{\nu}} \sup_{\nu \in \mathfrak{H}_s(x_0,\delta,D) \cap \mathfrak{B}_\gamma} \mathbb{P}_{(\nu,\mathcal{T})}\left(|\bar{\nu}(x_0) - \widehat{\nu}(x_0)| > \varepsilon h_n^s \log^{-1}(1/h_n)\right) > 0 \tag{4.10}$$

for $s \geq 1$, $\delta > 0$, $D > 0$, any $\varepsilon > 0$ and any sequence h_n satisfying

$$n\Delta^{-1}[\mathcal{L}'_\Delta(c \cdot h_n^{-\gamma})]^2 h_n^{2s+1} = O(1), \qquad n \to \infty,$$

in the case $\mu = 0$ and

$$n\Delta^{-1}[\mathcal{L}'_\Delta(c \cdot h_n^{-1})]^2 h_n^{2s+1} = O(1), \qquad n \to \infty,$$

in the case $\mu > 0$, with some positive constant $c > 0$.

4.6. *Extensions*.

One-dimensional time-changed Lévy models. Let us consider a class of one-dimensional time-changed Lévy models (2.7) with the known time change process, that is, the known function $\mathcal{L}_t$ for all $t > 0$. This class of models trivially includes Lévy processes without time change [by setting $\mathcal{L}_t(z) = \exp(-tz)$] studied in Neumann and Reiß (2009) and Comte and Genon-Catalot (2010). We have in this case

$$\psi_1''(u) = -\frac{\phi''(u)\mathcal{L}_\Delta'(-\psi_1(u)) - \phi'(u)\mathcal{L}_\Delta''(-\psi_1(u))/\mathcal{L}_\Delta'(-\psi_1(u))}{[\mathcal{L}_\Delta'(-\psi_1(u))]^2} \tag{4.11}$$

with

$$\psi_1(u) = -\mathcal{L}_\Delta^-(\phi(u)),$$

where $\mathcal{L}_\Delta^-$ is an inverse function for $\mathcal{L}_\Delta$. Thus, $\psi_1''(u)$ is again a ratio-type estimate involving the derivatives of the c.f. ϕ up to second order, that agrees with the one proposed in Comte and Genon-Catalot (2010) for the case of pure Lévy processes. Although we do not study the case of one-dimensional models in this work, our analysis can be easily adapted to this situation as well. In particular, the derivation of the pointwise convergence rates can be directly carried over to this situation.

The case of the unknown (σ_k). One way to proceed in the case of the unknown (σ_k) and $\nu_k \in \mathfrak{B}_\gamma$ with $\gamma < 2$ is to define $\widetilde{\nu}_k(x) = x^4\nu_k(x)$. Assuming $\int \widetilde{\nu}_k(x)\,dx < \infty$, we get

$$\psi_k^{(4)}(u) = \int_{\mathbb{R}} e^{\mathrm{i}ux}\widetilde{\nu}_k(x)\,dx.$$

Hence, in the above situation one can apply the regularized Fourier inversion formula to an estimate of $\psi_k^{(4)}(u)$ instead of $\psi_k''(u)$.

Estimation of $\mathcal{L}_\Delta$. Let us first estimate ψ_k. Set

$$\widehat{\psi}_k(u) = \Delta^{-1}\widehat{\phi}_l(\mathbf{0})\int_0^u \frac{\widehat{\phi}_k(v^{(k)})}{\widehat{\phi}_l(v^{(k)})}\,dv.$$

Under Assumptions (AL2), (AT1), (AT2), (AK) and (AH) we derive

$$\|\psi_k - \widehat{\psi}_k\|_{L_\infty(\mathbb{R},w)} = O_{\text{a.s.}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}}\right) \tag{4.12}$$

with a weighting function

$$w(u) = \left[\int_0^u \frac{1 + |\psi_k'(v)|}{|\mathcal{L}_\Delta'(-\psi_k(v))|}\,dv\right]^{-1}.$$

Now let us define an estimate for $\mathcal{L}_\Delta$ as a solution of the following optimization problem

$$\widehat{\mathcal{L}}_\Delta = \arg \inf_{\mathcal{L}\in\mathfrak{M}_\Delta} \sup_{u\in\mathbb{R}} \{w(u)|\mathcal{L}(-\widehat{\psi}_k(u)) - \widehat{\phi}(u^{(k)})|\}, \tag{4.13}$$

where $\mathfrak{M}_\Delta$ is the set of completely monotone functions $\mathcal{L}$ satisfying $\mathcal{L}(0) = 1$ and $\mathcal{L}'(0) = -\Delta$. Simple calculations and the bound (4.12) yield

$$\sup_{u\in\mathbb{R}} \{w(u)|\widehat{\mathcal{L}}_\Delta(-\psi_k(u)) - \mathcal{L}_\Delta(-\psi_k(u))|\} = O_{\text{a.s.}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}}\right). \tag{4.14}$$

Since any function $\mathcal{L}$ from $\mathfrak{M}_\Delta$ has a representation

$$\mathcal{L}(u) = \int_0^\infty e^{-ux}\, dF(x)$$

with some distribution function F satisfying $\int x\, dF(x) = \Delta$, we can replace the optimization over $\mathfrak{M}$ in (4.13) by the optimization over the corresponding set of distribution functions. The advantage of the latter approach is that herewith we can directly get an estimate for the distribution function of the r.v. $\mathcal{T}(\Delta)$. A practical implementation of the estimate (4.13) is still to be worked out, as the optimization over the set $\mathfrak{M}_\Delta$ is not feasible and should be replaced by the optimization over suitable approximation classes (sieves). Moreover, the "optimal" weights in (4.13) depend on the unknown $\mathcal{L}$. However, it turns out that it is possible to use any weighting function which is dominated by $w(u)$, that is, one needs only some lower bounds for $\mathcal{L}'_\Delta$.

REMARK 4.14. It is interesting to compare (4.12) and (4.14) with Theorem 3.2 in Horowitz and Mammen (2007). At first sight it may seem strange that, while the rates of convergence for our "link" function $\mathcal{L}_\Delta$ and the "components" ψ_k depend on the tail behavior of $\mathcal{L}'_\Delta$, the rates in Horowitz and Mammen (2007) rely only on the smoothness of the link function and the components. The main reason for this is that the derivative of the link function in the above paper is assumed to be uniformly bounded from below [assumption (A8)], a restriction that can be hardly justifiable in our setting. The convergence analysis in the unbounded case is, in our opinion, an important contribution of this paper to the problem of estimating composite functions that can be carried over to other setups and settings.

4.7. *Discussion.* As can be seen, the estimate $\widehat{\nu}_k$ can exhibit various asymptotic behavior depending on the underlying Lévy process L_t and the time-change $\mathcal{T}(t)$. In particular, if the Laplace transform $\mathcal{L}_t(z)$ of $\mathcal{T}$ dies off at exponential rate as $\operatorname{Re} z \to +\infty$ and $\mu_k = 0$, then the rates of convergence of $\widehat{\nu}_k$ are logarithmic and depend on the Blumenthal–Geetor index of the Lévy process L_t. The larger is the Blumenthal–Geetor index, the slower are the rates and the more difficult the estimation problem becomes. For the polynomially decaying $\mathcal{L}_t(z)$ one gets polynomial convergence rates that also depend on the Blumenthal–Geetor index of L_t. Let us also note that the uniform rates of convergence are usually rather slow, since $\beta < 1 - \gamma$ in most situations. The pointwise convergence rates for points $x_0 \neq 0$ can, on the contrary, be very fast. The rates obtained turn out to be optimal up to a logarithmic factor in the minimax sense over the classes $\mathfrak{S}_\beta \cap \mathfrak{B}_\gamma$ and $\mathfrak{H}_s(x_0, \delta, D) \cap \mathfrak{B}_\gamma$.

5. Simulation study. In our simulation study, we consider two models based on time-changed normal inverse Gaussian (NIG) Lévy processes. The NIG Lévy processes is a relatively new class of processes introduced in Barndorff-Nielsen (1998) as a model for log returns of stock prices. The processes of this type are characterized by the property that their increments have NIG distribution. Barndorff-Nielsen (1998) considered classes of normal variance–mean mixtures and defined the NIG distribution as the case when the mixing distribution is inverse Gaussian. Shortly after its introduction, it was shown that the NIG distribution fits very well the log returns on German stock market data, making the NIG Lévy processes of great interest for practioneers. A NIG distribution has in general four parameters: $\alpha \in \mathbb{R}_+$, $\varkappa \in \mathbb{R}$, $\delta \in \mathbb{R}_+$ and $\mu \in \mathbb{R}$ with $|\varkappa| < \alpha$. Each parameter in NIG$(\alpha, \varkappa, \delta, \mu)$ distribution can be interpreted as having a different effect on the shape of the distribution: α is responsible for the tail heaviness of steepness, $\varkappa$ has to do with symmetry, δ scales the distribution and μ determines its mean value. The NIG distribution is infinitely divisible with c.f.

$$\phi(u) = \exp\{\delta(\sqrt{\alpha^2 - \varkappa^2} - \sqrt{\alpha^2 - (\varkappa + \mathrm{i}u)^2} + \mathrm{i}\mu u)\}.$$

Therefore, one can define the NIG Lévy process $(L_t)_{t \geq 0}$ which starts at zero and has independent and stationary increments such that each increment $L_{t+\Delta} - L_t$ has NIG$(\alpha, \varkappa, \Delta\delta, \Delta\mu)$ distribution. The NIG process has no diffusion component making it a pure jump process with the Lévy density

$$\nu(x) = \frac{2\alpha\delta}{\pi} \frac{\exp(\varkappa x) K_1(\alpha|x|)}{|x|}, \tag{5.1}$$

where $K_\lambda(z)$ is the modified Bessel function of the third kind. Taking into account the asymptotic relations

$$K_1(z) \asymp 2/z, \qquad z \to +0, \quad \text{and} \quad K_1(z) \asymp \sqrt{\frac{\pi}{2z}} e^{-z}, \qquad z \to +\infty,$$

we conclude that $\nu \in \mathfrak{B}_1$ and $\nu \in \mathfrak{H}_s(x_0, \delta, D)$ for arbitrary large $s > 0$ and some $\delta > 0$, $D > 0$, if $x_0 \neq 0$. Moreover, assumption (AL2) is fulfilled for any $p > 0$. Furthermore, the identity

$$\frac{d^2}{du^2} \log \phi(u) = -\alpha^2 / (\alpha^2 - (\varkappa + \mathrm{i}u)^2)^{3/2}$$

implies $\nu \in \mathfrak{S}_{2-\delta}$ for arbitrary small $\delta > 0$. In the next sections are going to study two time-changed NIG processes: one uses the Gamma process as a time change and another employs the integrated CIR processes to model $\mathcal{T}$.

5.1. *Time change via a Gamma process.* Gamma process is a Lévy process such that its increments have Gamma distribution, so that $\mathcal{T}$ is a pure-jump increasing Lévy process with the Lévy density

$$\nu_{\mathcal{T}}(x) = \theta x^{-1} \exp(-\lambda x), \qquad x \ge 0,$$

where the parameter θ controls the rate of jump arrivals and the scaling parameter λ inversely controls the jump size. The Laplace transform of $\mathcal{T}$ is of the form

$$\mathcal{L}_t(z) = (1 + z/\lambda)^{-\theta t}, \qquad \operatorname{Re} z \ge 0.$$

It follows from the properties of the Gamma and the corresponding inverse Gamma distributions that assumptions (AT1) and (AT2) are fulfilled for the Gamma process $\mathcal{T}$, provided $\theta\Delta > 2/\gamma$. Consider now the time-changed Lévy process $Y_t = L_{\mathcal{T}(t)}$ where $L_t = (L_t^1, L_t^2, L_t^3)$ is a three-dimensional Lévy process with independent NIG components and $\mathcal{T}$ is a Gamma process. Note that the process Y_t is a multidimensional Lévy process since $\mathcal{T}$ was itself the Lévy process. Let us be more specific and take the Δ-increments of the Lévy processes L_t^1, L_t^2 and L_t^3 to have NIG$(1, -0.05, 1, -0.5)$, NIG$(3, -0.05, 1, -1)$ and NIG$(1, -0.03, 1, 2)$ distributions, respectively. Take also $\theta = 1$ and $\lambda = 1$ for the parameters of the Gamma process $\mathcal{T}$. Next, fix an equidistant grid on $[0, 10]$ of the length $n = 1{,}000$ and simulate a discretized trajectory of the process Y_t. Let us stress that the dependence structure between the components of Y_t is rather flexible (although they are uncorrelated) and can be efficiently controlled by the parameters of the corresponding Gamma process $\mathcal{T}$. Next, we construct an estimate $\widehat{\nu}_1$ as described in Section 3.2. We first estimate the derivatives ϕ_1, ϕ_2, ϕ_{11} and ϕ_{12} by means of (3.7) and (3.8). Then we estimate $\psi_1''(u)$ using the formula (3.10) with $k = 1$ and $l = 2$. Finally, we get $\widehat{\nu}_1$ from (3.12) where the kernel $\mathcal{K}$ is chosen to be the so-called flat-top kernel of the form

$$\mathcal{K}(x) = \begin{cases} 1, & |x| \le 0.05, \\ \exp\left(-\dfrac{e^{-1/(|x|-0.05)}}{1-|x|}\right), & 0.05 < |x| < 1, \\ 0, & |x| \ge 1. \end{cases}$$

The flat-top kernels obviously satisfy assumption (AK). Thus, all assumptions of Theorem 4.4 are fulfilled and Corollary 4.5 leads to the following convergence rates for the estimate $\widehat{\nu}_1$ of the function $\bar{\nu}_1(x) = x^2 \nu(x)$:

$$\|\bar{\nu}_1 - \widehat{\nu}_1\|_{L_\infty(\mathbb{R}, w)} = O_{\text{a.s.}}\bigl(n^{-(1-\delta')/(\theta\Delta+5/2)} \log^{(3+\epsilon')/(\theta\Delta+5/2)}(n)\bigr), \qquad n \to \infty,$$

with arbitrary small positive numbers δ' and ϵ', provided the sequence h_n is chosen as in Corollary 4.5. Let us turn to the finite sample performance of the estimate $\widehat{\nu}_1$.

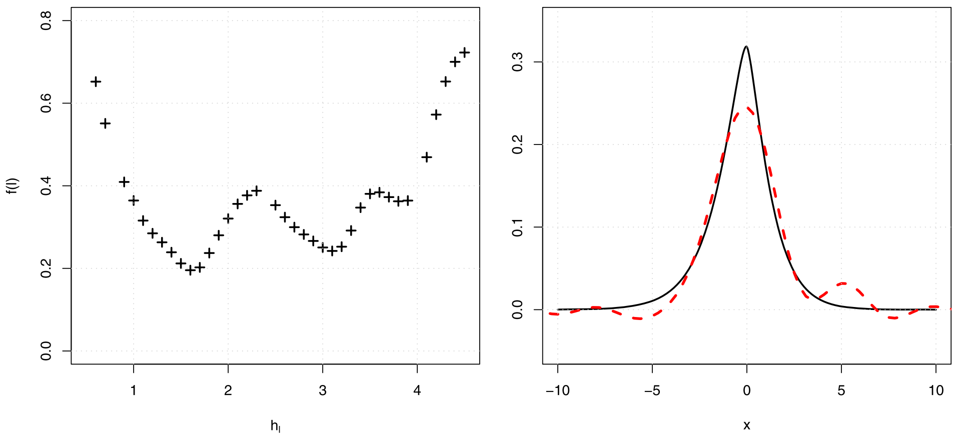

FIG. 1. *Left-hand side*: *objective function* $f(l)$ *for "quasi-optimality" approach versus the corresponding bandwidths* h_l, $l = 1, \dots, 40$. *Right-hand side*: *adaptive estimate* $\widetilde{\nu}_1$ (*dashed line*) *together with the true function* $\bar{\nu}_1$ (*solid line*).

It turns out that the choice of the sequence h_n is crucial for a good performance of ν_1. For this choice, we adopt the so called "quasi-optimality" approach proposed in Bauer and Reiß (2008). This approach is aimed to perform a model selection in inverse problems without taking into account the noise level. Although one can prove the optimality of this criterion on average only, it leads in many situations to quite reasonable results. In order to implement the "quasi-optimality" algorithm in our situation, we first fix a sequence of bandwidths $h_1, \dots, h_L$ and construct the estimates $\nu_1^{(1)}, \dots, \nu_1^{(L)}$ using the formula (3.12) with bandwidths $h_1, \dots, h_L$, respectively. Then one finds $l^\star = \arg\min_l f(l)$ with

$$f(l) = \|\widehat{\nu}_1^{(l+1)} - \widehat{\nu}_1^{(l)}\|_{L_1(\mathbb{R})}, \qquad l = 1, \dots, L.$$

Denote by $\widetilde{\nu}_1 = \widehat{\nu}_1^{l^*}$ a new adaptive estimate for $\bar{\nu}_1$. In our implementation of the "quasi-optimality" approach, we take $h_l = 0.5 + 0.1 \times l$, $l = 1, \dots, 40$. In Figure 1, the sequence $f(l)$, $l = 1, \dots, 40$, is plotted. On the right-hand side of Figure 1, we show the resulting estimate $\widetilde{\nu}_1$ together with the true function $\bar{\nu}_1$. Based on the estimate $\widetilde{\nu}_1$, one can estimate some functionals of $\bar{\nu}_1$. For example, we have $\int \widetilde{\nu}_1(x)\,dx = 1.049053$ $[\int \bar{\nu}_1(x)\,dx = 1.015189]$.

5.2. *Time change via an integrated CIR process.* Another possibility to construct a time-changed Lévy process from the NIG Lévy process L_t is to use a time change of the form (2.2) with some rate process $\rho(t)$. A possible candidate for the rate of the time change is given by the Cox–Ingersoll–Ross process (CIR process). The CIR process is defined as a solution of the following SDE:

$$dZ_t = \kappa(\eta - Z_t)\,dt + \zeta\sqrt{Z_t}\,dW_t, \qquad Z_0 = 1,$$

where W_t is a Wiener process. This process is mean reverting with $\kappa > 0$ being the speed of mean reversion, $\eta > 0$ being the long-run mean rate and $\zeta > 0$ controlling the volatility of Z_t. Additionally, if $2\kappa\eta > \zeta^2$ and Z_0 has Gamma distribution, then Z_t is stationary and exponentially α-mixing [see, e.g., Masuda (2007)]. The time change $\mathcal{T}$ is then defined as

$$\mathcal{T}(t) = \int_0^t Z_t \, dt.$$

Simple calculations show that the Laplace transform of $\mathcal{T}(t)$ is given by

$$\mathcal{L}_t(z) = \frac{\exp(\kappa^2 \eta t/\zeta^2) \exp(-2z/(\kappa + \gamma(z) \coth(\gamma(z)t/2)))}{(\cosh(\gamma(z)t/2) + \kappa \sinh(\gamma(z)t/2)/\gamma(z))^{2\kappa\eta/\zeta^2}}$$

with $\gamma(z) = \sqrt{\kappa^2 + 2\zeta^2 z}$. It is easy to see that $\mathcal{L}_t(z) \asymp \exp(-\frac{\sqrt{2z}}{\zeta}[1 + t\kappa\eta])$ as $|z| \to \infty$ with $\operatorname{Re} z \geq 0$. Moreover, it can be shown that $\mathrm{E}|\mathcal{T}(t)|^p < \infty$ for any $p \in \mathbb{R}$. Let L_t be again a three-dimensional NIG Lévy process with independent components distributed as in Section 5.1. Construct the time-changed process $Y_t = L_{\mathcal{T}(t)}$. Note that the process Y_t is not any longer a Lévy process and has in general dependent increments. Let us estimate $\bar{\nu}_1$, the transformed Lévy density of the first component of L_t. First, note that according to Theorem 4.4, the estimate $\widehat{\nu}_1$ constructed as described in Section 3.2, has the following logarithmic convergence rates

$$\|\bar{\nu}_1 - \widehat{\nu}_1\|_{L_\infty(\mathbb{R},w)} = O_{\text{a.s.}}(\log^{-2(2-\delta)}(n)), \qquad n \to \infty,$$

for arbitrary small $\delta > 0$, provided the bandwidth sequence is chosen in the optimal way. Finite sample performance of $\widehat{\nu}_1$ with the choice of h_n based on the "quasi-optimality" approach is illustrated in Figure 2 where the sequence of estimates $\widehat{\nu}_1^{(1)}, \dots, \widehat{\nu}_1^{(L)}$ was constructed from the time series $Y_\Delta, \dots, Y_{n\Delta}$ with $n = 5{,}000$ and $\Delta = 0.1$. The parameters of the used CIR process are $\kappa = 1$, $\eta = 1$ and $\zeta = 0.1$. Again we can compute some functionals of $\widetilde{\nu}_1$. We have, for example, following estimates for the integral and for the mean of $\bar{\nu}_1$: $\int \widetilde{\nu}_1(x)\, dx = 1.081376$ [$\int \bar{\nu}_1(x)\, dx = 1.015189$] and $\int x\widetilde{\nu}_1(x)\, dx = -0.4772505$ [$\int x\bar{\nu}_1(x)\, dx = -0.3057733$].

Let us now test the performance of estimation algorithm in the case of a time-changed NIG process (parameters are the same as before), where the time change is again given by the integrated CIR process with the parameters $\eta = 1$, $\zeta = 0.1$ and $\kappa \in \{0.05, 0.1, 0.5, 1\}$. Figure 3(left) shows the boxplots of the resulting error $\|\bar{\nu}_1 - \widetilde{\nu}_1\|_{L_\infty(\mathbb{R},w)}$ computed using 100 trajectories each of the length $n = 5{,}000$, where the time span between observation is $\Delta = 0.1$. Note that if our time units are days, then we get about two years of observations with about one mean reversion per month in the case $\kappa = 0.05$. As one can see, the performance of the algorithm remains reasonable for the whole range of κ. In Figure 3(right), we present the boxplots of the error $\|\bar{\nu}_1 - \widetilde{\nu}_1\|_{L_\infty(\mathbb{R},w)}$ in the case of $\eta = 1$, $\zeta = 0.1$, $\kappa = 1$ and $n \in$

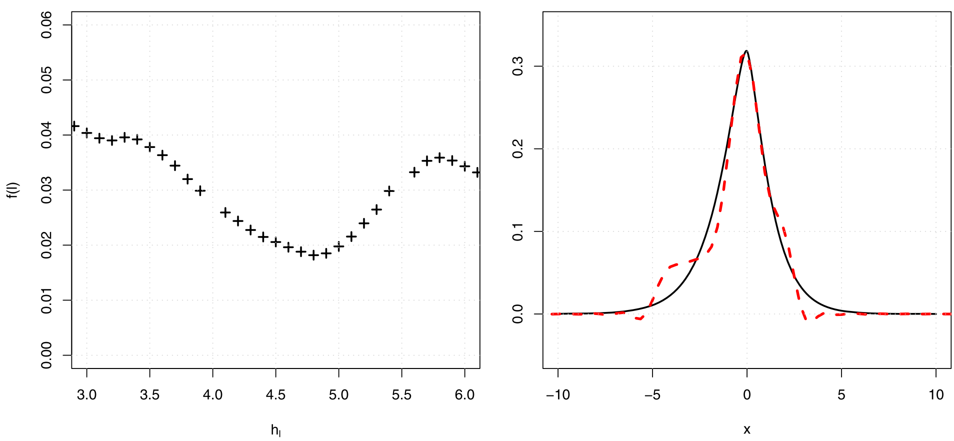

FIG. 2. *Left-hand side: objective function $f(l)$ for the "quasi-optimality" approach versus the corresponding bandwidths h_l. Right-hand side: adaptive estimate $\widetilde{\nu}_1$ (dashed line) together with the true function $\bar{\nu}_1$ (solid line).*

{500, 1,000, 3,000, 5,000}. As one can expect, the performance of the algorithm becomes worse as n decreases. However, the quality of the estimation remains reasonable even for $n = 500$.

6. Proofs of the main results.

6.1. *Proof of Theorem* 4.4. For simplicity, let consider the case of $\mu_l > 0$ and $\sigma_k = 0$. By Proposition 7.4 [take $G_n(u,z) = \exp(\mathrm{i}uz)$, $L_n = \bar{\mu}_n = \bar{\sigma}_n = 1$, $a = 0, b = 1$]

$$\mathbb{P}(|\widehat{\phi}_l(\mathbf{0})| \le \kappa/\sqrt{n}) \ge \mathbb{P}(|\widehat{\phi}_l(\mathbf{0}) - \phi_l(\mathbf{0})| > \mu_l) \le Bn^{-1-\delta}$$

for some constants $\delta > 0$, $B > 0$ and n large enough. Furthermore, simple calculations lead to the following representation:

$$\begin{aligned}\psi_k''(u) - \widehat{\psi}_{k,2}(u) &= \frac{\psi_k''(u)}{\psi_l'(0)}(\phi_l(\mathbf{0}) - \widehat{\phi}_l(\mathbf{0})) \\ &\quad + \mathcal{R}_0(u) + \mathcal{R}_1(u) + \mathcal{R}_2(u),\end{aligned} \tag{6.1}$$

where

$$\begin{aligned}\mathcal{R}_0(u) = {} & [V_1(u)\psi_k''(u) - V_2(u)\psi_k'(u)](\phi_l(u^{(k)}) - \widehat{\phi}_l(u^{(k)})) \\ & + V_2(u)(\phi_k(u^{(k)}) - \widehat{\phi}_k(u^{(k)})) \\ & - V_1(u)(\phi_{kk}(u^{(k)}) - \widehat{\phi}_{kk}(u^{(k)})) \\ & + V_1(u)\psi_k'(u)(\phi_{lk}(u^{(k)}) - \widehat{\phi}_{lk}(u^{(k)})),\end{aligned}$$

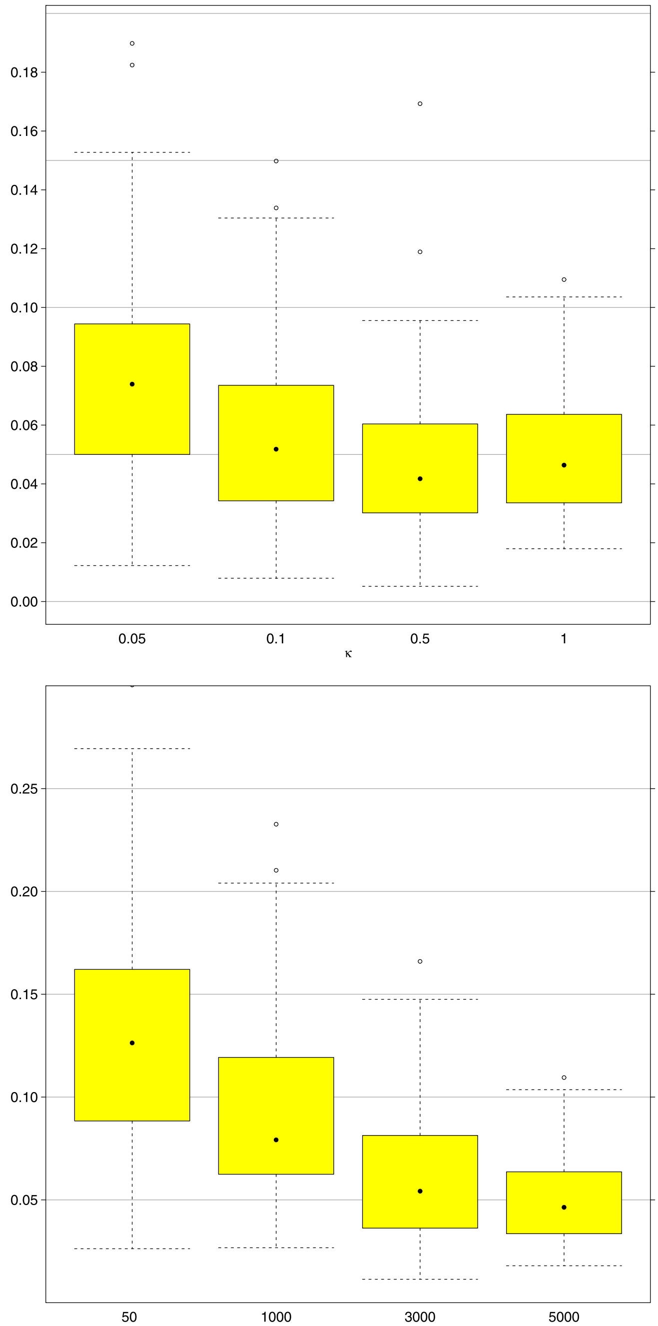

FIG. 3. *Boxplots of the error* $\|\bar{\nu}_1 - \widetilde{\nu}_1\|_{L_\infty(\mathbb{R},w)}$ *for different values of the mean reversion speed parameter* κ *and different numbers of observations* n.

$$\begin{aligned}\mathcal{R}_1(u) &= [\widetilde{V}_1(u)\psi_k''(u) - \widetilde{V}_2(u)\psi_k'(u)](\phi_l(u^{(k)}) - \widehat{\phi}_l(u^{(k)}))\\ &\quad + \widetilde{V}_2(u)(\phi_k(u^{(k)}) - \widehat{\phi}_k(u^{(k)}))\\ &\quad - \widetilde{V}_1(u)(\phi_{kk}(u^{(k)}) - \widehat{\phi}_{kk}(u^{(k)}))\\ &\quad + \widetilde{V}_1(u)\psi_k'(u)(\phi_{lk}(u^{(k)}) - \widehat{\phi}_{lk}(u^{(k)})),\\ \mathcal{R}_2(u) &= \Gamma^2(u)\frac{\phi_l(\mathbf{0})(\phi_{lk}(u^{(k)}) - \widehat{\phi}_{lk}(u^{(k)}))}{[\phi_l(u^{(k)})]^2}\\ &\quad \times [(\phi_l(u^{(k)}) - \widehat{\phi}_l(u^{(k)}))\psi_k'(u) - (\phi_k(u^{(k)}) - \widehat{\phi}_k(u^{(k)}))]\\ &\quad + \frac{(\widehat{\phi}_l(\mathbf{0}) - \phi_l(\mathbf{0}))}{\phi_l(u^{(k)})}\left[\frac{\mathcal{R}_0 + \mathcal{R}_1}{\phi_l(\mathbf{0})}\right]\end{aligned}$$

with

$$\begin{aligned} V_1(u) &= \frac{\phi_l(\mathbf{0})}{\Delta\phi_l(u^{(k)})} = -\frac{1}{\mathcal{L}_\Delta'(-\psi_k(u))},\\ V_2(u) &= \frac{\phi_l(\mathbf{0})\phi_{lk}(u^{(k)})}{\Delta[\phi_l(u^{(k)})]^2} = -V_1(u)\psi_k'(u)\frac{\mathcal{L}_\Delta''(-\psi_k(u))}{\mathcal{L}_\Delta'(-\psi_k(u))},\\ \widetilde{V}_1(u) &= (\Gamma(u) - 1)V_1(u), \qquad \widetilde{V}_2(u) = (\Gamma^2(u) - 1)V_2(u)\end{aligned}$$

and

$$\Gamma(u) = \left[1 - \frac{1}{\phi_l(u^{(k)})}(\phi_l(u^{(k)}) - \widehat{\phi}_l(u^{(k)}))\right]^{-1}.$$

The representation (6.1) and the Fourier inversion formula imply the following representation for the deviation $\bar{\nu}_k - \widehat{\nu}_k$:

$$\begin{aligned}\bar{\nu}_k(x) - \widehat{\nu}_k(x) &= \frac{1}{2\pi}\frac{(\phi_l(\mathbf{0}) - \widehat{\phi}_l(\mathbf{0}))}{\psi_l'(0)}\int_{\mathbb{R}} e^{-\mathrm{i}ux}\psi_k''(u)\mathcal{K}(uh_n)\,du\\ &\quad + \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux}\mathcal{R}_0(u)\mathcal{K}(uh_n)\,du\\ &\quad + \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux}\mathcal{R}_1(u)\mathcal{K}(uh_n)\,du\\ &\quad + \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux}\mathcal{R}_2(u)\mathcal{K}(uh_n)\,du\\ &\quad + \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux}(1 - \mathcal{K}(uh_n))\psi_k''(u)\,du.\end{aligned}$$

First, let us show that

$$\sup_{x\in\mathbb{R}}\left|\int_{\mathbb{R}} e^{-\mathrm{i}ux}\mathcal{R}_1(u)\mathcal{K}(uh_n)\,du\right| = o_{\mathrm{a.s}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}\int_{-1/h_n}^{1/h_n}\mathfrak{R}_k^2(u)\,du}\right)$$

and

$$\sup_{x\in\mathbb{R}}\left|\int_{\mathbb{R}} e^{-\mathrm{i}ux}\mathcal{R}_2(u)\mathcal{K}(uh_n)\,du\right| = o_{\mathrm{a.s}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}\int_{\mathbb{R}}\mathfrak{R}_k^2(u)\,du}\right).$$

We have, for example, for the first term in $\mathcal{R}_1(u)$

$$\begin{aligned}&\left|\int_{\mathbb{R}} e^{-\mathrm{i}uz}(\Gamma(u)-1)V_1(u)\psi_k''(u)(\phi_l(u^{(k)})-\widehat{\phi}_l(u^{(k)}))\mathcal{K}(uh_n)\,du\right|\\ &\qquad\le \sup_{|u|\le 1/h_n}|\Gamma(u)-1|\sup_{u\in\mathbb{R}}\big[w(|u|)\big|\phi_l(u^{(k)})-\widehat{\phi}_l(u^{(k)})\big|\big]w^{-1}(1/h_n)\\ &\qquad\quad\times\int_{-1/h_n}^{1/h_n}|V_1(u)||\psi_k''(u)|\,du\end{aligned}$$

with $w(u)=\log^{-1/2}(e+u)$, $u\ge 0$. Fix some $\xi>0$ and consider the event

$$\mathcal{A}=\left\{\sup_{\{|u|\le 1/h_n\}}\big[w(|u|)\big|\widehat{\phi}_l(u^{(k)})-\phi_l(u^{(k)})\big|\big]\le\xi\sqrt{\frac{\log n}{n}}\right\}.$$

By assumption (AH), it holds on $\mathcal{A}$ that

$$\begin{aligned}\sup_{|u|<1/h_n}\left|\frac{\phi_l(u^{(k)})-\widehat{\phi}_l(u^{(k)})}{\phi_l(u^{(k)})}\right| &\le \xi M_n w^{-1}(1/h_n)\sqrt{\log n/n}\\ &= o(\sqrt{h_n}),\qquad n\to\infty,\end{aligned}$$

and hence

$$\sup_{\{|u|\le 1/h_n\}}|1-\Gamma(u)| = o(\sqrt{h_n}),\qquad n\to\infty. \tag{6.2}$$

Therefore, one has on $\mathcal{A}$ that

$$\begin{aligned}&\sup_{x\in\mathbb{R}}\left|\int_{-1/h_n}^{1/h_n} e^{-\mathrm{i}ux}(\Gamma(u)-1)V_1(u)\psi_k''(u)(\phi_l(u^{(k)})-\widehat{\phi}_l(u^{(k)}))\mathcal{K}(uh_n)\,du\right|\\ &\quad = o\left(\sqrt{\frac{h_n\log^2 n}{n}}\int_{-1/h_n}^{1/h_n}\mathfrak{R}_k(u)\,du\right) = o\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}\int_{-1/h_n}^{1/h_n}\mathfrak{R}_k^2(u)\,du}\right)\end{aligned}$$

since $\psi_k''(u)$ and $\mathcal{K}(u)$ are uniformly bounded on $\mathbb{R}$. On the other hand, Proposition 7.4 implies [on can take $G_n(u,z)=\exp(\mathrm{i}uz)$, $L_n=\bar{\mu}_n=\bar{\sigma}_n=1$, $a=0$, $b=1$]

$$\mathbb{P}(\bar{\mathcal{A}})\lesssim n^{-1-\delta'},\qquad n\to\infty,$$

for some $\delta' > 0$. The Borel–Cantelli lemma yields

$$\sup_{x\in\mathbb{R}}\left|\int_{-1/h_n}^{1/h_n} e^{-\mathrm{i}ux}(\Gamma(u)-1)V_1(u)\psi_k''(u)\big(\phi_l(u^{(k)})-\widehat{\phi}_l(u^{(k)})\big)\mathcal{K}(uh_n)\,du\right|$$
$$= o_{\text{a.s.}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}\int_{-1/h_n}^{1/h_n}\mathfrak{R}_k^2(u)\,du}\right).$$

Other terms in $\mathcal{R}_1$ and $\mathcal{R}_2$ can be analyzed in a similar way. Turn now to the rate determining term $\mathcal{R}_0$. Consider, for instance, the integral

$$\begin{aligned}&\int_{-1/h_n}^{1/h_n} e^{-\mathrm{i}ux}V_1(u)\psi_k''(u)\big(\phi_l(u^{(k)})-\widehat{\phi}_l(u^{(k)})\big)\mathcal{K}(uh_n)\,du\\&\quad=\frac{1}{nh_n}\sum_{j=1}^{n}\left[Z_j^l K_n\left(\frac{x-Z_j^k}{h_n}\right)-\mathrm{E}\left\{Z^l\frac{1}{h_n}K_n\left(\frac{x-Z^k}{h_n}\right)\right\}\right]=\mathcal{S}(x)\end{aligned}\tag{6.3}$$

with

$$K_n(z)=\int_{-1}^{1}e^{-\mathrm{i}uz}V_1(u/h_n)\psi_k''(u/h_n)\mathcal{K}(u)\,du.$$

Now we are going to make use of Proposition 7.4 to estimate the term $\mathcal{S}(x)$ on the r.h.s. of (6.3). To this end, let

$$G_n(u,z)=\frac{1}{h_n}K_n\left(\frac{u-z}{h_n}\right).$$

Since $\nu_k,\nu_l\in\mathfrak{B}_\gamma$ for some $\gamma>0$ [assumption (AL1)], the Lévy processes L_t^k and L_t^l possess infinitely smooth densities $p_{k,t}$ and $p_{l,t}$ which are bounded for $t>0$ [see Sato (1999), Section 28] and fulfill [see Picard (1997)]

$$\sup_{x\in\mathbb{R}}\{p_{k,t}(x)\}\lesssim t^{-1/\gamma},\qquad t\to 0,\tag{6.4}$$

$$\sup_{x\in\mathbb{R}}\{p_{l,t}(x)\}\lesssim t^{-1/\gamma},\qquad t\to 0.\tag{6.5}$$

Moreover, under assumption (AL2) [see Luschgy and Pagès (2008)]

$$\int|x|^m p_{k,t}(x)\,dx=O(t),\qquad\int|x|^m p_{l,t}(x)\,dx=O(t),\qquad t\to 0,\tag{6.6}$$

and

$$\begin{aligned}&\int|x|^m p_{k,t}(x)\,dx=O(t^m),\\&\int|x|^m p_{l,t}(x)\,dx=O(t^m),\qquad t\to+\infty,\end{aligned}\tag{6.7}$$

for any $2 \le m \le p$. As a result, the distribution of (Z^k, Z^l) is absolutely continuous with uniformly bounded density q_{kl} given by

$$q_{kl}(y,z) = \int_0^\infty p_{k,t}(y) p_{l,t}(z)\, d\pi(dt),$$

where π is the distribution function of the r.v. $\mathcal{T}(\Delta)$. The asymptotic relations (6.4)–(6.7) and assumption (AT1) imply

$$\begin{aligned}
\mathrm{E}[|Z^l|^2 |G_n(u, Z^k)|^2] &= \frac{1}{h_n^2} \int_{\mathbb{R}} \left| K_n\left(\frac{u-y}{h_n}\right)\right|^2 \left\{ \int_{\mathbb{R}} |z|^2 q_{kl}(y,z)\, dz \right\} dy \\
&\le \frac{C_0}{h_n} \int_{\mathbb{R}} |K_n(v)|^2\, dv \\
&\le C_1 \int_{-1/h_n}^{1/h_n} |V_1(u)|^2\, du
\end{aligned}$$

with some finite constants $C_0 > 0$ and $C_1 > 0$. Similarly,

$$\begin{aligned}
\mathrm{E}[|Z^k|^2 |G_n(u, Z^k)|^2] &\le C_2 \int_{-1/h_n}^{1/h_n} |V_1(u)|^2\, du, \\
\mathrm{E}[|Z^k|^4 |G_n(u, Z^k)|^2] &\le C_3 \int_{-1/h_n}^{1/h_n} |V_1(u)|^2\, du, \\
\mathrm{E}[|Z^k|^2 |Z^l|^2 |G_n(u, Z^k)|^2] &\le C_4 \int_{-1/h_n}^{1/h_n} |V_1(u)|^2\, du
\end{aligned}$$

with some positive constants C_2, C_3 and C_4. Define

$$\begin{aligned}
\bar{\sigma}_n^2 &= C \int_{-1/h_n}^{1/h_n} |V_1(u)|^2\, du, \\
\bar{\mu}_n &= \|\mathcal{K}\|_\infty \|\psi''\|_\infty \int_{-1/h_n}^{1/h_n} |V_1(u)|\, du, \\
L_n &= \|\mathcal{K}\|_\infty \|\psi''\|_\infty \int_{-1/h_n}^{1/h_n} |u||V_1(u)|\, du,
\end{aligned}$$

where $C = \max_{k=1,2,3,4}\{C_k\}$. Since $|V_1(u)| \to \infty$ as $|u| \to \infty$ and $h_n \to \infty$, we get $\bar{\mu}_n / \bar{\sigma}_n^2 = O(1)$. Furthermore, due to assumption (AH)

$$\bar{\mu}_n \lesssim h_n^{-1/2} \bar{\sigma}_n \lesssim n^{1/2-\delta/2} \bar{\sigma}_n, \qquad L_n \lesssim h_n^{3/2} \bar{\sigma}_n \lesssim n^{3/2} \bar{\sigma}_n, \qquad n \to \infty, \tag{6.8}$$

and $\bar{\sigma}_n = O(h_n^{-1/2} M_n) = O(n^{1/2})$. Thus, assumptions (AG1) and (AG2) of Proposition 7.4 are fulfilled. Assumption (AZ1) follows from Lemma 7.1 and assumption (AT1). Therefore, we get by Proposition 7.4

$$\mathbb{P}\left(\sup_{z \in \mathbb{R}} [w(|z|) |\mathcal{S}(z)|] \ge \xi \sqrt{\frac{\bar{\sigma}_n^2 \log^{3+\varepsilon} n}{n}} \right) \lesssim n^{-1-\delta'}$$

for some $\delta' > 0$ and $\xi > \xi_0$. Noting that

$$\bar{\sigma}_n^2 \le C \int_{-1/h_n}^{1/h_n} \Re_k^2(u)\,du,$$

we derive

$$\sup_{z\in\mathbb{R}}[w(|z|)|\mathcal{S}(z)|] = O_{\text{a.s.}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}\int_{-1/h_n}^{1/h_n} \Re_k^2(u)\,du}\right).$$

Other terms in $\mathcal{R}_0$ can be studied in a similar manner. Finally,

$$\|\widehat{\nu}_k - \bar{\nu}_k\|_{L_\infty(\mathbb{R},w)} = O_{\text{a.s.}}\left(\sqrt{\frac{\log^{3+\varepsilon} n}{n}\int_{-1/h_n}^{1/h_n} \Re_k^2(u)\,du}\right) + \frac{1}{2\pi}\int_{\mathbb{R}} |1-\mathcal{K}(uh_n)||\psi_k''(u)|\,du. \tag{6.9}$$

The second, bias term on the r.h.s. of (6.9) can be easily bounded if we recall that $\nu_k \in \mathfrak{S}_\beta$ and $\mathcal{K}(u) = 1$ on $[-a_K, a_K]$

$$\begin{aligned}\frac{1}{2\pi}\int_{\mathbb{R}} |1-\mathcal{K}(uh_n)||\psi_k''(u)|\,du &\lesssim h_n^\beta \int_{\{|u|>a_K/h_n\}} |u|^\beta |\mathbf{F}[\bar{\nu}_k](u)|\,du \\ &\lesssim h_n^\beta \int_{\mathbb{R}} (1+|u|^\beta)|\mathbf{F}[\bar{\nu}_k](u)|\,du, \qquad n\to\infty.\end{aligned}$$

6.2. *Proof of Theorem* 4.7. We have

$$\begin{aligned}\widehat{\nu}_k(x_0) - \bar{\nu}_k(x_0) &= \left[\frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux_0}\psi_k''(u)\mathcal{K}(uh_n)\,du - \bar{\nu}_k(x_0)\right] \\ &\quad + \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux_0}\big(\widehat{\psi}_{k,2} - \psi_k''(u)\big)\mathcal{K}(uh_n)\,du \\ &= J_1 + J_2\end{aligned}$$

Introduce

$$K(z) = \frac{1}{2\pi}\int_{-1}^{1} e^{\mathrm{i}uz}\mathcal{K}(u)\,du,$$

then by the Fourier inversion formula

$$\mathcal{K}(u) = \int_{\mathbb{R}} e^{-\mathrm{i}uz} K(z)\,dz. \tag{6.10}$$

Assumption (AK) together with the smoothness of $\mathcal{K}$ implies that $K(z)$ has finite absolute moments up to order $m \ge s$ and it holds that

$$\int K(z)\,dz = 1, \qquad \int z^k K(z)\,dz = 0, \qquad k = 1,\dots,m. \tag{6.11}$$

Hence

$$J_1 = \int_{-\infty}^{\infty} \bar{\nu}_k(x_0 + h_n v) K(v)\, dv - \bar{\nu}_k(x_0)$$

and

$$\begin{aligned} |J_1| &\leq \left| \int_{|v|>\delta/h_n} [\bar{\nu}_k(x_0) - \bar{\nu}_k(x_0 + h_n v)] K(v)\, dv \right| \\ &\quad + \left| \int_{|v|\leq\delta/h_n} [\bar{\nu}_k(x_0) - \bar{\nu}_k(x_0 + h_n v)] K(v)\, dv \right| \\ &= I_1 + I_2. \end{aligned}$$

Since $\|\bar{\nu}\|_\infty \leq C_{\bar{\nu}}$ for some constant $C_{\bar{\nu}} > 0$, we get

$$I_1 \leq 2C_{\bar{\nu}} \int_{|v|>\delta/h_n} |K(v)|\, dv \leq C_{\bar{\nu}} C_K (h_n/\delta)^m$$

with $C_K = \int_{\mathbb{R}} |K(v)||v|^m\, dv$. Further, by the Taylor expansion formula,

$$\begin{aligned} I_2 &\leq \left| \sum_{j=0}^{s-1} \frac{h_n^j \bar{\nu}_k^{(j)}(x_0)}{j!} \int_{|v|\leq\delta/h_n} K(v) v^j\, dv \right| \\ &\quad + \left| \int_{|v|\leq\delta/h_n} K(v) \left[\int_{x_0}^{x_0+h_n v} \frac{\bar{\nu}_k^{(s)}(\zeta)(\zeta - x_0)^{s-1}}{(s-1)!}\, d\zeta \right] dv \right| \\ &= I_{21} + I_{22}. \end{aligned}$$

First, let us bound I_{21} from above. Note that, due to (6.11),

$$I_{21} = \left| \sum_{j=0}^{s-1} \frac{h_n^j \bar{\nu}_k^{(j)}(x_0)}{j!} \int_{|v|>\delta/h_n} K(v) v^j\, dv \right|.$$

Hence,

$$\begin{aligned} I_{21} &\leq \left(\frac{h_n}{\delta}\right)^m \sum_{j=0}^{s-1} \frac{\delta^j |\bar{\nu}_k^{(j)}(x_0)|}{j!} \int_{|v|>\delta/h_n} |K(v)||v|^m\, dv \\ &\leq \left(\frac{h_n}{\delta}\right)^m L C_K \exp(\delta). \end{aligned}$$

Furthermore, we have for I_{22}

$$I_{22} \leq \frac{L h_n^s}{s!} \int_{|v|\leq\delta/h_n} |K(v)||v|^s\, dv.$$

Combining all previous inequalities and taking into account the fact that $m \geq s$, we derive

$$|J_1| \lesssim h_n^s, \qquad n \to \infty.$$

The stochastic term J_2 can handled along the same lines as in the proof of Theorem 4.4.

6.3. *Proof of Theorem* 4.9. Define

$$K_0(x) = \prod_{k=1}^{\infty} \left(\frac{\sin(a_k x)}{a_k x} \right)^2$$

with $a_k = 2^{-k}, k \in \mathbb{N}$. Since $K_0(x)$ is continuous at 0 and does not vanish there, the function

$$K(x) = \frac{1}{2\pi} \frac{\sin(2x)}{\pi x} \frac{K_0(x)}{K_0(0)}$$

is well defined on $\mathbb{R}$. Next, fix two positive numbers β and γ such that $\gamma \in (0, 1)$ and $0 < \beta < 1 - \gamma$. Consider a function

$$\Phi(u) = \frac{e^{\mathrm{i}x_0 u}}{(1 + u^2)^{(1+\beta)/2} \log^2(e + u^2)}$$

for some $x_0 > 0$ and define

$$\mu_h(x) = \int_{-\infty}^{\infty} \mu(x + zh) K(z) \, dz$$

for any $h > 0$, where

$$\mu(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\mathrm{i}xu} \Phi(u) \, du.$$

In the next lemma, some properties of the functions μ and μ_h are collected.

LEMMA 6.1. *Functions μ and μ_h have the following properties*:

(i) *μ and μ_h are uniformly bounded on $\mathbb{R}$*,
(ii) *for any natural $n > 0$*

$$\max\{\mu(x), \mu_h(x)\} \lesssim |x|^{-n}, \qquad |x| \to \infty, \tag{6.12}$$

that is, both functions $\mu(x)$ and $\mu_h(x)$ decay faster than any negative power of x,
(iii) *it holds*

$$x_0^2 \mu(x_0) - x_0^2 \mu_h(x_0) \geq D h^{\beta} \log^{-1}(1/h) \tag{6.13}$$

for some constant $D > 0$ and h small enough.

Fix some $\varepsilon > 0$ and consider two functions

$$\nu_1(x) = \nu_\gamma(x) + \frac{1 - \varepsilon}{(1 + x^2)^2} + \varepsilon \mu(x),$$

$$\nu_2(x) = \nu_\gamma(x) + \frac{1 - \varepsilon}{(1 + x^2)^2} + \varepsilon \mu_h(x),$$

where $\nu_\gamma(x)$ is given by

$$\nu_\gamma(x) = \frac{1}{(1+x^2)}\left[\frac{1}{x^{1+\gamma}}1\{x \geq 0\} + \frac{1}{|x|^{1+\gamma}}1\{x < 0\}\right].$$

Due to statements (i) and (ii) of Lemma 6.1, one can always choose ε in such a way that ν_1 and ν_2 stay positive on $\mathbb{R}_+$ and thus they can be viewed as the Lévy densities of some Lévy processes $L_{1,t}$ and $L_{2,t}$, respectively. It directly follows from the definition of ν_1 and ν_2 that $\nu_1, \nu_2 \in \mathfrak{B}_\gamma$. The next lemma describes some other properties of $\nu_1(x)$ and $\nu_2(x)$. Denote $\bar{\nu}_1(x) = x^2\nu_1(x)$ and $\bar{\nu}_2(x) = x^2\nu_2(x)$.

LEMMA 6.2. *Functions $\bar{\nu}_1(x)$ and $\bar{\nu}_2(x)$ satisfy*

$$\sup_{x\in\mathbb{R}} |\bar{\nu}_1(x) - \bar{\nu}_2(x)| \geq \varepsilon D h^\beta \log^{-1}(1/h) \tag{6.14}$$

and

$$\int_{-\infty}^{\infty} (1 + |u|^\beta)|\mathbf{F}[\bar{\nu}_i](u)|\, du < \infty, \qquad i = 1, 2, \tag{6.15}$$

that is, both functions $\nu_1(x)$ and $\nu_2(x)$ belong to the class $\mathfrak{S}_\beta$.

Let us now perform a time change in the processes $L_{1,t}$ and $L_{2,t}$. To this end, introduce a time change $\mathcal{T}(t)$, such that the Laplace transform of $\mathcal{T}(t)$ has following representation:

$$\mathcal{L}_t(z) = \mathrm{E}\big[e^{-z\mathcal{T}(t)}\big] = \int_0^\infty e^{-zy}\, dF_t(y),$$

where $(F_t, t \geq 0)$ is a family of distribution functions on $\mathbb{R}_+$ satisfying

$$1 - F_t(y) \leq 1 - F_s(y), \qquad y \in \mathbb{R}_+,$$

for any $t \leq s$. Denote by $\widetilde{p}_{1,t}$ and $\widetilde{p}_{2,t}$ the marginal densities of the resulting time-changed Lévy processes $Y_{1,t} = L_{1,\mathcal{T}(t)}$ and $Y_{2,t} = L_{2,\mathcal{T}(t)}$, respectively. The following lemma provides us with an upper bound for the χ^2-divergence between $\widetilde{p}_{1,t}$ and $\widetilde{p}_{2,t}$, where for any two probability measures P and Q the χ^2-divergence between P and Q is defined as

$$\chi^2(P, Q) = \begin{cases} \displaystyle\int \left(\frac{dP}{dQ} - 1\right)^2 dQ, & \text{if } P \ll Q, \\ +\infty, & \text{otherwise.} \end{cases}$$

LEMMA 6.3. *Suppose that the Laplace transform of the time change $\mathcal{T}(t)$ fulfills*

$$\big|\mathcal{L}_\Delta^{(k+1)}(z)/\mathcal{L}_\Delta^{(k)}(z)\big| = O(1), \qquad |z| \to \infty, \tag{6.16}$$

for $k = 0, 1, 2$, and uniformly in $\Delta \in [0, 1]$. Then

$$\chi^2(\widetilde{p}_{1,\Delta}, \widetilde{p}_{2,\Delta}) \lesssim \Delta^{-1}[\mathcal{L}'_\Delta(ch^{-\gamma})]^2 h^{(2\beta+1)}, \qquad h \to 0,$$

with some constant $c > 0$.

The proofs of Lemmas 6.1, 6.2 and 6.3 can be found in the preprint version of our paper Belomestny (2010a). Combining Lemma 6.3 with inequality (6.14) and using the well-known Assouad lemma [see, e.g., Theorem 2.6 in Tsybakov (2004)], one obtains

$$\liminf_{n\to\infty} \inf_{\widehat{\nu}} \sup_{\nu\in\mathfrak{B}_\gamma\cap\mathfrak{S}_\beta} \mathbb{P}\Big(\sup_{x\in\mathbb{R}} |\bar{\nu}(x) - \widehat{\nu}(x)| > h_n^\beta \log^{-1}(1/h_n)\Big) > 0$$

for any sequence h_n satisfying

$$n\Delta^{-1}[\mathcal{L}_t'(c \cdot h_n^{-\gamma})]^2 h_n^{(2\beta+1)} = O(1), \qquad n \to \infty.$$

7. Auxiliary results.

7.1. *Some results on time-changed Lévy processes.*

LEMMA 7.1. *Let L_t be a d-dimensional Lévy process with the Lévy measure ν and let $\mathcal{T}(t)$ be a time change independent of L_t. Fix some $\Delta > 0$ and consider two sequences $T_k = \mathcal{T}(\Delta k) - \mathcal{T}(\Delta(k-1))$ and $Z_k = Y_{\Delta k} - Y_{\Delta(k-1)}$, $k = 1, \dots, n$, where $Y_t = L_{\mathcal{T}(t)}$. If the sequence $(T_k)_{k\in\mathbb{N}}$ is strictly stationary and α-mixing with the mixing coefficients $(\alpha_T(j))_{j\in\mathbb{N}}$, then the sequence $(Z_k)_{k\in\mathbb{N}}$ is also strictly stationary and α-mixing with the mixing coefficients $(\alpha_Z(j))_{j\in\mathbb{N}}$, satisfying*

$$\alpha_Z(j) \le \alpha_T(j), \qquad j \in \mathbb{N}. \tag{7.1}$$

PROOF. Fix some natural k, l with $k + l < n$. Using the independence of increments of the Lévy process L_t and the fact that $\mathcal{T}$ is a nondecreasing process, we get $\mathrm{E}[\phi(Z_1, \dots, Z_k)] = \mathrm{E}[\widetilde{\phi}(T_1, \dots, T_k)]$ and

$$\begin{aligned}&\mathrm{E}[\phi(Z_1, \dots, Z_k)\psi(Z_{k+l}, \dots, Z_n)]\\ &\quad= \mathrm{E}[\widetilde{\phi}(T_1, \dots, T_k)\widetilde{\psi}(T_{k+l}, \dots, T_n)], \qquad k, l \in \mathbb{N},\end{aligned}$$

for any two functions $\phi : \mathbb{R}^k \to [0, 1]$ and $\psi : \mathbb{R}^{n-l-k} \to [0, 1]$, where $\widetilde{\phi}(t_1, \dots, t_k) = \mathrm{E}[\phi(L_{t_1}, \dots, L_{t_k})]$ and $\widetilde{\psi}(t_1, \dots, t_k) = \mathrm{E}[\psi(L_{t_1}, \dots, L_{t_k})]$. This implies that the sequence Z_k is strictly stationary and α-mixing with the mixing coefficients satisfying (7.1). □

7.2. *Exponential inequalities for dependent sequences.*

The following theorem can be found in Merlevéde, Peligrad and Rio (2009).

THEOREM 7.2. *Let $(Z_k, k \ge 1)$ be a strongly mixing sequence of centered real-valued random variables on the probability space $(\Omega, \mathcal{F}, P)$ with the mixing coefficients satisfying*

$$\alpha(n) \le \bar{\alpha} \exp(-cn), \qquad n \ge 1, \bar{\alpha} > 0, c > 0. \tag{7.2}$$

Assume that $\sup_{k\geq 1}|Z_k| \leq M$ *a.s., then there is a positive constant* C *depending on* c *and* $\bar{\alpha}$ *such that*

$$\mathbb{P}\left\{\sum_{i=1}^{n} Z_i \geq \zeta\right\} \leq \exp\left[-\frac{C\zeta^2}{nv^2 + M^2 + M\zeta\log^2(n)}\right]$$

for all $\zeta > 0$ *and* $n \geq 4$, *where*

$$v^2 = \sup_i\left(\mathrm{E}[Z_i]^2 + 2\sum_{j\geq i}\mathrm{Cov}(Z_i, Z_j)\right).$$

COROLLARY 7.3. *Denote*

$$\rho_j = \mathrm{E}\big[Z_j^2 \log^{2(1+\varepsilon)}(|Z_j|^2)\big], \qquad j = 1, 2, \ldots,$$

with arbitrary small $\varepsilon > 0$ *and suppose that all* ρ_j *are finite. Then*

$$\sum_{j\geq i}\mathrm{Cov}(Z_i, Z_j) \leq C\max_j \rho_j$$

for some constant $C > 0$, *provided* (7.2) *holds. Consequently, the following inequality holds*:

$$v^2 \leq \sup_i \mathrm{E}[Z_i]^2 + C\max_j \rho_j.$$

The proof can be found in Belomestny (2010a).

7.3. *Bounds on large deviations probabilities for weighted sup norms.* Let $Z_j = (X_j, Y_j)$, $j = 1, \ldots, n$, be a sequence of two-dimensional random vectors and let $G_n(u, z)$, $n = 1, 2, \ldots,$ be a sequence of complex-valued functions defined on $\mathbb{R}^2$. Define

$$\widehat{m}_1(u) = \frac{1}{n}\sum_{j=1}^{n} X_j G_n(u, X_j),$$

$$\widehat{m}_2(u) = \frac{1}{n}\sum_{j=1}^{n} Y_j G_n(u, X_j),$$

$$\widehat{m}_3(u) = \frac{1}{n}\sum_{j=1}^{n} X_j^2 G_n(u, X_j),$$

$$\widehat{m}_4(u) = \frac{1}{n}\sum_{j=1}^{n} X_j Y_j G_n(u, X_j).$$

PROPOSITION 7.4. *Suppose that the following assumptions hold*:

(AZ1) *The sequence* Z_j, $j = 1, \dots, n$, *is strictly stationary and is* α*-mixing with mixing coefficients* $(\alpha_Z(k))_{k \in \mathbb{N}}$ *satisfying*

$$\alpha_Z(k) \le \bar{\alpha}_0 \exp(-\bar{\alpha}_1 k), \qquad k \in \mathbb{N},$$

for some $\bar{\alpha}_0 > 0$ *and* $\bar{\alpha}_1 > 0$.

(AZ2) *The r.v.* X_j *and* Y_j *possess finite absolute moments of order* $p > 2$.

(AG1) *Each function* $G_n(u, z), n \in \mathbb{N}$ *is Lipschitz in* u *with linearly growing* (*in* z) *Lipschitz constant, that is, for any* $u_1, u_2 \in \mathbb{R}$

$$|G_n(u_1, z) - G_n(u_2, z)| \le L_n(a + b|z|)|u_1 - u_2|,$$

where a, b *are two nonnegative real numbers not depending on* n *and the sequence* L_n *does not depend on* u.

(AG2) *There are two sequences* $\bar{\mu}_n$ *and* $\bar{\sigma}_n$, *such that*

$$|G_n(u, z)| \le \bar{\mu}_n, \qquad (u, z) \in \mathbb{R}^2,$$

and all the functions

$$\mathrm{E}[(|X|^2 + |Y|^2)|G_n(u, X)|^2], \qquad \mathrm{E}[|X|^4|G_n(u, X)|^2], \qquad \mathrm{E}[|X|^2|Y|^2|G_n(u, X)|^2]$$

are uniformly bounded on $\mathbb{R}$ *by* $\bar{\sigma}_n^2$. *Moreover, assume that the sequences* $\bar{\mu}_n$, L_n *and* $\bar{\sigma}_n$ *fulfill*

$$\bar{\mu}_n/\bar{\sigma}_n^2 = O(1), \qquad \bar{\mu}_n/\bar{\sigma}_n = O(n^{1/2-\delta/2}), \qquad \bar{\sigma}_n^2 = O(n),$$
$$L_n/\bar{\sigma}_n = O(n^{3/2}), \qquad n \to \infty,$$

for some δ *satisfying* $2/p < \delta \le 1$.

Let w *be a symmetric, Lipschitz continuous, positive, monotone decreasing on* $\mathbb{R}_+$ *function such that*

$$0 < w(z) \le \log^{-1/2}(e + |z|), \qquad z \in \mathbb{R}. \tag{7.3}$$

Then there is $\delta' > 0$ *and* $\xi_0 > 0$, *such that the inequality*

$$\mathbb{P}\left\{\log^{-(1+\varepsilon)}(1 + \bar{\mu}_n)\sqrt{\frac{n}{\bar{\sigma}_n^2 \log n}}\|\widehat{m}_k - \mathrm{E}[\widehat{m}_k]\|_{L_\infty(\mathbb{R}, w)} > \xi\right\} \le Bn^{-1-\delta'} \tag{7.4}$$

holds for any $\xi > \xi_0$, *any* $k \in \{1, \dots, 4\}$, *some positive constant* B *depending on* ξ *and arbitrary small* $\varepsilon > 0$.

The proof of the proposition can be found in Belomestny (2010a).

REFERENCES

BARNDORFF-NIELSEN, O. E. (1998). Processes of normal inverse Gaussian type. *Finance Stoch*. **2** 41–68. MR1804664

BASAWA, I. V. and BROCKWELL, P. J. (1982). Nonparametric estimation for nondecreasing Lévy processes. *J. R. Stat. Soc. Ser. B Stat. Methodol*. **44** 262–269. MR0676217

BAUER, F. and REISS, M. (2008). Regularization independent of the noise level: An analysis of quasi-optimality. *Inverse Problems* **24** 055009. MR2438944

BELOMESTNY, D. (2010a). Statistical inference for multidimensional time-changed Lévy processes based on low-frequency data. Available at arXiv:1003.0275.

BELOMESTNY, D. (2010b). Spectral estimation of the fractional order of a Lévy process. *Ann. Statist*. **38** 317–351. MR2589324

BOCHNER, S. (1949). Diffusion equation and stochastic processes. *Proc. Natl. Acad. Sci. USA* **35** 368–370. MR0030151

CARR, P. and WU, L. (2004). Time-changed Lévy processes and option pricing. *Journal of Financial Economics* **71** 113–141.

CARR, P., GEMAN, H., MADAN, D. B. and YOR, M. (2003). Stochastic volatility for Lévy processes. *Math. Finance* **13** 345–382. MR1995283

CLARK, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* **41** 135–155. MR0415944

COMTE, F. and GENON-CATALOT, V. (2010). Nonparametric adaptive estimation for pure jump Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat*. **46** 595–617. MR2682259

ERDÉLYI, A. (1956). *Asymptotic Expansions*. Dover, New York. MR0078494

FIGUEROA-LÓPEZ, J. E. (2004). Nonparametric estimation of Lévy processes with a view towards mathematical finance. Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA.

FIGUEROA-LÓPEZ, J. E. (2009). Nonparametric estimation of time-changed Lévy models under high-frequency data. *Adv. in Appl. Probab*. **41** 1161–1188. MR2663241

HOROWITZ, J. L. and MAMMEN, E. (2007). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Ann. Statist*. **35** 2589–2619. MR2382659

JUDITSKY, A. B., LEPSKI, O. V. and TSYBAKOV, A. B. (2009). Nonparametric estimation of composite functions. *Ann. Statist*. **37** 1360–1404. MR2509077

KAMAE, T. and KRENGEL, U. (1978). Stochastic partial ordering. *Ann. Probab*. **6** 1044–1049. MR0512419

LUSCHGY, H. and PAGÈS, G. (2008). Moment estimates for Lévy processes. *Electron. Commun. Probab*. **13** 422–434. MR2430710

MASUDA, H. (2007). Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl*. **117** 35–56. MR2287102

MCNEIL, A. J. and NEŠLEHOVÁ, J. (2009). Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. *Ann. Statist*. **37** 3059–3097. MR2541455

MERLEVÉDE, F., PELIGRAD, M. and RIO, E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High Dimensional Probability V: The Luminy Volume* (C. Houdrè, et al., eds.) 153–178. *IMS Collections* **5** 273–292. IMS, Beachwood, Ohio.

MONROE, I. (1978). Processes that can be embedded in Brownian motion. *Ann. Probab*. **6** 42–56. MR0455113

NEUMANN, M. H. and REISS, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15** 223–248. MR2546805

PICARD, J. (1997). Density in small time for Levy processes. *ESAIM Probab. Stat*. **1** 357–389 (electronic). MR1486930

ROSENBAUM, M. and TANKOV, P. (2010). Asymptotic results and statistical procedures for time-changed Lévy processes sampled at hitting times. Available at arXiv:1007.1414.

RUBIN, H. and TUCKER, H. G. (1959). Estimating the parameters of a differential process. *Ann. Math. Statist.* **30** 641–658. MR0110174

SATO, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions. Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. Translated from the 1990 Japanese original, Revised by the author. MR1739520

STEUTEL, F. W. and VAN HARN, K. (2004). *Infinite Divisibility of Probability Distributions on the Real Line. Monographs and Textbooks in Pure and Applied Mathematics* **259**. Dekker, New York. MR2011862

TSYBAKOV, A. B. (2004). *Introduction à L'estimation Non-paramétrique. Mathématiques and Applications* (*Berlin*) [*Mathematics and Applications*] **41**. Springer, Berlin. MR2013911

WOERNER, J. H. C. (2003). Variational sums and power variation: A unifying approach to model selection and estimation in semimartingale models. *Statist. Decisions* **21** 47–68. MR1985651

DUISBURG-ESSEN UNIVERSITY
FORSTHAUSWEG 2, D-47057 DUISBURG
GERMANY
E-MAIL: denis.belomestny@uni-due.de