Journal of Econometrics 157 (2010) 257–271

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Nonparametric estimation for a class of Lévy processes

Song X. Chen [a,b,*], Aurore Delaigle [b,c], Peter Hall [b]

[a] *Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100871, China*

[b] *Department of Mathematics and Statistics, University of Melbourne, VIC, 3010, Australia*

[c] *Department of Mathematics, University of Bristol, Bristol, BS8 4JS, UK*

ARTICLE INFO

Article history:
Received 29 December 2008
Received in revised form
16 December 2009
Accepted 21 December 2009
Available online 11 January 2010

JEL classification:
C13

Keywords:
Deconvolution
Empirical characteristic function
Errors in variables
Financial data
Inverse problem
Jump process
Lévy process
Rates of convergence
Regression
Stable law

ABSTRACT

We consider estimation for a class of Lévy processes, modelled as a sum of a drift, a symmetric stable process and a compound Poisson process. We propose a nonparametric approach to estimating unknown parameters of our model, including the drift, the scale and index parameters in the stable law, the mean of the Poisson process and the underlying jump size distribution. We show that regression and nonparametric deconvolution methods, based on the empirical characteristic function, can be used for inference. Interesting connections are shown to exist between properties of our estimators and of those found in conventional deconvolution.

1. Introduction

Lévy processes (Sato, 1999) are popular models for analysing econometric and financial data, due to their ability to accommodate both continuous evolution and abrupt jumps in the underlying state variables. A major attraction of the Lévy-process approach is that it addresses a particularly wide range of non-continuous settings, yet it is sufficiently mathematically tractable to permit estimators of parameters and distributions to be constructed relatively simply. Indeed, the ability to accommodate jumps makes Lévy processes particularly suitable for modelling data arising in finance, since the price of a financial security can jump significantly upon receiving unexpected news. Jumps are known to contribute to non-normality, and excessive skewness and kurtosis in both the return and relative return distributions in financial data (Johannes, 2004). These features, and the importance of potential applications in finance, have motivated a recent surge of interest in statistical inference for Lévy processes. See, for example, the work of Jedidi (2001), Rachev (2003), Aït-Sahalia (2004), Woerner (2004, 2006) and Aït-Sahalia and Jacod (2007, 2008).

In its most general form, a Lévy process is a sum of a drift, a Gaussian component and a jump component, and the model for the jump measure is quite general. When the jump measure is modelled more narrowly it is often considered to be of either finite or infinite activity, depending on whether the Lévy measure on $\mathbb{R}$ is finite or not. Examples of the first type are compound Poisson processes, and examples of the second type include non-Gaussian stable processes. However, some financial processes are very complex and could involve jumps of both types. As indicated by Aït-Sahalia and Jacod (2007), a process of infinite activity, which displays many small jumps, could represent frictions due to the mechanics of trading, whereas a compound Poisson process could model infrequent arrival of information related to the asset, and both types of jumps may be present. In this paper we consider processes which can be modelled as a sum of a drift, a stable process (either Gaussian or of infinite activity) and a compound Poisson process. The relatively simple form of our model allows us

* Corresponding author at: Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100871, China. Tel.: +86 1062750427; fax: +86 1062753624.

E-mail addresses: csx@gsm.pku.edu.cn (S.X. Chen), A.Delaigle@ms.unimelb.edu.au (A. Delaigle), halpstat@ms.unimelb.edu.au (P. Hall).

0304-4076/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.12.005

to make simple inferences. We suggest an elementary, regression-based method for estimating the unknowns describing the drift, the scale of the stable process and the Poisson mean, and a simple, nonparametric deconvolution argument for accessing the distribution of jump size.

The deconvolution aspect of our methodology is motivated by the simple nature of the Lévy–Khintchine representation. That feature also has made characteristic-function based methods attractive in parametric approaches to inference in the same problem. See, for example, the discussion by Aït-Sahalia and Jacod (2007) of methods based on using the characteristic function as the basis for estimating equations, and work of Feuerverger and McDunnough (1981) and Singleton (2001). The connection to deconvolution gives the problem a number of intriguing theoretical aspects, which supplement the practical and statistical interest noted earlier.

Although the modelling problem that we treat is more difficult than standard deconvolution, in the sense that it requires estimation of a parametrised form for the "noise" distribution (in addition to nonparametric estimation of the "signal" distribution), the information–theoretic difficulty of estimating the unknown underlying distribution of the compound Poisson process turns out to be identical to that of estimating the "signal" distribution in conventional deconvolution. In particular, the distribution and density estimators typically converge at the same rate. On the other hand, estimation of the scale parameter can be actually a little simpler than in the case of deconvolution. This feature is reflected in the fact that the parameter estimators can converge a little faster. Our regression-based method is significantly simpler than techniques that are generally proposed for estimating scale in deconvolution problems. Indeed, the approach we develop is an attractive alternative that can also be used in the deconvolution setting, where it can produce improved convergence rates under slightly more stringent, but nevertheless realistic, conditions. Importantly, financial data are often available in significant quantity. Sample sizes of many hundreds are not uncommon, and much larger samples can sometimes be generated simply by gathering the data over a longer period, or on a finer scale. Therefore nonparametric methods, rather than the parametric approaches that are commonly advocated, are attractive.

Inference for special Lévy processes has been addressed in the past. Akritas and Johnson (1981) and Basawa and Brockwell (1982) considered estimation for pure jump processes where the jump sizes are directly observable. In this setting, no deconvolution is involved since the jump process is directly observable. Earlier, Rubin and Tucker (1959) considered estimation of the distribution of the Gaussian part plus the jump part, based on data observed virtually in the continuum. In the present paper we propose methods that are semiparametric in nature, in that they involve both explicitly defined parameters and other quantities that are defined only in terms of smoothness conditions. In particular, we suggest estimators of parameters in each component of the Lévy process, as well as nonparametric estimators of the underlying distribution of the compound Poisson process.

Related problems have been treated recently by Belomestny and Reiss (2006) and Neumann and Reiss (2009). Belomestny and Reiss (2006) employ a method that has similarities with our regression method, but they consider a particular case of ours, where the stable process is only permitted to be Gaussian. Moreover, they use spline-based characteristic function estimators to develop methodology that employs option data, which is quite different from our approach. Neumann and Reiss (2009) consider the more general form of Lévy processes with an abstract model for the jump measure. They use empirical characteristic functions as the basis for methods for time-series data to construct estimators of the drift parameter and of a function that combines the Gaussian volatility and the underlying jump measure. However, the more abstract form of Neumann and Reiss's approach results in a particularly challenging technique for estimating the jump measure. We develop methodology using a similar type of data to them, but the model simplifications that we make lead us to estimators which, we feel, are substantially simpler to apply.

The paper is organised as follows. In Section 2 we outline a Lévy model framework for our methodology for inference about parameters and about the jump size distribution in the compound Poisson process. Theoretical properties of our estimators are discussed in Section 3. Section 4 reports results of a simulation study and a case study on a real financial dataset. Technical details are deferred to the Appendix.

2. Model and methodology

2.1. Lévy process model

Suppose we observe n independent increments $X_1, \ldots, X_n$ of a Lévy process, where each X_j is distributed as X, which, via a representation of Lévy–Khintchine type, has characteristic function

$$\phi(t) = E\{\exp(\mathrm{i}tX)\} = \exp\Big[\mathrm{i}t\,b - |t|^{\alpha}c + \lambda\{\chi(t) - 1\}\Big]. \tag{2.1}$$

Here $\mathrm{i} = \sqrt{-1}$, $0 < \alpha \le 2$, b (the drift parameter) is a real number, c (which we call the scale parameter) and λ (the Poisson jump rate parameter) denote strictly positive constants, and χ is another characteristic function. Also, since we do not assume that a random variable with characteristic function χ has finite first moment, it is of interest to quantify its location in terms of a truncated mean:

$$a = a(p) = E\{ZI(|Z| \le p)\}, \tag{2.2}$$

where $p > 0$. The model (2.1) is in some respects more statistically tractable than that for a general Lévy process which uses an abstract jump measure, and at the same time allows both a general symmetric stable process with $\alpha \in (0, 2]$ and a compound Poisson process. By permitting a stable law at this point we allow the model to capture heavy-tailedness of the distribution, reflecting a feature that is commonly observed in financial data. In Section 2.4 we shall generalise this aspect of the specification.

In view of (2.1), X can be expressed as

$$X = b + W + Z_1 + \cdots + Z_N, \tag{2.3}$$

where N, W and $Z_1, Z_2, \ldots$ are independent random variables, W has a symmetric stable distribution with exponent α, the Z_js are distributed as Z, and N is Poisson-distributed with mean λ. Our goal is to estimate a, b, λ and the distribution of Z. In Section 2.2, we develop estimators for the case where α is known, and in Section 2.3 we discuss estimation of α, and argue that in many instances, replacing α by an estimator has negligible effect. In Section 2.4, we consider more general problems where the distribution of W is less tightly specified.

When $\alpha = 2$, W in (2.3) is Gaussian distributed and the Lévy model has a diffusion component; but when $\alpha \in (0, 2)$ the diffusion is replaced by an infinite activity stable process. In both settings the terms $Z_1 + \cdots + Z_N$ in (2.3) represent a finite activity compound Poisson process. Pure jump Lévy processes have been used previously to model financial data. Examples include hyperbolic and generalised hyperbolic processes, the normal inverse Gaussian, variance gamma and of course stable processes. See Cont and Tankov (2003) for discussion.

Remark 1 (*Alternative to* $\lambda\{\chi(t)-1\}$). In the particular case where $\alpha = 2$ it is conventional to replace the term $\lambda\{\chi(t) - 1\}$ in (2.1) by:

$$J(t) = \int_{-\infty}^{\infty} \left\{ e^{itx} - 1 - it\,xI(|x| \le 1) \right\} d\nu(x),$$

where the nonnegative Lévy measure ν is restricted only by the condition $\int x^2(1+x^2)^{-1}\nu(dx) < \infty$. When $\alpha = 2$ this form is appropriate for our proposed regression-based estimation of (b, c, λ), since the condition on ν guarantees that $J(t) = o(t^2)$ as $|t| \to \infty$, and so $J(t)$ is negligible relative to the term $t^2 c$ which appears in (2.1). However, in our formulation of the Lévy process model this is not appropriate in general. In particular, $J(t)$ can behave like $|t|^{\xi}$ for any $0 < \xi < 2$, and so if $\alpha < 2$, in particular if W, in (2.3), has a nonnormal stable distribution, $J(t)$ can be of the same size as the term in $|t|^{\alpha}c$. This confounds nonparametric inference, particularly in the more general cases taken up in Section 3.1.

2.2. Suggested methodology

Let $\psi = \log\phi$. It is clear from (2.1) that the real part of $\psi(t)$ is finite for all finite values of t, and so ϕ never vanishes. Therefore ψ can be estimated by taking logarithms of the empirical characteristic function, $\hat\phi$, of the data:

$$\hat\psi(t) = \log\{\hat\phi(t)\} = \log\left\{ \frac{1}{n}\sum_{j=1}^{n} \exp(it\,X_j) \right\}. \tag{2.4}$$

We shall write

$$\hat\psi(t) = it\,b - |t|^{\alpha}c + \lambda\{\chi(t) - 1\} + \epsilon(t), \tag{2.5}$$

where

$$\epsilon(t) = \hat\psi(t) - \psi(t), \tag{2.6}$$

which can also be written as

$$\Im\,\hat\psi(t) = \arg\hat\phi(t) = tb + \lambda\Im\,\chi(t) + \Im\,\epsilon(t),$$
$$\Re\,\hat\psi(t) = \log\left|\hat\phi(t)\right| = -|t|^{\alpha}c + \lambda\{\Re\,\chi(t) - 1\} + \Re\,\epsilon(t),$$

where $\Re$ and $\Im$ denote the real and imaginary part operators, respectively, and $\epsilon(t)$ is as in (2.6). This suggests two naive regressions that can be used to estimate b, λ and c:

$$\Im\,\hat\psi(t) = tb + \epsilon_1(t), \qquad \Re\,\hat\psi(t) = -\lambda - |t|^{\alpha}c + \epsilon_2(t), \tag{2.7}$$

where

$$\epsilon_1(t) = \lambda\Im\,\chi(t) + \Im\,\epsilon(t), \qquad \epsilon_2(t) = \lambda\Re\,\chi(t) + \Re\,\epsilon(t). \tag{2.8}$$

In practice, each regression needs to be implemented over a set $\mathcal{T}$ of values t for which $|t|$ is large but not too large. In particular, provided $|t|$ is not too large, the value of $\epsilon(t)$ in (2.6) will be small; and provided $|t|$ is not too small, the values of tb, in the first regression in (2.7), and of $-|t|^{\alpha}c$ in the second regression, will be of larger order than $\epsilon_1(t)$ and $\epsilon_2(t)$, respectively. In the two regressions at (2.7) we minimise

$$\int_{\mathcal{T}} \{\Im\,\hat\psi(t) - tb\}^2 dt, \qquad \int_{\mathcal{T}} \left\{\Re\,\hat\psi(t) - (-\lambda - |t|^{\alpha}c)\right\}^2 dt, \tag{2.9}$$

obtaining in the first case an estimator $\hat b$ of b, the latter given by (2.1), and, in the second case, estimators $\hat c$ and $\hat\lambda$ of c and λ. Note that, with probability 1, $\hat\psi$ diverges at only a finite number of places within any given interval, and that if $\hat t$ is one those places then $|\hat\psi(t)| \to \infty$ at a rate that is slower than any polynomial in $|t - \hat t|^{-1}$, as $t \to \hat t$. Therefore, with probability 1 the integrals in (2.9) are finite and well-defined.

In principle a weighted least-squares method could be used in place of the ordinary least-squares approach suggested by (2.9). However, estimation of the covariance functions of $\Im\,\hat\psi(t)$ and $\Re\,\hat\psi(t)$ requires estimation of ratios of unknown quantities, where the denominator is not bounded away from zero, and so presents empirical challenges which will outweigh performance advantages that the method might enjoy.

Substituting these estimators of b, c and λ into (2.5), ignoring the term $\epsilon(t)$ there, and solving for $\chi(t)$, we obtain an estimator of the latter function:

$$\hat\chi(t) = \begin{cases} 1 + \hat\lambda^{-1}\left\{\hat\psi(t) - it\hat b + |t|^{\alpha}\hat c\right\} & \text{if } |t| \le t_1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.10}$$

The truncation point $t_1 > 0$ is introduced here so as to assist removal of spurious oscillations in the tails, and to ensure integrability of $\hat\chi$. The value of t_1 will be permitted to diverge as sample size increases.

Our initial estimator of the density f of Z is given by

$$\hat f(x) = \frac{1}{2\pi}\Re\int_{-\infty}^{\infty} e^{-itx}\hat\chi(t)dt. \tag{2.11}$$

It can be enhanced, by truncation, to remove oscillations in the tails, and converted to a proper density $\tilde f$ by taking $\tilde f = \max(\hat f, 0)/\int\max(\hat f, 0)$. An estimator of the constant $a = a(p)$, at (2.2), is obtainable by Fourier inversion:

$$\hat a = \frac{1}{\pi}\int_{-\infty}^{\infty}\left(\frac{\sin pt}{t^2} - \frac{p\cos pt}{t}\right)\Im\,\hat\chi(t)dt, \tag{2.12}$$

where, on this occasion, the constant t_1 used to construct $\hat\chi$ is potentially different from that employed to define $\hat f$ and $\tilde f$.

An estimator $\widehat F$ of the distribution function F of Z is given by

$$\widehat F(x_1) - \widehat F(x_2) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\exp(-it\,x_1) - \exp(-it\,x_2)}{it}\hat\chi(t)dt, \tag{2.13}$$

where $\hat\chi$ is as at (2.10). Like $\hat f$, at (2.11), $\widehat F$ is based on a deconvolution argument which removes most of the influence that the stable part of the Lévy process has on the distribution of X.

Remark 2 (*Regression Type Estimators*). The regression models used at (2.7) and (2.8) to define estimators seem perhaps rather naive, and in particular, the polynomial models considered there may seem to be crude approximations. However, in practice they turn out to be effective. Asymptotically, this is clear since $\epsilon_1(t)$ and $\epsilon_2(t)$ tend to zero in the tails as $n \to \infty$, whereas the main regression terms are bounded away from zero. In practice, too, $|t|$ does not need to be very large for the terms $\epsilon_1(t)$ and $\epsilon_2(t)$ to be small compared to the main regression terms, unless b, c and λ are very small, where the problem clearly becomes almost unidentifiable.

2.3. Estimation of α

We could estimate α by extending the methodology in Section 2.2, for example by considering the second regression problem at (2.7) as one involving the three parameters c, λ and α. However, a better estimator of α, in fact an estimator with a polynomial rather than logarithmic convergence rate, is often achieved by considering the distribution of X to be regularly varying, and estimating the exponent of regular variation. In our work, where X denotes a random variable whose distribution is dominated, in the sense of both the characteristic function and the survival function, by a stable law, the exponent is identical to α. See, for example, the representation at (2.3), where W denotes the "component" of X with a stable distribution. If the common distribution of the Z_js in (2.3) is sufficiently light tailed and sufficiently smooth then the characteristic functions of X and W have the same behaviour for large t, and likewise the functions $P(|X| > x)$ and $P(|W| > x)$ have the same regular variation behaviour for large x.

The Hill (1975) estimator is a very popular method for estimating exponents of regular variation, although other techniques are also used. See also Csörgő et al. (1985) and Groeneboom et al. (2003), who developed kernel methods, and Aït-Sahalia and Jacod (2009) and Belomestny (in press), who considered alternative approaches for models that are moderately close to our own. To construct the Hill estimator, rank the absolute values $|X_j|$ of the data as $|X_{(n)}| \geq |X_{(n-1)}| \geq \cdots$, and take

$$\hat{\alpha} = \left(\frac{1}{r} \sum_{j=1}^{r} \log |X_{(n-j+1)}/X_{(n-r)}| \right)^{-1}, \tag{2.14}$$

where choice of r is at the discretion of the experimenter (see below). The ratio r/n is often referred to the "sample fraction".

It is known (see e.g. Hall, 1982) that if the distribution of $|X|$ satisfies

$$P(|X| > x) = D_1 x^{-\alpha} \left\{ 1 + O\left(x^{-D_2}\right) \right\} \tag{2.15}$$

as $x \to \infty$, for constants $D_1, D_2 > 0$, and if r is chosen so that $r/n^{\eta_1} \to \infty$ and $r/n^{1-\eta_2} \to 0$ for constants $\eta_1, \eta_2 > 0$, then there exists $\eta > 0$ such that

$$\hat{\alpha} = \alpha + O_p\left(n^{-\eta}\right). \tag{2.16}$$

The asymptotically optimal choice of r, as a function of α and D_2, is $r \sim \text{const.}\, n^{2D_2/(2D_2+\alpha)}$, and for this choice, $\eta = D_2/(2D_2 + \alpha)$. There is a variety of methods for choosing the smoothing parameter r, which generally achieve the convergence rate in (2.16) for this value of η. They include those suggested by Resnick and Stărică (1997), Drees and Kaufmann (1998), Danielsson et al. (2001), Guillou and Hall (2001) and De Sousa and Michailidis (2004), for example. In our numerical work (Section 4), we chose r according to the procedure proposed by Beirlant et al. (2002).

Theorem B.1 in section B.1 of a longer version of this paper (Chen et al., 2009) shows that, for a general class of distributions of W and for the Poisson jump model introduced in Section 2.1, property (2.15) holds and therefore (2.16) is valid. (Note that estimation of α from data on X requires the distribution of the Poisson jump to be sufficiently light tailed. It is sufficient for the jump distribution to have absolute moments of order at least α.) Hence we can construct an estimator of α (e.g. the Hill estimator) which enjoys a convergence rate that is polynomial in the sample size, n. Since, in the properties of our estimators of c, λ and the distribution of jump size, α appears only as an exponent of a quantity whose value does not exceed a power of $\log n$, in order of magnitude terms; and since in so-called "ordinary smooth" cases, discussed in Section 3.1, the convergence rates of those estimators are no faster than logarithmic; then, in such instances, replacing α by $\hat{\alpha}$ has no effect on first-order properties. In particular, having to estimate α does not impair the properties of asymptotic optimality that our methods enjoy. See Section 3.4 for more theoretical details.

In so-called "supersmooth" cases, where the characteristic function χ decreases at rate $\exp(-c_1|t|^\beta)$ as $|t| \to \infty$, with $\beta > \alpha$, the value of α can be estimated at rate $O_p(n^{\eta-(1/2)})$ for each $\eta > 0$. The procedure for inference amounts to treating the second regression in (2.8) as one that involves three unknown parameters, c, α and λ, and estimating those quantities simultaneously by least-squares. Theoretical results establishing the convergence rate $O_p(n^{\eta-(1/2)})$ are virtually identical to those treated in Case 2 in Section 3.1, and proofs are the same as those in Appendix A.1.

2.4. Generalisation of the stable-law part

In the previous sections we introduced estimators based on the model at (2.1), which was motivated by the assumption that the random variable W has exactly a stable law with shape parameter α. In particular,

$$E\left(e^{it W}\right) = \exp\left(-c|t|^{\alpha}\right) \tag{2.17}$$

for all t. Empirical studies have shown that financial data can have more complex features than those prescribed by (2.1). To accommodate more of these features we consider a generalised version of the model at (2.1), motivated by cases where the distribution of W is a mixture of that at (2.17) and other distributions:

$$P(W \leq w) = \pi_0 P(W_0 \leq w) + \pi_1 P(Q_1 \leq w) + \cdots + \pi_r P(Q_r \leq w). \tag{2.18}$$

Here the mixing proportions $\pi_0, \ldots, \pi_r$ are assumed to be all strictly positive and add to 1, W_0 has a non-normal stable distribution with characteristic function given by (2.17), the random variables $Q_1, \ldots, Q_r$ all have a finite moment $E|Q_j|^\eta$ for an appropriate $\eta > 0$, and their characteristic functions decrease to zero at rate $o\{\exp(-c|t|^\alpha)\}$ as $|t| \to \infty$. These Q_is can be viewed as "residual" disturbances of the stable law, and add more model generality than is provided by the symmetric stable law specification at (2.1). For example, we can take the Q_is to be themselves random variables having stable distributions with respective index parameters $\alpha_i > \alpha$, and means μ_i. In particular they allow us to accommodate asymmetry and more complex tail behaviour, and therefore they allow the model to adapt more readily to real financial datasets. It is even possible for the tails of the common distribution of the Q_js to be strictly heavier than those of W_0 but for the tails of the characteristic functions of W_0 and W to be asymptotic to one another. For example, this would be the case if each Q_j were distributed as $Q^{(1)} + Q^{(2)}$, where $Q^{(1)}$ was normal $N(0, 1)$, $Q^{(2)}$ had a stable law with exponent strictly less than α, and $Q^{(1)}$ and $Q^{(2)}$ were independent. (However, in this instance the Hill estimator suggested in Section 2.3 would not be appropriate for estimating α.)

The case where W is normal can be considered too, but under more restrictive conditions on the Q_i's. The latter condition ensures that the characteristic function of W has many of the high-frequency properties of its counterpart at (2.17), and so in at least some respects we should expect similar properties of estimators of b, c, λ, and the Poisson jump distribution, for the distribution of W determined by either (2.17) or (2.18). In Section 3.1 we shall investigate the extent of these similarities.

Note that with (2.18) specified as in the discussion above, the condition $E|Q_j|^\eta < \infty$ for some $\eta > 0$ and all j guarantees that $E\{\exp(it\, Q_j)\} = 1 + O(|t|^\eta)$ as $t \to 0$. Moreover, as $|t| \to \infty$,

$$\log\left\{E\left(e^{it W}\right)\right\} = \log\left[\pi_0 \exp\left(-c|t|^\alpha\right)\{1 + o(1)\}\right] = -c|t|^\alpha - \lambda_1 + O\{g(t)\},$$

where $\lambda_1 = -\log \pi_0$ and $g(t) \to 0$ as $|t| \to \infty$. Motivated by these considerations, we generalise (2.17) to the condition

$$\log\left\{E\left(e^{it W}\right)\right\} = -c|t|^\alpha + \begin{cases} O(|t|^\eta) & \text{as } t \to 0 \\ -\lambda_1 + O\{g(t)\} & \text{as } |t| \to \infty, \end{cases} \tag{2.19}$$

where $0 < \alpha \leq 2$, λ_1 is an arbitrary constant and $g(t)$ denotes a nonnegative function that decreases to zero as $|t| \to \infty$. In (2.19), η can be any fixed, arbitrarily small positive constant; in a neighbourhood of the origin we require only that $E(e^{it W}) = 1 + O(|t|^\eta)$ for some $\eta > 0$. An example of a distribution of W for which (2.19) holds is one where W is a mixture of a non-normal, stably distributed random variable with characteristic function $\exp(-c_1|t|^\alpha)$, and a normally distributed variate with characteristic function $\exp(-c_2 t^2)$, in proportions π and $1-\pi$, respectively. There (2.19) holds with $c = c_1\pi$ and $\lambda_1 = -\log \pi$.

With the changes indicated above, our representation for the characteristic function at (2.1) alters to:

$$\begin{aligned}\log\{\phi(t)\} &= \log[E\{\exp(\mathrm{i}t X)\}] \\ &= \mathrm{i}t\, b - |t|^\alpha c + \lambda\{\chi(t) - 1\} \\ &\quad + \begin{cases} O(|t|^\eta) & \text{as } t \to 0 \\ -\lambda_1 + O\{g(t)\} & \text{as } |t| \to \infty, \end{cases}\end{aligned} \tag{2.20}$$

where $\eta > 0$, $-\infty < \lambda_1 < \infty$ and $g(t) \to 0$ as $|t| \to \infty$. We continue to define all estimators as in Section 2.2, working under the fictitious assumption that ϕ is given by (2.1), but in Section 3.1 we discuss theoretical properties when ϕ satisfies the more general condition (2.20). If the model (2.20) is valid for a value $\eta > \alpha$ then the methods developed in Section 2.3 can be used to estimate α in that model. However, if (2.20) holds only for $\eta < \alpha$ then our estimator of α will generally not be consistent. In this case, as discussed in Section 2.3, we could estimate α by extending the regression methodology in Section 2.2, where, in the second regression problem at (2.7), we would treat α as an unknown parameter. Alternative approaches to estimating α have been suggested by Aït-Sahalia and Jacod (2009) and Belomestny (in press).

As we shall see there, the main difficulty that can arise with this generalisation is that the distribution of the stable-law part can be confounded with that of the Poisson jump part, resulting in the latter not being estimated consistently. However, estimators of the location, scale and shape of the stable part are consistent, and have good convergence rates, under general conditions. In these terms, our methodology is robust.

3. Theoretical properties

3.1. Properties of estimators of b, c and λ

Shortly, we shall derive properties of estimators under the general model (2.20), where α is known. Properties under the simpler model (2.1) are immediately obtained by taking $\lambda_1 = g(t) = 0$. Let $\mathcal{T}$ be the set of values t for which the two squared-error criteria at (2.9) are defined, and, given $s_2 > 0$, put

$$\delta_{n1} = \left(n^{-1}\log n\right)^{1/2} \exp\left(c s_2^\alpha\right), \tag{3.1}$$

$$\delta_{n2} = \int_{\mathcal{T}} \{|\chi(t)| + g(t)\}\mathrm{d}t + n^{-1/2} \int_{\mathcal{T}} \exp\left(c|t|^\alpha\right)\mathrm{d}t, \tag{3.2}$$

where c and α are as in (2.20). The estimators $\hat{b}$, $\hat{c}$ and $\hat{\lambda}$ are those defined by minimising the integrated squares at (2.9). We assume that:

(a) the independent and identically distributed data X_j have characteristic function ϕ, at (2.20);
(b) the random variable Z, with characteristic function χ, satisfies $E|Z|^\eta < \infty$ for some $\eta > 0$; and
(c) for constants s_1, s_2 and s_3 depending on n, $\mathcal{T} = [s_1, s_2]$, where both s_1 and s_2 diverge to infinity as $n \to \infty$, $s_1 = s_2 - s_3$ for some $s_3 \in (0, \rho s_2)$ with $\rho \in (0,1)$ denoting a fixed constant, and $\delta_{n1} \to 0$ as $n \to \infty$. (3.3)

Condition (a) just reminds us that we work under model (2.20); condition (b) is a mild assumption on the jump distribution; condition (c) states that we apply the two linear regressions on a positive interval, whose lower end s_1 is far enough from zero for $|\chi(t)|$ to be small, thus for the two linear relations to be approximately satisfied; and whose upper end s_2 is not too large so that $\hat{\psi}$ is sufficiently reliable on the interval. Moreover, s_1 and s_2 increase with the sample size because $\hat{\psi}(t)$ becomes reliable further in the tails as n increases, but s_2 should increase slowly enough for $\hat{\psi}(t)$ to remain sufficiently accurate.

The next theorem quantifies asymptotic distances between our estimators and their target values. See Appendix A.1 for a proof.

Theorem 3.1. *If (3.3) holds then*

$$\begin{aligned}&\hat{b} - b = O_p\left(s_2^{-1}s_3^{-1}\delta_{n2}\right), \qquad \hat{c} - c = O_p\left(s_2^{2-\alpha}s_3^{-3}\delta_{n2}\right), \\ &\hat{\lambda} - (\lambda + \lambda_1) = O_p(s_2^2 s_3^{-3}\delta_{n2}).\end{aligned} \tag{3.4}$$

In the next two sections we discuss the main implications of the theorem.

3.1.1. Consistency of estimators

It is immediately clear from the result for $\hat{\lambda}$ in (3.4) that passing from the model (2.1) to its more general form at (2.20) has confused the methodology as to the value of the intensity of the Poisson jump process. In particular, $\hat{\lambda}$ is treating λ as though it were the difference between the true intensity and the logarithm of the mixture probability π_0 in (2.18). In fact, in this case, λ is unidentifiable, as the example below will illustrate. This, however, is the main casualty; both b and c are estimated consistently in many cases of interest. If $\pi_0 = 1$, namely in the absence of the other mixture components in (2.18), $\hat{\lambda}$ is consistent for λ.

To appreciate the origins of the confusion, note that one extension of the finite mixture model at (2.18) is to the infinite mixture

$$\begin{aligned}P(W \le w) &= \pi_0 P(W_0 \le w) + \pi_1 P(W_0 + Z_1 \le w) + \cdots \\ &\quad + \pi_r P(W_0 + Z_1 + \cdots + Z_r \le w) + \cdots,\end{aligned}$$

where W_0 has the characteristic function at (2.17), each Z_j has characteristic function χ, the variables $W_0, Z_1, Z_2, \ldots$ are independent, and $\pi_j = \lambda_1^j \mathrm{e}^{-\lambda_1}/j!$. On this occasion the stably distributed random variable W_0 is contaminated by the sum of a Poisson number of jumps, the Poisson process here having intensity λ_1; and the model at (2.19) becomes:

$$\log\left\{E\left(\mathrm{e}^{\mathrm{i}t W}\right)\right\} = -c|t|^\alpha + \lambda_1\{\chi(t) - 1\}.$$

Correspondingly, the model at (2.20) for the distribution of the data X becomes:

$$\begin{aligned}\log\{\phi(t)\} &= \log[E\{\exp(\mathrm{i}t X)\}] \\ &= \mathrm{i}t\, b - |t|^\alpha c + (\lambda + \lambda_1)\{\chi(t) - 1\}.\end{aligned}$$

Clearly, the contamination of W_0 is of the same type as, and is confounded with, the Poisson jump process. It is perhaps not so surprising, then, that in more general cases the estimator $\hat{\lambda}$ captures the "overall intensity" $\lambda + \lambda_1$, rather than just the intensity of the Poisson jump process.

3.1.2. Rates of convergence of estimators

Using the notation $V = X$, $U = b + W$ and $T = Z$, our problem can be formulated as one where the goal is to estimate functionals of ϕ_T from observations on a variable V whose characteristic function satisfies $\phi_V = \exp\{\lambda(\phi_T - 1)\}\phi_U$. Thus, our problem has analogies with deconvolution problems, where the goal is to estimate functionals of the unknown characteristic function ϕ_T of a variable T, based on a sample of observations of a variable V whose characteristic function ϕ_V satisfies $\phi_V = \phi_T\phi_U$, and ϕ_U is the characteristic function of a variable U, usually referred to as a measurement error.

In the deconvolution literature, it is common to classify distributions in two types, ordinary smooth and supersmooth, respectively. In the terminology of Fan (1991), the distribution of Z is ordinary smooth if it characteristic function χ satisfies

$$|\chi(t)| \leq C_1(1+|t|)^{-\beta} \quad \text{for } -\infty < t < \infty, \tag{3.5}$$

where $C_1 > 0$ and $\beta > 0$. For example, Laplace and Gamma distributions, and their convolutions are ordinary smooth. The distribution of Z is supersmooth if χ satisfies

$$|\chi(t)| \leq C_1 \exp\left(-c_1|t|^{\beta}\right) \quad \text{for } -\infty < t < \infty, \tag{3.6}$$

where $c_1, C_1, \beta > 0$. For example, stable distributions, and their convolutions with other distributions, are supersmooth.

As in deconvolution problems, the rates of convergence of our estimators are dictated by the relative behaviour of g in (2.20), χ and the characteristic function of W in the tails. We distinguish between the cases where the distribution of Z is ordinary smooth or supersmooth, and where g decreases to zero polynomially or exponentially fast. Since we cannot possibly consider every possible combination here, we treat the most interesting cases. Other cases can be treated similarly.

Case 1. *Distribution of Z is ordinary smooth.* In the ordinary smooth setting, we interpret Theorem 3.1 in the case $g(t) = |t|^{-\beta_1}$. Here, the first term of $\delta_{2,n}$ is of order $O\{s_1^{-(\beta\wedge\beta_1)+1}\}$, and the second term is of order $O\{s_3 n^{-1/2}\exp(cs_2^{\alpha})\}$, with c as in (2.20). Therefore, for each of the three O_p quantities in (3.4), optimal choice of s_1 and s_2 involves a trade-off between two terms. Similarly as in Fan (1991), it is not hard to check that if we take $s_3 \asymp s_2 = (c_0 \log n)^{1/\alpha}$, where $c_0 < (2c)^{-1}$, then $\delta_{n2} = O\{s_1^{-(\beta\wedge\beta_1)+1}\}$. Defining $k_c = \alpha$, $k_b = 1$ and $k_{\lambda+\lambda_1} = 0$, and using (3.4), we deduce the following corollary.

Corollary 3.1 (*Z Ordinary Smooth*)**.** *Assume* (3.5) *and let* $g(t) = |t|^{-\beta_1}$. *Then, if* $s_3 \asymp s_2 = (c_0 \log n)^{1/\alpha}$, *where* $c_0 < (2c)^{-1}$, *we have, for* $\theta = b, c$ *and* $\lambda + \lambda_1$,

$$\hat{\theta} - \theta = O_p\left[(\log n)^{-\{k_\theta + (\beta\wedge\beta_1)\}/\alpha}\right]. \tag{3.7}$$

We shall show in Section 3.3 that when $\beta_1 \geq \beta$ these rates are optimal. In particular, the logarithmic convergence rate in (3.7) is not caused by the inefficiency of our method, but is inherent to the difficulty of the problem. In the financial setting which motivates our work, data are often in plentiful supply. Therefore logarithmic convergence rates are not necessarily a problem. In the case $\beta_1 < \beta$ the added term $O\{g(t)\}$ in (2.20) can be confounded with the Poisson jump contribution, and so the target quantities are unidentifiable. Indeed, as discussed earlier, the $O\{g(t)\}$ term could itself be the result of a Poisson jump process, even dominating the first in terms of its prominence in the inferential problem.

Case 2. *Distribution of Z is supersmooth.* In the supersmooth setting, we interpret Theorem 3.1 with $g(t) = \exp(-D|t|^{\beta_1})$. In this case it is not hard to check that $\delta_{n2} = O\{s_3\{\exp(-c_1 s_1^{\beta}) + \exp(-D s_1^{\beta_1})\}\} + O\{s_3 n^{-1/2}\exp(cs_2^{\alpha})\}$. This setting is more favourable since, as we shall see, polynomial convergence rates are possible when $0 < \alpha \leq \min(\beta, \beta_1)$. When $\alpha > \min(\beta, \beta_1)$, optimal convergence rates of our estimators are generally slower than $n^{-\eta}$ for any $\eta > 0$, but faster than $(\log n)^{-C}$ for all $C > 0$. Below we treat only the case $\alpha \leq \min(\beta, \beta_1)$, since it is distinctly different from Case 1 where only logarithmic convergence rates are possible. Further, since, as in Case 1, the target quantities are not identifiable when $\beta > \beta_1$, we simplify the discussion below by assuming that $\beta_1 > \beta$. In this case (or when $\beta = \beta_1$ with $D > c_1$), our estimators achieve optimal rates. The following corollary summarises the convergence rates of our estimators in this setting. It can be proved by standard arguments.

Corollary 3.2 (*Z Supersmooth*)**.** *Assume* (3.6) *and let* $g(t) = \exp(-D|t|^{\beta_1})$ *with* $\beta_1 > \beta$. (i) *If* $0 < \alpha < \beta$, $s_1 = (c_2 \log n)^{1/\beta}$ *and* $s_2 = (c_2 \log n)^{1/\beta} + c_3$ *where* $c_2, c_3 > 0$, *then provided* $c_1 c_2 \geq \frac{1}{2}$, *we have, for* $\theta = b, c$ *and* $\lambda + \lambda_1$,

$$\hat{\theta} - \theta = O_p\left(n^{\eta-(1/2)}\right) \tag{3.8}$$

for all $\eta > 0$. (ii) *If* $\beta = \alpha$, *then by taking* $c_2 = \{2(c+c_1)\}^{-1}$, $s_1 = (c_2 \log n)^{1/\alpha}$ *and* $s_2 = (c_2 \log n)^{1/\alpha} + c_3$ *where* $c_3 > 0$, *we have, for* $\theta = b, c$ *and* $\lambda + \lambda_1$,

$$\hat{\theta} - \theta = O_p\left(n^{\eta - c_1/\{2(c+c_1)\}}\right) \tag{3.9}$$

for all $\eta > 0$. *More generally, when using any positive value of* c_2 *for which* $c\,c_2 > \frac{1}{2}$, *the convergence rate is* $O_p(n^{\eta-\min\{c_1c_2,(1/2)-c\,c_2\}})$ *for all* $\eta > 0$.

Result (3.8) implies that the estimators $\hat{b}$, $\hat{c}$ and $\hat{\lambda}$ of b, c and $\lambda + \lambda_1$, are root-n consistent, up to an arbitrary polynomial factor. Property (3.9) entails similar performance, in that the exponent, $c_1/\{2(c+c_1)\}$, on the right-hand side of (3.9), increases to $1/2$ as the value of c_1, in (3.6), increases. As in the case of deconvolution kernel density estimators, the above choices of smoothing parameters suffice to establish optimality of our estimators (see Section 3.3). Of course, in practice, we need to choose the smoothing parameters in a data-driven way (see Section 4).

3.2. Properties of estimators of the distribution of Z

In the previous section we noted that, in perturbing the model at (2.1) to that at (2.20), we did not necessarily disturb the basic properties of estimators of b and c. However, estimators of λ could be inconsistent because the perturbation could be confused with (indeed, confounded with) the intrinsic variation acquired through the Poisson jump process. This is similarly a problem when estimating the distribution function F of Z. In fact, under the conditions imposed in Corollaries 3.1 and 3.2 above, where we take the perturbation $g(t)$ to equal $|t|^{-\beta}$ and $\exp(-c_1|t|^{\beta})$, respectively, F is not necessarily identifiable when $\lambda_1 \neq 0$. Therefore in this section we revert to the model at (2.1), and in particular, in references below to condition (3.3), it is assumed that (2.20) there is replaced by (2.1).

Theorem 3.2 below gives rates of convergence for the density estimator, $\hat{f}$, at (2.11), the distribution estimator $\widehat{F}$, at (2.13), and the estimator $\hat{a}$, at (2.12), of the truncated moment $a = a(p)$, at (2.2), in the case of ordinary smooth distributions. Theorem 3.3 treats the supersmooth case. We state the latter results only in the case where polynomial convergence rates are possible. In Theorems 3.2 and 3.3 we take the estimators $\hat{b}$, $\hat{c}$ and $\hat{\lambda}$, implicit in the definitions of $\hat{a}$, $\hat{f}$ and $\widehat{F}$ through the definition of $\hat{\chi}$ at (2.10), to be those obtained by minimising the integrated squares at (2.9). Note that the condition $\beta > 1$ used in Theorem 3.2 is not a strong condition as it merely amounts to assuming that the characteristic function χ is absolutely integrable.

Theorem 3.2 (*Z Ordinary Smooth*)**.** *Assume* (3.3)(a) *and* (3.3)(b), *and that* χ *satisfies* (3.5) *with* $C_1, \beta > 0$. *Put* $t_1 = (c_0 \log n)^{1/\alpha}$ *in* (2.10), *with* $c\,c_0 < \frac{1}{2}$. *Then*

$$\sup_{-\infty<x<\infty}\left|\widehat{F}(x) - F(x)\right| = O_p\left\{(\log n)^{-\beta/\alpha}\right\},$$
$$|\hat{a} - a| = O_p\left\{(\log n)^{-\beta/\alpha}\right\}. \tag{3.10}$$

Moreover, if $\beta > 1$ *then, for* $t_1 = (c_0 \log n)^{1/\alpha}$ *in* (2.10), *with* $c\,c_0 < \frac{1}{2}$,

$$\sup_{-\infty<x<\infty}\left|\hat{f}(x) - f(x)\right| = O_p\left\{(\log n)^{-(\beta-1)/\alpha}\right\}. \tag{3.11}$$

Theorem 3.3 (*Z Supersmooth*). *Assume (3.3)(a) and (3.3)(b), and that χ satisfies (3.6) with $c_1, C_1, \beta > 0$. Put $t_1 = (c_2 \log n)^{1/\beta}$ in (2.10).* (i) *If $0 < \alpha < \beta$ and $c_1 c_2 > \frac{1}{2}$ then for all $\eta > 0$,*

$$\sup_{-\infty<x<\infty} \left|\widehat{F}(x) - F(x)\right| = O_p\left(n^{\eta-(1/2)}\right),$$
$$|\hat{a} - a| = O_p\left(n^{\eta-(1/2)}\right), \tag{3.12}$$

$$\sup_{-\infty<x<\infty} \left|\hat{f}(x) - f(x)\right| = O_p\left(n^{\eta-(1/2)}\right). \tag{3.13}$$

(ii) *If $\alpha = \beta$ and $c_2 = \{2(c + c_1)\}^{-1}$ then (3.12) and (3.13) hold for all $\eta > 0$, provided we replace $O_p(n^{\eta-(1/2)})$ by $O_p(n^{\eta-c_1/\{2(c+c_1)\}})$ at each appearance.*

A proof of Theorem 3.2 is given in Appendix A.2. Theorem 3.3 can be derived similarly.

3.3. Optimality

The implications of lower bounds to convergence rates are strengthened if they are stated for relatively narrow classes of distributions. Therefore, to derive lower bounds it is appropriate to revert to the simpler model for ϕ at (2.1), ignoring the generalisation to (2.20) treated in Section 3.1.

Recall from Section 2.1 that the main problem we are addressing is that of estimating b, c, λ and the distribution with characteristic function χ, from data $X_1, \ldots, X_n$ with characteristic function $\phi = \gamma\tau$, where

$$\gamma(t) = \exp\left(it\,b - |t|^\alpha c\right), \qquad \tau(t) = \exp[\lambda\{\chi(t) - 1\}] \tag{3.14}$$

are respectively the characteristic functions of a stable law, centred at b rather than at the origin, and a Poisson number of variables with the distribution of Z, which has characteristic function χ. The problem of estimating the truncated moment a, defined at (2.2), is intrinsically as difficult as the problem of estimating the distribution function F of Z, and is a little more difficult than the problem of estimating b. This is reflected in the somewhat slower rate of convergence of estimators of a, relative to those of b, in the ordinary smooth case.

As we have seen in Section 3.1, the problem of estimating b, c, F and the corresponding density $f = F'$, from n independent data with characteristic function $\phi = \gamma\tau$, is connected to the conventional deconvolution problem, where the goal is to estimate those quantities from the same number of independent data from the distribution with characteristic function $\gamma\chi$. Typically, in deconvolution, γ is assumed to be completely known; that is, b and c are known. In this context the optimal convergence rates of estimators of F and f are well-known, and are achieved by kernel estimators. See, for example, Fan (1991) and Hall and Lahiri (2008). See also Li and Vuong (1998) and Butucea and Tsybakov (2007).

The optimal convergence rates that we obtain below in our context are the same as the rates for conventional deconvolution problems mentioned in the previous paragraph, and we will see that they are attained by our estimators. For $C_1, \beta > 0$ fixed, let $\mathcal{C}^{\text{os}}_{\text{dist}}(\beta)$ and $\mathcal{C}^{\text{os}}_{\text{dens}}(\beta)$ be the classes of distribution and density functions, respectively, for which the corresponding characteristic function χ satisfies (3.5), and let $\mathcal{C}^{\text{ss}}_{\text{dist}}(\beta)$ and $\mathcal{C}^{\text{ss}}_{\text{dens}}(\beta)$ be as $\mathcal{C}^{\text{os}}_{\text{dist}}(\beta)$ and $\mathcal{C}^{\text{os}}_{\text{dens}}(\beta)$, but with (3.5) replaced by (3.6). Here the subscripts os and ss stand for ordinary smooth and supersmooth, respectively.

Theorem 3.4. *Let $\inf_{\check{F}}$ and $\inf_{\check{f}}$ denote infima over all measurable functionals (regarded as estimators of F and of f, respectively) of n independent data drawn from the distribution with characteristic function ϕ at (2.1). Keep b, c, α and λ fixed in the definition of ϕ, but permit the characteristic function χ to vary. Then*

(i) *in the ordinary smooth case, for $C > 0$ sufficiently small, we have*

$$\limsup_{n\to\infty} \inf_{\check{F}} \sup_{F\in\mathcal{C}^{\text{os}}_{\text{dist}}(\beta)} P\left\{\sup_{-\infty<x<\infty} \left|\check{F}(x) - F(x)\right| > C(\log n)^{-\beta/\alpha}\right\} > 0, \tag{3.15}$$

$$\limsup_{n\to\infty} \inf_{\check{f}} \sup_{f\in\mathcal{C}^{\text{os}}_{\text{dens}}(\beta)} P\left\{\sup_{-\infty<x<\infty} \left|\check{f}(x) - f(x)\right| > C(\log n)^{-(\beta-1)/\alpha}\right\} > 0; \tag{3.16}$$

(ii) *in the supersmooth case, for all sufficiently small $\eta_1 > 0$ and all $\eta_2 > 0$, we have*

$$\limsup_{n\to\infty} \inf_{\check{F}} \sup_{F\in\mathcal{C}^{\text{ss}}_{\text{dist}}(\beta)} P\left\{\sup_{-\infty<x<\infty} \left|\check{F}(x) - F(x)\right| > \eta_1 n^{-1/2}\right\} > 0, \tag{3.17}$$

$$\limsup_{n\to\infty} \inf_{\check{f}} \sup_{f\in\mathcal{C}^{\text{ss}}_{\text{dens}}(\beta)} P\left\{\sup_{-\infty<x<\infty} \left|\check{f}(x) - f(x)\right| > n^{-\eta_2 - c_1/\{2(c+c_1)\}}\right\} > 0. \tag{3.18}$$

See Appendix A.3 for a proof. The arguments used in Appendix A.2 to prove Theorems 3.2 and 3.3 are readily employed to establish a version of those results uniformly over distributions and densities in $\mathcal{C}^{\text{os}}_{\text{dist}}(\beta)$ and $\mathcal{C}^{\text{os}}_{\text{dens}}(\beta)$ or $\mathcal{C}^{\text{ss}}_{\text{dist}}(\beta)$ and $\mathcal{C}^{\text{ss}}_{\text{dens}}(\beta)$, respectively. Note that the lower-bound convergence rates in (3.15)–(3.18) are identical to the upper-bound rates given in (3.10)–(3.13), respectively. (In the case of (3.12), (3.13), (3.17) and (3.18) we are asserting this equivalence only up to a factor n^η, where $\eta > 0$ is arbitrarily small.) Together, those results establish optimality of the convergence rates for distribution and density estimators at (3.10) and (3.11).

The situation is quite different as far as estimation of b, c and λ is concerned. In particular, there is no analogue of λ in the conventional deconvolution problem, and b does not really have a counterpart there either, since, there, the so-called error distribution (corresponding to the distribution with characteristic function γ, in (3.14)) is typically assumed to have zero mean. Estimation of c is sometimes treated in conventional deconvolution, but in cases where the distribution of Z is ordinary smooth, optimal convergence rates of estimators of c are slower, at $(\log\log n)/(\log n)$, than the rate $(\log n)^{-1-\beta/\alpha}$ given by (3.7). See Butucea and Matias (2005). Therefore, in the ordinary smooth case, estimation of c is a little "less difficult" in the problem discussed in this paper than in conventional deconvolution. In the supersmooth case the convergence rates of estimators of c are identical in the two problems, up to the factor $n^{-\eta}$ for $\eta > 0$ arbitrarily small.

In Theorem 3.5, below, we show that the convergence rates given in (3.7) for the estimators $\hat{b}$, $\hat{c}$ and $\hat{\lambda}$, defined in Section 2.2, are optimal. (Those upper-bound rates are readily extended so that they hold uniformly over the class of $\mathcal{C}$ of distributions defined below.) Likewise it can be proved that the rate given in (3.10) for $\hat{a}$ is optimal, and also that the rates given earlier in the supersmooth case are optimal up to the factor n^η for arbitrarily small η.

Let $\mathcal{C}$ be the class of distributions G whose characteristic functions ϕ have the form at (2.1), with the characteristic function χ satisfying (3.5) and with $c \in \mathcal{I}_+$ and $\lambda \in \mathcal{I}_+$, where $\mathcal{I}$ and $\mathcal{I}_+$ denote, respectively, a bounded interval and a strictly positive bounded interval. As in the discussion of Case 1 in Section 3.1, let $k_b = 1$, $k_c = \alpha$ and $k_\lambda = 0$. See section 5.4 for a proof of the next theorem.

Theorem 3.5. *Let θ denote either b, c or λ, and write $\inf_{\check{\theta}}$ for the infimum over functions $\check{\theta}$ of n independent data drawn from the distribution G with characteristic function ϕ at (2.1). Then for $C > 0$ sufficiently small and C_1 in (2.5) sufficiently large,*

$$\limsup_{n\to\infty} \inf_{\check{\theta}} \sup_{G\in\mathcal{C}} P\left\{|\hat{\theta} - \theta| > C(\log n)^{-(k_\theta+\beta)/\alpha}\right\} > 0. \quad (3.19)$$

3.4. Estimation of α

In a longer version of this paper (Chen et al., 2009) we show that, under assumptions on the characteristic function of X that are substantially less restrictive than requiring the distribution of W to be stable, the distribution of X, defined at (2.3), satisfies (2.15). As explained in Section 2.3, this means that α can be estimated at a polynomial rate; see (2.16). From that property it is straightforward to show that all our conclusions in the ordinary-smooth case continue to hold if, in the construction of our estimators in Section 2.2, α is estimated from data. In particular, if we replace α by $\hat{\alpha}$, defined at (2.14), in the second regression at (2.7) which was used to define c and λ, and employ the resulting slightly altered definitions of all our estimators, then results (3.7), (3.10) and (3.11) continue to hold. Likewise, the uniform versions of those properties, discussed in the paragraph below Theorem 3.4, remain valid.

4. Numerical properties

4.1. Simulation settings

We simulated data from various models, taking $\lambda = 0.1, 0.5$ or 1, W with a symmetric stable distribution having exponent $\alpha \in (0, 2]$, (i) $Z \sim N(-4, 25)$ and $b = 2$; (ii) $Z \sim$ Cauchy $(-5, 5)$, $b = 1$; (iii) $Z \sim \text{Lap}_3(7) - \mu_{\text{Lap}}$, $b = 10$, $\mu_{\text{Lap}} = 4$ or 15, where $\text{Lap}_k(\sigma)$ denotes the sum of k independent Laplace (i.e. double exponential) random variables with variance $2\sigma^2$. Number of simulation is 200; samples sizes $n = 500$, 1000 and 2000.

4.2. Data-driven procedures

To calculate $\hat{b}$, $\hat{c}$ and $\hat{\lambda}$ we need to choose $\mathcal{T}$. For estimation of b this can be done by plotting t against $\Im\,\hat{\psi}(t)$ on a range of values of t, and finding an interval close to $t = 0$ where the relation looks linear. That could be done automatically, but in practice it is most likely to be done visually.

Once the interval has been identified, $\hat{b}$ can be found by performing a simple least-squares linear regression (without intercept) on the set of observations $(t_i, \Im\,\hat{\psi}(t_i))_{i=1,\ldots,G}$, where $t_1, \ldots, t_G$ denotes a grid of equispaced points in that interval. In our simulations we take $G = 100$. The same method can be used to estimate c and λ, where we plot $|t|^\alpha$ against $\Re\,\hat{\psi}(t)$ to identify the interval nearest to the origin where the relation looks linear (the fact that c and λ should both be positive is helpful in identifying the interval), and then perform a least-squares linear regression on that interval. Visual procedures of this sort are not uncommon in problems of this type; see, for example, Crovella et al. (1998) and De Sousa and Michailidis (2004).

Finding such intervals is not difficult, as we illustrate in Appendix A.5. However, in our numerical work, since we could not possibly inspect the graphs for every Monte–Carlo replication, we used the same intervals for every replication, fixed as follows: we generated a few "test samples" and took the intervals such that, for most of the test samples, the graphs of $(t, \Im\,\hat{\psi}(t))$ and $(|t|^\alpha, \Re\,\hat{\psi}(t))$ looked linear. Even though, in most cases, this implied that we did not select the best possible intervals, the automatic method gave good results.

The smoothing parameter t_1 at (2.10) was chosen by minimising a bootstrap estimator of the mean integrated squared error, $E\int(\hat{f}-f)^2$, of $\hat{f}$. Since $\int f^2$ does not depend on t_1, it suffices to select t_1 via minimisation of an estimator of $E(\int\hat{f}^2) - 2E(\int\hat{f}f)$, constructed as follows: (1) Generate B bootstrap resamples $\{X^*_{b,1}, \ldots, X^*_{b,n}\}_{b=1,\ldots,B}$, of size n, by drawing randomly with replacement from the sample $X_1, \ldots, X_n$; (2) Select t_1 to minimise

$$B^{-1}\sum_{b=1}^{B}\int \hat{f}^*_b(x; t_1)^2 \mathrm{d}x - 2B^{-1}\sum_{b=1}^{B}\int \hat{f}^*_b(x; t_1)\hat{f}_P(x)\mathrm{d}x,$$

where $\hat{f}^*_b(x; t_1)$ denotes the estimator at (2.11) calculated from the resample $X^*_{b,1}, \ldots, X^*_{b,n}$, and using the truncation point t_1, and $\hat{f}_P(x)$ denotes a pilot estimator of f. To construct the latter we combine the use of a pilot truncating point t^*_1 with the idea of Neumann and Reiss (2009) of truncating $\hat{\chi}(t)$ to 0 when it gets too unreliable (knowing that we can not estimate χ at a rate faster than $\sqrt{n}$). We took

$$\hat{f}_P(x) = (2\pi)^{-1}\left[\int_{-\infty}^{\infty}\cos(tx)\Re\,\hat{\chi}(t; t^*_1)\cdot 1_{\{|\Re\,\hat{\chi}(t;t^*_1)|>n^{-1/2}\}}\mathrm{d}t + \int_{-\infty}^{\infty}\sin(tx)\Im\,\hat{\chi}(t; t^*_1)\cdot 1_{\{|\Im\,\hat{\chi}(t;t^*_1)|>n^{-1/2}\}}\mathrm{d}t\right],$$

where $\hat{\chi}(t; t^*_1)$ denotes the estimator (2.10) with $t_1 = t^*_1$.

To choose t^*_1, note that, from (2.7), (2.8) and (2.10), once $\Re\hat{\psi}$ and $\Im\hat{\psi}$ drift away from their respective fitted models, $\hat{\chi}$ is unreliable. Therefore, we should take $t^*_1 \le \min(T_1, T_2)$, where T_1 and T_2 are the largest zeroes of $\Re\,\hat{\psi}(t) - (-|t|^\alpha\hat{c} - \hat{\lambda})$ and $\Im\,\hat{\psi}(t) - t\hat{b}$, respectively. Moreover, for most common densities, the magnitude of $\Re\,\chi(t)$ at successful local maxima decreases as $|t|$ increases, at least for $|t|$ large enough (we could also use $\Im\,\chi(t)$, but at the pilot stage $\Re\,\chi(t)$ suffices). In particular, let $t_0 = \min\{t \ge 0 \text{ s.t. } |\Re\hat{\chi}(t)| \le n^{-1/2}\}$ (thus, t_0 determines a zone where we suspect $\Re\,\hat{\chi}(t)$ to become unreliable); and let $M_1 < M_2 < M_3 < \cdots$ denote the successful local maxima of $\Re\hat{\chi}(t)$ for $t \ge t_0$. If $\Re\hat{\chi}(M_j) < \Re\hat{\chi}(M_{j+1})$ for some j, this indicates that $\Re\hat{\chi}$ has become too unreliable. Therefore, we take $t^*_1 = \min(T_1, T_2, T_3)$, with T_1 and T_2 as above and $T_3 = \min\{M_j \text{ s.t. } \Re\hat{\chi}(M_j) < \Re\hat{\chi}(M_{j+1})\}$.

4.3. Simulation results

In Fig. 1, we present the results for estimating density (i) for various sample sizes, when $\lambda = 0.1$ or 0.5, $\alpha = 2$ and $c = 2$ or 4.5. The bottom row, which shows estimated quartile curves in various cases, illustrates the fact that the quality of the estimator improves as n and/or λ increases. The top row shows boxplots of 200 values of $\hat{\theta} - \theta$, for $\theta = b, c$ and λ for various settings. In each case, the results improve as the sample size increases and the problem is more difficult as c gets larger and as λ, the mean number of jumps, increases.

Fig. 2 shows similar results from density (ii), when $\alpha = 1.5$ or $\alpha = 2$, $\lambda = 0.5$ and c takes different values. Again, the estimators improve as n increases and/or as c decreases, and the data-driven procedure is seen to work well.

Finally, in Fig. 3 we examine the effect of estimating α, by comparing the results obtained when estimating density (iii) with $\alpha = 0.5$ or 0.75. Clearly, the estimators are less good when α is estimated, but the results remain reasonable, especially when n is large. Of course, this is not surprising since α is estimated from extreme observations, and the latter are more numerous for n large.

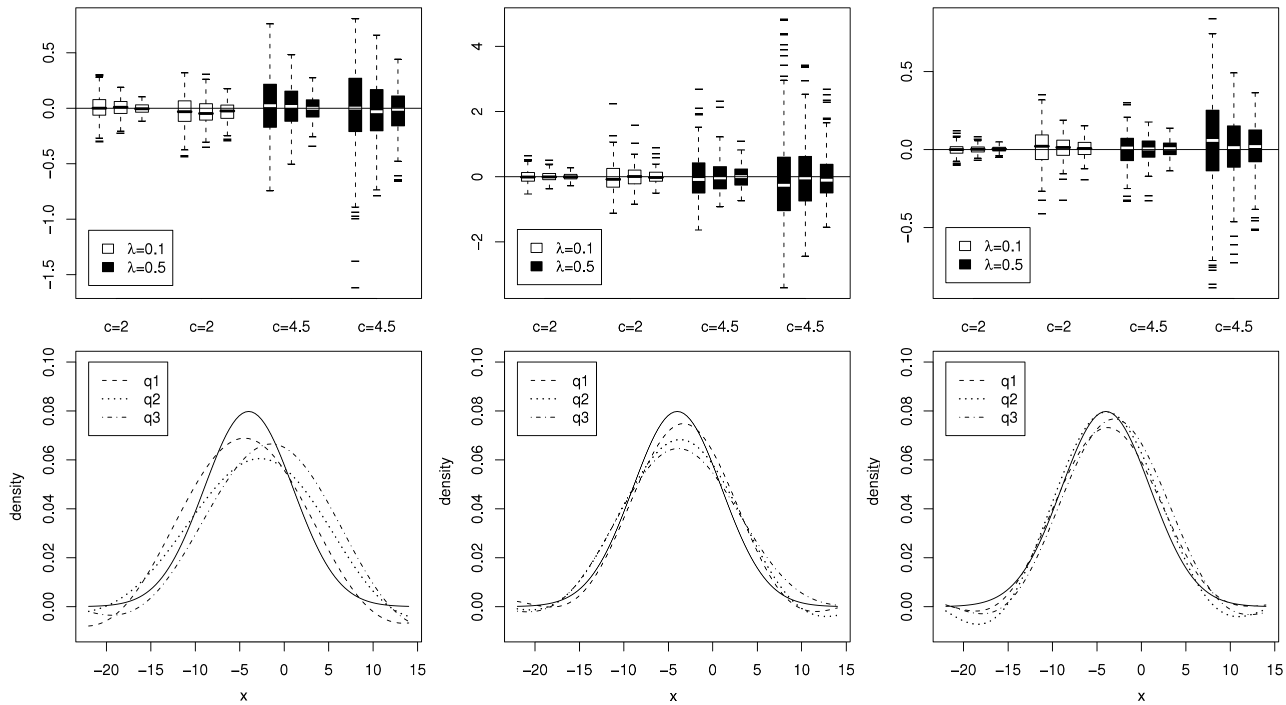

Fig. 1. Results for density (i). Top: Boxplots of $\hat{\theta} - \theta$, for $\theta = b$ (left), $\theta = c$ (middle) and $\theta = \lambda$ (right). In each case, $n = 500$ for boxes 1, 4, 7, 10, $n = 1000$ for boxes 2, 5, 8, 11, $n = 2000$ otherwise. Bottom: First ($q1$), second ($q2$) and third ($q3$) quartile curves when $c = 2$, with $(n, \lambda) = (1000, 0.1)$ (left), $(n, \lambda) = (1000, 0.5)$ (middle) and $(n, \lambda) = (2000, 0.5)$ (right). The solid lines represent the target curve.

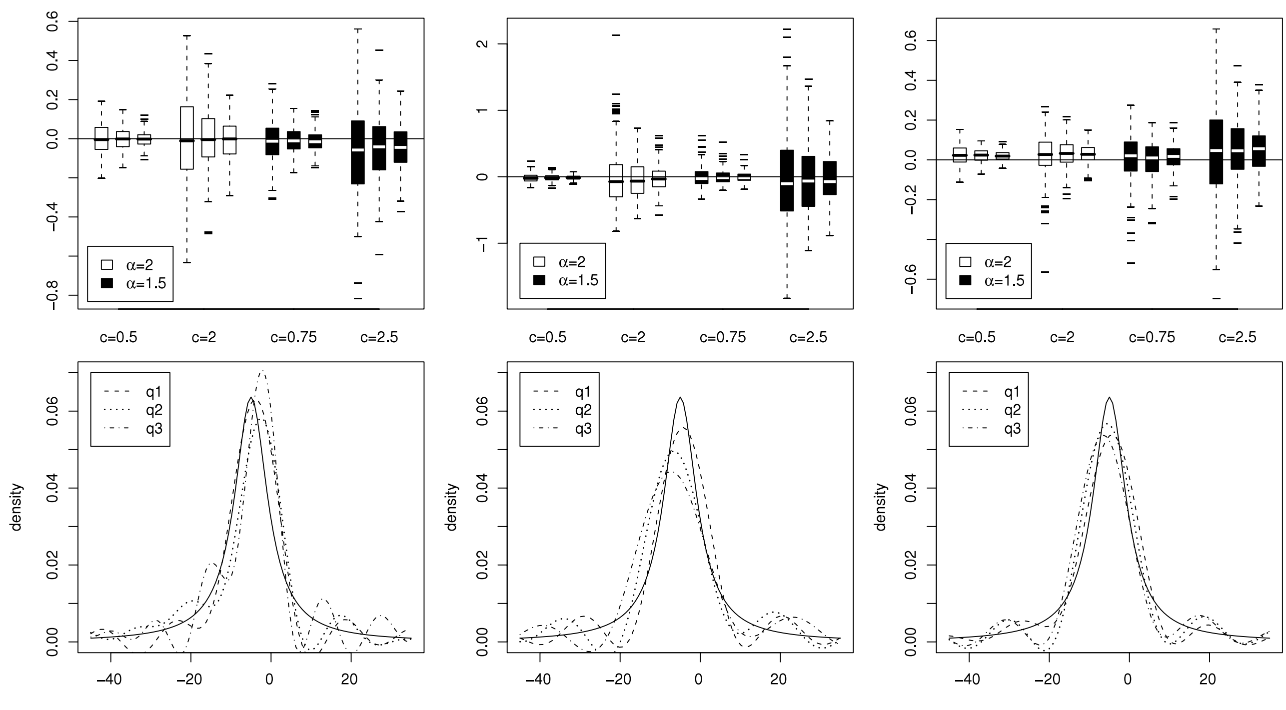

Fig. 2. Results for density (ii). Top: Boxplots of $\hat{\theta} - \theta$, for $\theta = b$ (left), $\theta = c$ (middle) and $\theta = \lambda$ (right) when $\lambda = 0.5$ and $\alpha = 2$ or 1.5. In each case, $n = 500$ for boxes 1, 4, 7, 10, $n = 1000$ for boxes 2, 5, 8, 11, $n = 2000$ otherwise. Bottom: First ($q1$), second ($q2$) and third ($q3$) quartile curves when $(\alpha, c, \lambda) = (1.5, 0.75, 0.5)$, with $n = 500$ (left), $n = 1000$ (middle) and $n = 2000$ (right). The solid lines represent the target curve.

4.4. Real data example

We applied our method to daily closing stock prices of PokerTek Inc (PTEK), downloaded from http://www.euroinvestor.fr. PokerTek is a US based company that specialises in electronic poker tables and related amusement products. The X_i's were taken to be the daily log returns between January 1 2002 and December 31, 2007, resulting in a sample of size $n = 1272$. The stocks were not traded for a period in 2005, but this did not cause a major discontinuity in the stock prices (between trading days 736 and 737). The data are plotted in Fig. 4 and indicate considerable activity,

We used the method of Hill (1975) to estimate α. In particular, we combined a plot of $\hat{\alpha}$, at (2.14), against r with a plot of $|t|^{\alpha}$ against $\Re\hat{\psi}(t)$. In the first plot we searched for zones where the

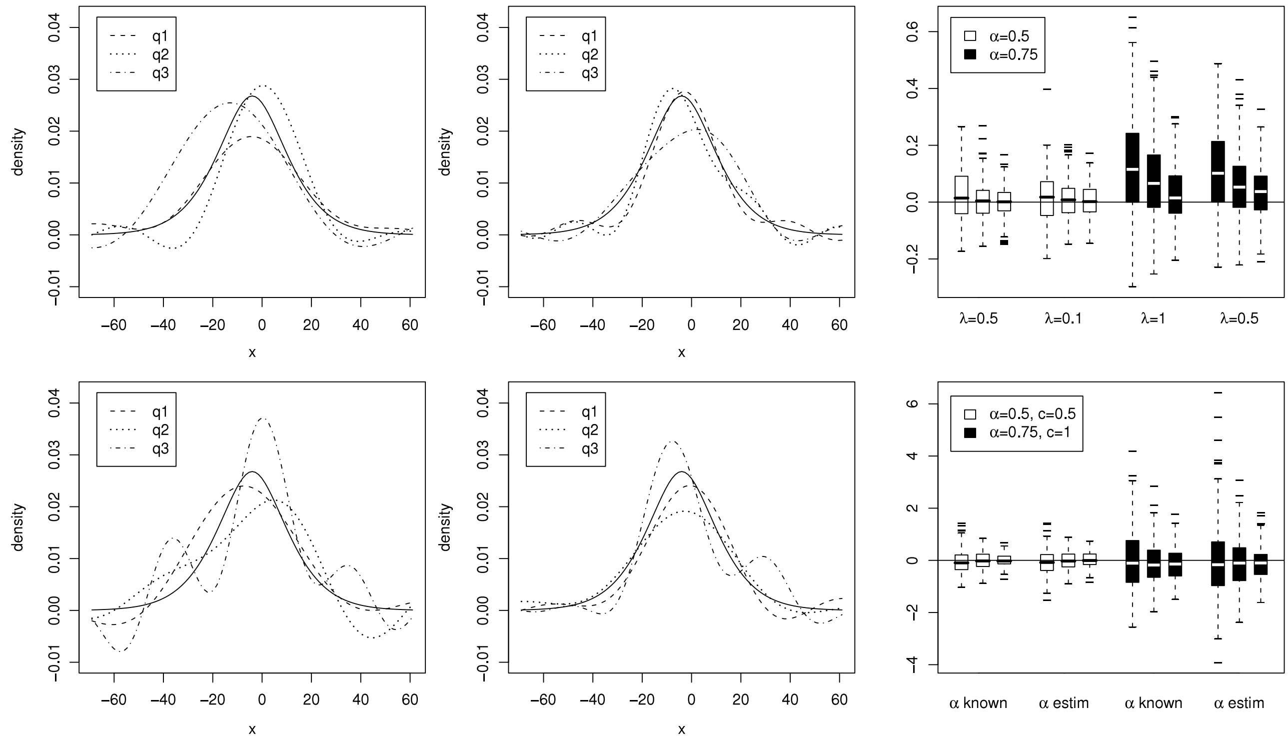

Fig. 3. Results for density (iii) with $(\alpha, \mu_{\text{Lap}}) = (0.5, 4)$ or $(0.75, 15)$. First ($q1$), second ($q2$) and third ($q3$) quartile curves when $(\alpha, \lambda, c) = (0.5, 0.5, 1.5)$, for $n = 500$ (left) or $n = 2000$ (middle), when α is known (top) or estimated (bottom). The solid lines represent the target curve; and boxplots of $\hat{\theta} - \theta$ for $\theta = \alpha$ (top right) or $\theta = c$ when α is known or estimated (bottom right). In each case, $\lambda = 0.1$ for boxes 1–3, 7–9 and $\lambda = 0.5$ otherwise; $n = 500$ for boxes 1, 4, 7, 10, $n = 1000$ for boxes 2, 5, 8, 11, $n = 2000$ otherwise.

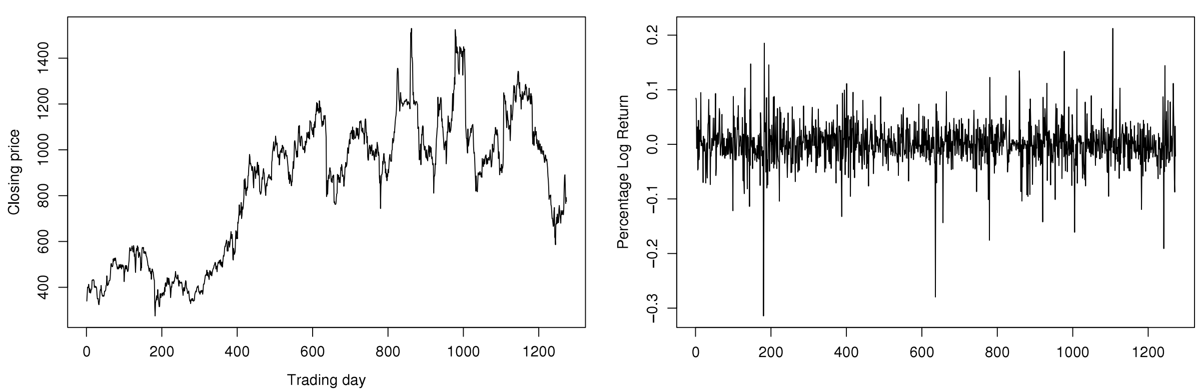

Fig. 4. Daily closing price for PTEK (left) and percentage daily log returns (right) for each trading day between 2002 and 2007.

curve was approximately horizontal, whereas in the second we looked for values α such that there was an approximately linear relationship between $|t|^\alpha$ and $\Re\hat{\psi}(t)$ (except near zero or too far out in the tails). Both approaches indicated that the value of α should be large, between 1.8 and 2. The graphs of $|t|^{1.8}$ and $|t|^2$ versus $\Re\hat{\psi}(t)$ are shown in Fig. 5, together with $-\hat{c}|t|^\alpha - \hat{\lambda}$, for $t \in [0, 120]$, with $\alpha = 1.8$ and 2.0 respectively. We see that away from zero and before the graph of $\Re\hat{\psi}(t)$ curves upward (apparently reflecting degradation due to noise), the relationship is approximately linear. We also tried the method of Belomestny (in press), which suggested $\alpha \approx 1.6$. However, we found that the corresponding estimates of $\chi(t)$ when $\alpha < 1.8$ did not have the properties one would expect of characteristic functions (e.g. $\chi(t)$ decreasing on $[0, \epsilon]$ for some $\epsilon > 0$). Therefore, we tend to favour the suggestion that $1.8 \le \alpha \le 2$.

We then ran our regression procedure with $\alpha = 1.8$ and 2.0 respectively. For $\alpha = 1.8$ we found that $\hat{c} = 6.22 \times 10^{-4}$ and $\hat{\lambda} = 0.260$; for $\alpha = 2$ we found that $\hat{c} = 2.34 \times 10^{-4}$ and $\hat{\lambda} = 0.412$. The estimate of b was $\hat{b} = 7.23 \times 10^{-4}$, which is not affected by the value of the index parameter α since $\hat{b}$ is calculated using the first equation in (2.9) and that equation does not involve α. Note too that when the underlying Lévy process is viewed as having a diffusion part ($\alpha = 2.0$), the estimated compound Poisson rate λ is much larger than its counterpart when the diffusion is replaced by a stable process ($\alpha = 1.8$). This reflects the fact that the stable process prescribes an infinite activity jump process which absorbs some smaller size jumps and hence reduces the rate of the larger jumps that come from the compound Poisson process.

The estimated jump size density, $\hat{f}$, of the Poisson process is shown in Fig. 6 for $\alpha = 1.8$ and $\alpha = 2$, respectively. In both cases we see that the estimated density had its mode very near zero, and was slightly skewed to the left, due to the two large negative jumps occurring near time points 200 and 600.

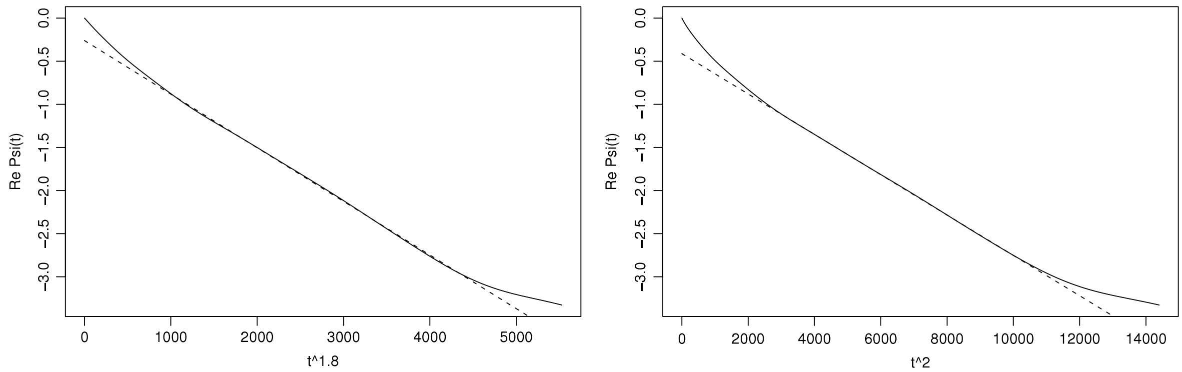

Fig. 5. Graph of $|t|^\alpha$ versus $\Re\hat{\psi}(t)$ when $\alpha = 1.8$ (left) or $\alpha = 2$ (right). In both cases, the continuous line shows $\Re\hat{\psi}(t)$ and the dashed line shows the fitted value of $-c|t|^\alpha - \lambda$.

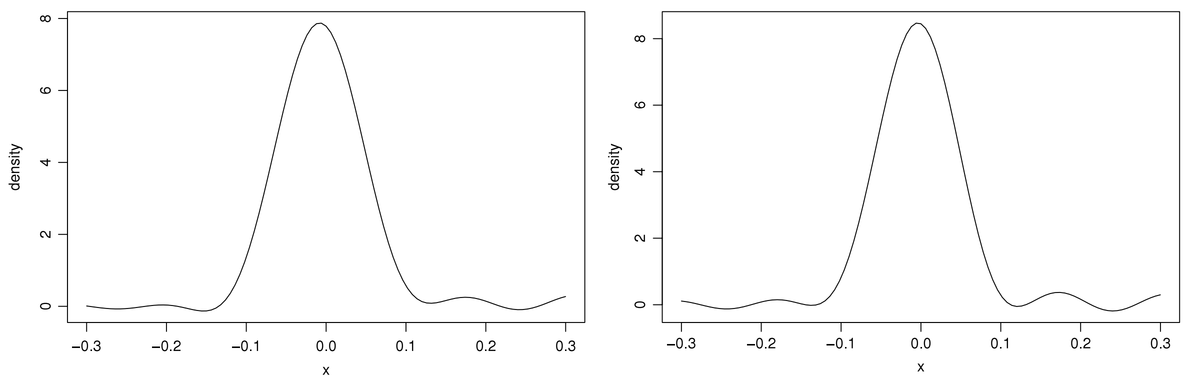

Fig. 6. Estimator of the density f of the Poisson jump when $\alpha = 1.8$ (left) or $\alpha = 2$ (right).

Acknowledgements

We are grateful to two referees, the Associate Editor and the Editor for constructive comments and suggestions that have improved the presentation of the paper. The work of Aurore Delaigle and Peter Hall was partially supported by a grant from the Australian Research Council.

Appendix. Proofs

A.1. Proof of Theorem 3.1

For brevity we treat only the case where λ_1, in (2.20), is nonzero. The case $\lambda_1 = 0$ is similar. To capture the generality of the model at (2.20), as opposed to that at (2.1), we introduce a function ω defined by

$$\begin{aligned} \psi(t) &= \log\phi(t) \\ &= \mathrm{i}t\, b - |t|^\alpha c + \lambda\{\chi(t) - 1\} + \lambda_1\{\omega(t) - 1\}, \end{aligned} \tag{A.1}$$

where $\lambda_1\{\omega(t) - 1\} = O(|t|^\eta)$ as $t \to 0$ and $\lambda_1\omega(t) = O\{g(t)\}$ as $|t| \to \infty$. In this notation, (2.7) and (2.8) alter to:

$$\begin{aligned} &\Im\,\hat{\psi}(t) = tb + \epsilon_1(t), \qquad \Re\,\hat{\psi}(t) = -(\lambda + \lambda_1) - |t|^\alpha c + \epsilon_2(t), \\ &\epsilon_1(t) = \Im\,\{\lambda\chi(t) + \lambda_1\omega(t)\} + \Im\,\epsilon(t), \\ &\epsilon_2(t) = \Re\,\{\lambda\chi(t) + \lambda_1\omega(t)\} + \Re\,\epsilon(t), \end{aligned} \tag{A.2}$$

respectively. As before we put $\epsilon(t) = \hat{\psi}(t) - \psi(t)$, where $\hat{\psi}$ is defined as at (2.4), but now $\psi = \log\phi$ with ϕ defined at (A.1) rather than (2.1).

With ϵ_1 and ϵ_2 as at (A.2), and for integers j, put $I_j = \int_{\mathcal{T}} t^{j\alpha}\mathrm{d}t$ and $J_j = \int_{\mathcal{T}} t^{j\alpha}\epsilon_2(t)\mathrm{d}t$, and note that in view of (3.3)(a) the minimisers $\hat{b}$, $\hat{c}$ and $\hat{\lambda}$ of the integrals at (2.9) are given by:

$$\hat{b} - b = \left(\int_{\mathcal{T}} t^2\mathrm{d}t\right)^{-1}\int_{\mathcal{T}} t\epsilon_1(t)\mathrm{d}t, \tag{A.3}$$

$$\begin{aligned} \begin{pmatrix} \hat{\lambda} - (\lambda + \lambda_1) \\ \hat{c} - c \end{pmatrix} &= -\begin{pmatrix} I_0 & I_1 \\ I_1 & I_2 \end{pmatrix}^{-1}\begin{pmatrix} J_0 \\ J_1 \end{pmatrix} \\ &= \left(I_0I_2 - I_1^2\right)^{-1}\begin{pmatrix} I_1J_1 - I_2J_0 \\ I_1J_0 - I_0J_1 \end{pmatrix}. \end{aligned} \tag{A.4}$$

Let $\Delta = (\hat{\phi} - \phi)/\phi$ and, given positive functions A and B of s_1 and s_2, write $A \asymp B$ to denote that A/B is bounded away from zero and infinity as s_1 and s_2 increase. In this notation,

$$\int_{\mathcal{T}} t^2\mathrm{d}t \asymp s_2^2 s_3, \qquad I_j \asymp s_2^{j\alpha}s_3, \qquad I_0I_2 - I_1^2 \asymp s_2^{2\alpha-2}s_3^4. \tag{A.5}$$

Moreover, since

$$\begin{aligned} \epsilon_1(t) &= \lambda\Im\,\chi(t) + \lambda_1\Im\,\omega(t) + \Im\,\epsilon(t) \\ &= \lambda\Im\,\chi(t) + \lambda_1\Im\,\omega(t) + \Im\,\{\hat{\psi}(t) - \psi(t)\} \\ &= \lambda\Im\,\chi(t) + \lambda_1\Im\,\omega(t) + \Im\,[\log\{1 + \Delta(t)\}], \end{aligned}$$

and the same equality holds for ϵ_2 when we replace $\Im$ by $\Re$, we have

$$\begin{aligned} &\int_{\mathcal{T}} t|\epsilon_1(t)|\mathrm{d}t \le \lambda_2\int_{\mathcal{T}} t\chi_1(t)\mathrm{d}t + R_1, \\ &|J_j| \le \lambda_2\int_{\mathcal{T}} t^{j\alpha}\chi_1(t)\mathrm{d}t + R_2(j), \end{aligned} \tag{A.6}$$

where $\lambda_2 = 4\max(\lambda, |\lambda_1|)$, $\chi_1 = \max(|\chi|, |\omega|)$,

$$R_1 = \int_{\mathcal{T}} t|\log\{1 + \Delta(t)\}|\mathrm{d}t,$$

$$R_2(j) = \int_{\mathcal{T}} t^{j\alpha}|\log\{1 + \Delta(t)\}|\mathrm{d}t.$$

At this point it is convenient to re-define δ_{n1}, at (3.1), by introducing a constant multiplier $C > 0$ (this does not have any impact on the validity of the theorem):

$$\delta_{n1} = C\left(n^{-1}\log n\right)^{1/2}\exp\left(cs_2^{\alpha}\right). \tag{A.7}$$

Shortly we shall prove that, if $C > 0$ is chosen sufficiently large,

$$P\left\{\sup_{t\in\mathcal{T}}|\Delta(t)| > \delta_{n1}\right\} \to 0. \tag{A.8}$$

If $|\Delta(t)| \le \delta_{n1} \le \frac{1}{2}$ for all $t \in \mathcal{T}$ then $|\log\{1+\Delta(t)\}| \le 2|\Delta(t)|$, and so $R_1 \le 2s_2\int_{\mathcal{T}}|\Delta(t)|\mathrm{d}t$. This result, and its analogue for $R_2(j)$, imply that, provided (A.8) holds,

$$R_1 = O_p\left\{s_2\int_{\mathcal{T}}E|\Delta(t)|\mathrm{d}t\right\}, \qquad R_2(j) = O_p\left\{s_2^{j\alpha}\int_{\mathcal{T}}E|\Delta(t)|\mathrm{d}t\right\}. \tag{A.9}$$

Also, $E|\hat{\phi}-\phi|^2 \le 2n^{-1}$, and so,

$$E|\Delta(t)| = |\phi(t)|^{-1}E\left|\hat{\phi}(t)-\phi(t)\right| \le 2\exp\left(c|t|^{\alpha}+\lambda_2\right)n^{-1/2}. \tag{A.10}$$

Combining we deduce that

$$R_1 = O_p\left\{s_2 n^{-1/2}\int_{\mathcal{T}}\exp\left(c|t|^{\alpha}\right)\mathrm{d}t\right\}, \quad R_2(j) = O_p\left\{s_2^{j\alpha}n^{-1/2}\int_{\mathcal{T}}\exp\left(c|t|^{\alpha}\right)\mathrm{d}t\right\}. \tag{A.11}$$

Together, (A.6) and (A.11) give:

$$\int_{\mathcal{T}}t|\epsilon_1(t)|\mathrm{d}t = O_p\left\{s_2\int_{\mathcal{T}}|\chi_1(t)|\mathrm{d}t + s_2 n^{-1/2}\int_{\mathcal{T}}\exp\left(c|t|^{\alpha}\right)\mathrm{d}t\right\} = O_p(s_2\delta_{n2}), \tag{A.12}$$

$$|J_j| = O_p\left\{s_2^{j\alpha}\int_{\mathcal{T}}|\chi_1(t)|\mathrm{d}t + s_2^{j\alpha}n^{-1/2}\int_{\mathcal{T}}\exp\left(c|t|^{\alpha}\right)\mathrm{d}t\right\} = O_p\left(s_2^{j\alpha}\delta_{n2}\right). \tag{A.13}$$

Result (3.4) follows from (A.3)–(A.5), (A.12) and (A.13). When applying (A.4) we bound the term $I_iJ_j - I_kJ_\ell$ by simply $|I_i||J_j| + |I_k||J_\ell|$.

It remains to establish (A.8). Note that, by (3.3)(c), $\delta_{n1} \to 0$ as $n \to \infty$. We claim that, for all sufficiently large n,

$$\sup_{t\in\mathcal{T}}P\left\{|\Delta(t)| > \frac{1}{2}\delta_{n1}\right\} \le n^{-C_0C^2}, \tag{A.14}$$

where C is as in (A.7) and $C_0 > 0$ is an absolute constant. Indeed, with δ_{n1} given by (A.7), and for $t \in \mathcal{T}$, we have for all $n \ge n_0$, where n_0 does not depend on t,

$$P\{|\Delta(t)| > \frac{1}{2}\delta_{n1}\} \le P\left\{\left|\hat{\phi}(t)-\phi(t)\right| > \frac{1}{2}C\left(n^{-1}\log n\right)^{1/2}\right\} \le \exp\left(-C_0C^2\log n\right),$$

where the last inequality follows from Bernstein's inequality.

To complete the proof of (A.8), note that a random variable U with characteristic function $\exp(\mathrm{i}tb - |t|^{\alpha}c)$ has $E|U|^{\eta_1} < \infty$ for each $\eta_1 \in (0,\alpha)$. Observe too that, in view of the implication from (3.3)(b) that $E|Z|^{\eta_2} < \infty$ for some $\eta_2 \in (0, 1]$, the characteristic function χ of Z satisfies $\chi(t) - 1 = O(t^{\eta_2})$ as $t \to 0$. Therefore, $\exp[\lambda\{\chi(t) - 1\}] - 1 = O(t^{\eta_2})$ as $t \to 0$. Hence, a random variable V with characteristic function $\exp[\lambda\{\chi(t) - 1\}]$ satisfies $E|V|^{\eta_3} < \infty$ for each $\eta_3 \in (0, \eta_2)$. Taking $\eta_4 \in (0, \min(\eta_1, \eta_3, \eta))$, where η is as at (2.20), we deduce that a random variable X, with the characteristic function at (2.20), satisfies $E|X|^{\eta_4} < \infty$. For this choice of η_4, and for any $t, t' \in \mathcal{T}$, we have, using the fact that $|\cos x - 1| \le x^2$ for all x:

$$\begin{aligned}\left|\hat{\phi}(t)-\hat{\phi}(t')\right| &\le \frac{1}{n}\sum_{j=1}^{n}|\exp\{\mathrm{i}(t-t')X_j\}-1| \\ &\le \frac{2}{n}\sum_{j=1}^{n}\min\{|(t-t')X_j|, 1\} \\ &\le \frac{2}{n}\sum_{j=1}^{n}\min\left\{|(t-t')X_j|^{\eta_4}, 1\right\} \\ &\le |t-t'|^{\eta_4}\frac{2}{n}\sum_{j=1}^{n}|X_j|^{\eta_4}.\end{aligned}$$

(Here we have employed the fact that $0 < \eta_4 < 1$.) Therefore, using Markov's inequality we deduce that, for $\xi_1, \xi_2 > 0$,

$$P\left\{\sup_{t,t'\in\mathbb{R}:|t-t'|\le n^{-\xi_1}}\left|\hat{\phi}(t)-\hat{\phi}(t')\right| > n^{-\xi_2}\right\} \le P\left(\frac{2}{n^{1+\eta_4\xi_1-\xi_2}}\sum_{j=1}^{n}|X_j|^{\eta_4} > 1\right) \le 2n^{\xi_2-\eta_4\xi_1}E|X|^{\eta_4}. \tag{A.15}$$

More simply, $|\phi(t) - \phi(t')| \le \text{const.}\,|t-t'|^{\eta_4}$. Therefore, if $|t-t'| \le n^{-\xi_1}$ and $|\hat{\phi}(t)-\hat{\phi}(t')| \le n^{-\xi_2}$ then

$$\begin{aligned}|\Delta(t)-\Delta(t')| &= \left|\frac{\hat{\phi}(t)-\hat{\phi}(t')}{\phi(t)} + \frac{\hat{\phi}(t')}{\phi(t')}\left\{\frac{\phi(t')}{\phi(t)}-1\right\}\right| \\ &\le |\phi(t)|^{-1}n^{-\xi_2} + |\phi(t')|^{-2}n^{-\eta_4\xi_1} \\ &\le \text{const.}\exp\left(2cs_2^{\alpha}\right)n^{-\min(\eta_4\xi_1,\xi_2)} \\ &\le \text{const.}\,n^{1-\min(\eta_4\xi_1,\xi_2)},\end{aligned} \tag{A.16}$$

where only the before-last inequality requires $t, t' \in \mathcal{T}$ and we have used the fact that, in this case, δ_{n1} at (2.17) satisfies $\delta_{n1} \le 1$ for all sufficiently large n; see (3.3)(c).

From (A.15) and (A.16) we deduce that, for a constant $C_1 > 0$ depending on ξ_1 and ξ_2,

$$P\left\{\sup_{t,t'\in\mathcal{T}:|t-t'|\le n^{-\xi_1}}|\Delta(t)-\Delta(t')| > C_1n^{1-\min(\eta_4\xi_1,\xi_2)}\right\} \le 2n^{\xi_2-\eta_4\xi_1}E|X|^{\eta_4}. \tag{A.17}$$

Since ξ_1 and ξ_2 are arbitrarily large then, provided $\xi_1 > 0$ is chosen sufficiently large, (A.17) implies that

$$P\left\{\sup_{t,t'\in\mathcal{T}:|t-t'|\le n^{-\xi_1}}|\Delta(t)-\Delta(t')| > n^{-1}\right\} \to 0. \tag{A.18}$$

Let $\mathcal{T}_n$ denote a partition of $\mathcal{T}$ into $m(n) = O(s_3n^{\xi_1})$ equally-spaced points in $\mathcal{T}$, with adjacent points distant $n^{-\xi_1}$ apart. For $t \in \mathcal{T}$, let $t' \in \mathcal{T}$ such that $|t-t'| \le n^{-\xi_1}$. Using (A.14) and the fact that $n^{-1} = o(\delta_{n1})$, we have

$$\begin{aligned}P\left\{\sup_{t\in\mathcal{T}}|\Delta(t)| > \delta_{n1}\right\} &\le P\left\{\sup_{t'\in\mathcal{T}_n}|\Delta(t')| > \frac{1}{2}\delta_{n1}\right\} + P\left\{\sup_{t\in\mathcal{T},t'\in\mathcal{T}_n}|\Delta(t)-\Delta(t')| > \frac{1}{2}\delta_{n1}\right\} \\ &\le \#\mathcal{T}_n\sup_{t\in\mathcal{T}}P\left\{|\Delta(t)| > \frac{1}{2}\delta_{n1}\right\} + P\left\{\sup_{t,t'\in\mathcal{T}:|t-t'|\le n^{-\xi_1}}|\Delta(t)-\Delta(t')| > \frac{1}{2}\delta_{n1}\right\} \\ &\le m(n)n^{-C_0C^2} + P\left\{\sup_{t,t'\in\mathcal{T}:|t-t'|\le n^{-\xi_1}}|\Delta(t)-\Delta(t')| > n^{-1}\right\},\end{aligned}$$

which converges to zero by (A.18), as long as C is chosen sufficiently large. This establishes (A.8).

A.2. *Proof of* Theorem 3.2

Recall that $\hat{\chi}$ and $\widehat{F}$ are given by (2.10) and (2.13), respectively, and that the role played by t_1 is described by (2.10). Let G be the function with Fourier–Stieltjes transform equal to $\chi(t)-1$, for $|t|\le t_1$, and equal to zero for $|t|>t_1$. Then, using (2.5) and (2.6) to expand $\hat{\psi}$ in (2.10), we obtain:

$$\begin{aligned}A &\equiv \sup_{-\infty<x<\infty}\left|\widehat{F}(x)-F(x)+\hat{\lambda}^{-1}(\hat{\lambda}-\lambda)G(x)\right|\\ &\le \frac{1}{2\pi}\int_{-t_1}^{t_1}|t|^{-1}\left|\hat{\chi}(t)-\chi(t)+\hat{\lambda}^{-1}(\hat{\lambda}-\lambda)\{\chi(t)-1\}\right|\mathrm{d}t\\ &\quad+\frac{1}{\pi}\int_{t_1}^{\infty}|t|^{-1}|\chi(t)|\mathrm{d}t\\ &=\frac{1}{2\pi\hat{\lambda}}\int_{-t_1}^{t_1}|t|^{-1}\left|it(b-\hat{b})+|t|^{\alpha}(\hat{c}-c)+\hat{\psi}(t)-\psi(t)\right|\mathrm{d}t\\ &\quad+\frac{1}{\pi}\int_{t_1}^{\infty}|t|^{-1}|\chi(t)|\mathrm{d}t.\end{aligned}$$

Choosing $t_1=(c_0\log n)^{1/\alpha}$, where $c_0>0$; noting that χ satisfies (3.5), which implies that $\int_{t>t_1}t^{-1}|\chi(t)|\mathrm{d}t=O(t_1^{-\beta})$; and taking $\hat{c}$, $\hat{b}$ and $\hat{\lambda}$ to be estimators that satisfy (3.7); we obtain:

$$\begin{aligned}A &= O_p\left\{t_1|\hat{b}-b|+t_1^{\alpha}|\hat{c}-c|+t_1^{-\beta}\right.\\ &\quad\left.+\int_0^{t_1}|t|^{-1}\left|\hat{\psi}(t)-\psi(t)\right|\mathrm{d}t\right\}\\ &=O_p\left(t_1^{-\beta}+I^{(1)}\right),\end{aligned}\tag{A.19}$$

where, for $j=0,1$,

$$I^{(j)}=\int_0^{t_1}|t|^{-j}\left|\hat{\psi}(t)-\psi(t)\right|\mathrm{d}t.$$

The function whose Fourier–Stieltjes transform equals $I(1\le|t|\le t_1)$ is $\kappa(t_1x)-\kappa(x)$, where $\kappa(x)=\pi^{-1}\int_{v\le x}v^{-1}\sin v\mathrm{d}v$. Therefore, the function whose Fourier–Stieltjes transform equals $\{\chi(t)-1\}I(|t|\le t_1)+I(1\le|t|\le t_1)$ is $G(x)+\kappa(t_1x)-\kappa(x)$. Hence,

$$\begin{aligned}&\sup_{-\infty<x<\infty}|G(x)+\kappa(t_1x)-\kappa(x)|\\ &\le\frac{1}{2\pi}\int_{-t_1}^{t_1}|t|^{-1}|\chi(t)-I(|t|\le 1)|\mathrm{d}t\\ &\le\frac{1}{2\pi}\int_{-\infty}^{\infty}|t|^{-1}|\chi(t)-I(|t|\le 1)|\mathrm{d}t\\ &<\infty.\end{aligned}$$

It follows that $|G(x)|$ is bounded uniformly in x and n, and so

$$\sup_{-\infty<x<\infty}\left|\hat{\lambda}^{-1}(\hat{\lambda}-\lambda)G(x)\right|=O_p\left(\left|\hat{\lambda}-\lambda\right|\right)=O_p\left(t_1^{-\beta}\right),\tag{A.20}$$

where the last identity follows from (3.7) with $\theta=\lambda$. Together, (A.19) and (A.20) imply that

$$B\equiv\sup_{-\infty<x<\infty}\left|\widehat{F}(x)-F(x)\right|=O_p\left(t_1^{-\beta}+I^{(1)}\right).$$

And arguments in Appendix A.1 can be used to prove that for all $\eta>0$, $I^{(1)}=O_p(n^{\eta+cc_0-(1/2)})$. Therefore, if c_0, in the definition of t_1 above, satisfies $c\,c_0<\frac{1}{2}$, then $I^{(1)}$ is negligibly small relative to $t_1^{-\beta}$. Therefore $B=O_p(t_1^{-\beta})$, which implies the first part of (3.10). The second part can be derived similarly.

To prove (3.11) we note that, analogously to the bounds for A,

$$\begin{aligned}&\sup_{-\infty<x<\infty}\left|\hat{f}(x)-f(x)\right|\\ &\le\frac{1}{2\pi}\int_{-t_1}^{t_1}\left|\hat{\chi}(t)-\chi(t)\right|\mathrm{d}t+\frac{1}{\pi}\int_{t_1}^{\infty}|\chi(t)|\mathrm{d}t\\ &=\frac{1}{2\pi\hat{\lambda}}\int_{-t_1}^{t_1}\left|it(b-\hat{b})+|t|^{\alpha}(\hat{c}-c)\right.\\ &\quad\left.+(\hat{\lambda}-\lambda)\{\chi(t)-1\}+\hat{\psi}(t)-\psi(t)\right|\mathrm{d}t+\frac{1}{\pi}\int_{t_1}^{\infty}|\chi(t)|\mathrm{d}t\\ &=O_p\left\{t_1^2|\hat{b}-b|+t_1^{\alpha+1}|\hat{c}-c|+t_1\left|\hat{\lambda}-\lambda\right|\right.\\ &\quad\left.+\int_0^{t_1}\left|\hat{\psi}(t)-\psi(t)\right|\mathrm{d}t\right\}\\ &=O_p\left(t_1^{1-\beta}+I^{(0)}\right)=O_p\left(t_1^{1-\beta}\right),\end{aligned}$$

where the second-last identity follows from (3.7) and the final identity holds provided $c\,c_0<\frac{1}{2}$. (The latter condition ensures that $I^{(0)}=O_p(n^{-\eta})$ for some $\eta>0$.) This implies (3.11).

A.3. *Proof of* Theorem 3.4

For brevity we treat only case (i). We take the characteristic function, $\gamma(t)=\exp(it\,b-|t|^{\alpha}c)$ in (3.14), to be fixed, and consider two different characteristic functions, χ_1 and χ_2, for Z. These give rise to respective versions, $\tau_j=\exp\{\lambda(\chi_j-1)\}$ for $j=1$ and 2, of the characteristic function τ in (3.14). Let F_j be the distribution function, and f_j the density, of the distribution with characteristic function χ_j, with f_1 and f_2 equal to, respectively, f_0 and f_n of Fan (1991), page 1269, with δ^k there equal to $(\log n)^{-(\beta-1)/\alpha}$. Write G_j for the distribution function, and g_j for the density, of the distribution with characteristic function $\phi_j=\gamma\tau_j$. The data come from either G_1 or G_2. Similarly as in the conventional-deconvolution dual of this problem treated by Fan (1991), the main technical step in the proof is showing that

$$\int\frac{(g_1-g_2)^2}{g_1}\le\text{const.}\,n^{-1}.\tag{A.21}$$

To do this, take H_j to be the distribution (for which the density is h_j) with characteristic function $\xi_j=\phi_1\chi_j$, for $j=1$ and 2. Note that ϕ_1 corresponds to a supersmooth error in the terminology of Fan (1991), and thus data from H_j are data from the conventional supersmooth deconvolution problem treated by Fan (1991) (there, h_j is denoted by f_{Y_j}). Thus from Fan (1991) we have: $\int(h_1-h_2)^2\le$ const. n^{-c_1} for some $c_1>1$. By Parseval's identity, we also have

$$\begin{aligned}\int(h_1-h_2)^2&=\frac{1}{2\pi}\int|\xi_1-\xi_2|^2=\frac{1}{2\pi}\int|\phi_1|^2|\chi_1-\chi_2|^2\\ &=\frac{1}{2\pi\lambda^2}\int|\phi_1|^2|\lambda(\chi_1-\chi_2)|^2\\ &\sim\frac{1}{2\pi\lambda^2}\int|\phi_1|^2|1-\exp\{\lambda(\chi_2-\chi_1)\}|^2\\ &\asymp\frac{1}{2\pi\lambda^2}\int|\phi_1-\phi_2|^2=\lambda^{-2}\int(g_1-g_2)^2.\end{aligned}\tag{A.22}$$

Result (A.22) implies that $\int(g_1-g_2)^2\le$ const. n^{-c_1}, and, using arguments borrowed from Fan (1991) (see particularly lines 13–16 and 19–21 of Fan's page 1270) and Butucea and Matias (2005), it follows that (A.21) holds (see also Appendix A.4 for more details in a similar but more complicated context). This implies that the

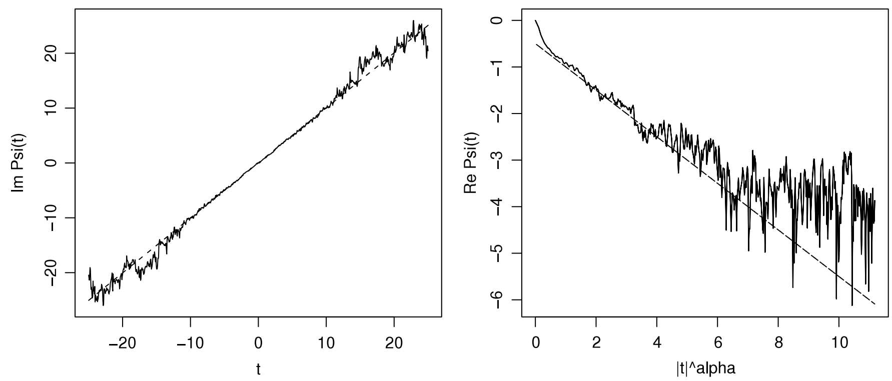

Fig. 7. Results for a typical sample of size $n = 1000$ from case (iii) when $n = 1000$, $\lambda = 0.5$, $c = 0.5$ and $\alpha = 0.5$. Left: graph of $(t, \Im\hat{\psi}(t))$ (continuous line) and $(t, b \cdot t)$ (dashed line). Right: graph of $(|t|^\alpha, \Re\hat{\psi}(t))$ (continuous line) and $(|t|^\alpha, -\lambda - c|t|^\alpha)$ (dashed line).

information in a dataset of size n from G_1 or G_2 is marginal for discrimination, which implies (3.15) and (3.16).

A.4. Proof of Theorem 3.5

Butucea and Matias (2005, Lemma 3) established the existence of a function q, with continuous Fourier transform q^{Ft}, such that: (a) q is an even function, (b) q^{Ft} vanishes outside $[-2, 2]$, (c) $q^{\mathrm{Ft}}(t) = 1$ for all $t \in [-1, 1]$, and (d) q^{Ft} has four continuous derivatives on the real line. Let $\eta_1, \eta_2 > 0$ be fixed; put $\delta = \eta_1(\log n)^{-1/\alpha}$; let $\alpha \in (0, 2]$ and $\beta > 0$ be as in (2.1) and (3.5), respectively; let c_0, b_0 and λ_0 be interior points of $\mathcal{I}_+$, $\mathcal{I}$ and $\mathcal{I}_+$ respectively; let k_b, k_c and k_λ be fixed constants; and put

$$\begin{aligned}\tau(t) &= \frac{\delta^\beta}{\lambda_0}\Big[\mathrm{i}t\,\delta k_b - |t\delta|^\alpha k_c + k_\lambda\{\chi_0(t) - 1\}\Big]q^{\mathrm{Ft}}(\eta_2\delta t),\\ \chi_\star(t) &= \chi_0(t) + \tau(t), \\ \phi_\star(t) &= \exp\Big[\mathrm{i}tb_0 - |t|^\alpha c_0 + \lambda_0\{\chi_\star(t) - 1\}\Big],\end{aligned} \tag{A.23}$$

where χ_0, satisfies (3.5) with C_1 there replaced by $\frac{1}{2}C_1$, and is a characteristic function such that the distribution with characteristic function

$$\phi_0(t) = \exp\Big[\mathrm{i}t\, b_0 - |t|^\alpha c_0 + \lambda_0\{\chi_0(t) - 1\}\Big]$$

corresponds to a density g_1 for which

$$g_1(x) > \text{const.}\,(1 + |x|)^{-(1+\zeta)} \tag{A.24}$$

for some $\zeta \in (0, 1)$. (We may choose C_1, in the definition of the function class $\mathcal{C}$ in (3.19), arbitrarily large.) Property (A.24) can be ensured by taking the distribution with characteristic function χ_0 to be mixed with a smooth distribution that has sufficiently heavy tails.

Define $b_\delta = b_0 + k_b\delta^{1+\beta}$, $c_\delta = c_0 + k_c\delta^{\alpha+\beta}$ and $\lambda_\delta = \lambda_0 + k_\lambda\delta^\beta$, and let ϕ_δ be the version of ϕ when $b = b_\delta$, $c = c_\delta$, $\lambda = \lambda_\delta$ and $\chi = \chi_0$:

$$\begin{aligned}\phi_\delta(t) &= \exp\Big[\mathrm{i}t\, b_\delta - |t|^\alpha c_\delta + \lambda_\delta\{\chi_0(t) - 1\}\Big]\\ &= \phi_0(t)\exp\{\xi(t)\},\end{aligned} \tag{A.25}$$

where both identities hold for all t, and

$$\xi(t) = \mathrm{i}t\, k_b\delta^{1+\beta} - |t|^\alpha k_c\delta^{\alpha+\beta} + k_\lambda\delta^\beta\{\chi_0(t) - 1\}.$$

In view of (A.23), (A.25) and property (c) noted in the previous paragraph,

$$\begin{aligned}&\phi_\delta(t) = \phi_\star(t) \quad \text{for } |t| \le (\eta_2\delta)^{-1},\\ &\phi_\delta(t) - \phi_\star(t) = \phi_0(t)\Big[\exp\{\xi(t)\} - \exp\big\{\xi(t)q^{\mathrm{Ft}}(\eta_2\delta t)\big\}\Big]\end{aligned} \tag{A.26}$$

for all t.

The second relation in (A.26) implies that the densities g_δ and $g_\star$ corresponding to characteristic functions ϕ_δ and $\phi_\star$, respectively, satisfy $g_\delta(x) - g_\star(x) = \delta^{-c_1}h_n(x)$, for some real c_1, where, in view of properties (a)–(d) in the first paragraph of this proof, the function h_n satisfies $\sup_{n,x}(1 + |x|^r)|h_n(x)| < \infty$ for some $r > 1$. Compare, for example, the first displayed formula on p. 1269 of Fan (1991).

Properties (a), (b) and (d), and (A.24), together with arguments of Butucea and Matias (2005), can be used to show that for appropriate choices of η_1 and η_2, which allow $\eta_1\eta_2$ to be arbitrarily small; and for all sufficiently small $\delta > 0$, i.e. for all sufficiently large n; $\chi_\star$ is a characteristic function and satisfies (3.5), and so $\phi_\star$ is the characteristic function of a distribution in class $\mathcal{C}$. The argument uses the fact that $|\tau(t)| \le \text{const.}\,\delta^\beta \le \text{const.}\,(1+|t|)^{-\beta}$ for all t, and $\tau(t) = 0$ for $|t| > 2(\eta_2\delta)^{-1}$.

For all sufficiently large n, the distribution with characteristic function $\phi_\star$ is absolutely continuous. Parseval's identity, (A.26), and properties (b) and (c) in the first paragraph of this proof, imply that

$$\begin{aligned}2\pi\int(g_\delta - g_\star)^2 &= \int_{1/(\eta_2\delta)\le|t|\le 2/(\eta_2\delta)}|\phi_\delta(t) - \phi_\star(t)|^2\mathrm{d}t\\ &\sim \int_{1/(\eta_2\delta)\le|t|\le 2/(\eta_2\delta)}\exp\big(-2|t|^\alpha c_0 - 2\lambda_0\big)|\xi(t)|^2\\ &\quad\times\Big|1 - q^{\mathrm{Ft}}(\eta_2\delta t)\Big|^2\,\mathrm{d}t.\end{aligned}$$

Therefore, provided $(\eta_1\eta_2)^{-\alpha} > \frac{1}{2}c_0^{-1}c_2$, where $c_2 > 1$, we have $\int(g_\delta - g_\star)^2 = O(n^{-c_2})$. Let $M_n = n^{c_3}$ with $c_3 > 0$. Then since $g_\delta > \text{const.}\,(1 + |x|)^{-(1+\zeta)}$, we have

$$\begin{aligned}\int\frac{(g_\delta - g_\star)^2}{g_\delta} &= \int_{-M_n}^{M_n}\frac{(g_\delta - g_\star)^2}{g_\delta}\\ &\quad + \int_{-\infty}^{-M_n}\frac{(g_\delta - g_\star)^2}{g_\delta} + \int_{M_n}^{\infty}\frac{(g_\delta - g_\star)^2}{g_\delta}\\ &\le \text{const.}\,M_n^{1+\zeta}\int_{-M_n}^{M_n}(g_\delta - g_\star)^2\\ &\quad + \text{const.}\,\delta^{-2c_1}\int_{M_n}^{\infty}x^{1+\zeta-2r}\\ &\le \text{const.}\,M_n^{1+\zeta}n^{-c_2} + \text{const.}\,\delta^{-2c_1}M_n^{2+\zeta-2r}\\ &= O(n^{-1})\end{aligned}$$

if $c_3 \leq (1 + c_2)/(1 + \zeta)$ and $c_3(2r - 2 - \zeta) > 1$, which is satisfied for c_2 large enough as long as we take χ_0 so that $0 < \zeta < 2r - 2$.

A.5. Illustration of the visual procedure of Section 4

Fig. 7(left) shows the graphs of $(t, \Im\hat{\psi}(t))$ and $(t, b \cdot t)$ for a typical sample of size $n = 1000$ from case (ii) with $\lambda = 0.5$, $c = 0.5$ and $\alpha = 0.5$. It appears clearly that there is a wide range of intervals for $|t|$, all included in about $[1, 10]$, where the graph is clearly linear. Each interval gives roughly the same fitted $\hat{b}$, which is close to b. Fig. 7(right) shows the graphs of $(|t|^{\alpha}, \Re\hat{\psi}(t))$ and $(|t|^{\alpha}, -\lambda - c|t|^{\alpha})$. There, too, it is not hard to find an interval close to zero where the relation looks linear and non-erratic, and the result is robust against the choice of the interval. This flexibility is reflected by the good results of the automatic method we used in the simulations, which of course did usually not select the best possible interval for a given sample.

References

Aït-Sahalia, Y., 2004. Disentangling diffusions from jumps. Journal of Financial Economics 74, 487–528.
Aït-Sahalia, Y., Jacod, J., 2007. Volatility estimators for discretely sampled Lévy processes. Annals of Statistics 35, 355–392.
Aït-Sahalia, Y., Jacod, J., 2008. Fisher's information for discretely sampled Lévy processes. Econometrica 76, 727–761.
Aït-Sahalia, Y., Jacod, J., 2009. Estimating the degree of activity of jumps in high frequency financial data. Annals of Statistics 37, 2202–2244.
Akritas, M.G., Johnson, R.A., 1981. Asymptotic inference in Lévy processes of the discontinuous type. Annals of Statistics 9, 604–614.
Basawa, I.V., Brockwell, P.J., 1982. Non-parametric estimation of non-decreasing Lévy processes. Journal of the Royoal Statistical Society Series B 44, 262–269.
Beirlant, J., Dierickx, G., Guilloui, A., Starica, C., 2002. On exponential representations of log-spacings of extreme order statistics. Extremes 5, 157–180.
Belomestny, D., 2009. Spectral estimation of the fractional order of a Lévy process. Annals of Statistics (in press).
Belomestny, D., Reiss, M., 2006. Spectral calibration of exponential Lévy models. Finance and Stochastics 10, 449–474.
Butucea, C., Matias, C., 2005. Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. Bernoulli 11, 309–340.
Butucea, C., Tsybakov, A.B., 2007. Sharp optimality for density deconvolution with dominating bias. II. Theoretical Probability and Applications 52, 336–349.
Chen, S., Delaigle, A., Hall, P., 2009. Nonparametric estimation for a class of Lévy processes (long version). Manuscript.
Cont, R., Tankov, P., 2003. Financial Modelling with Jump Processes. Chapman and Hall, CRC.
Crovella, M., Taqqu, M., Bestavros, A., 1998. Heavy tailed probability distributions in the world wide web. In: Adler, R.J., Feldman, R.E., Taqqu, M.S. (Eds.), A Practical Guide to Heavy Tails: Statistical Techniques and Applications. Birkhäuser, Boston.
Csörgő, S., Deheuvels, P., Mason, D., 1985. Kernel estimates of the tail index of a distribution. Annals of Statistics 13, 1050–1077.
Danielsson, J., De Haan, L., Peng, L., 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. Journal of Multivariate Analysis 76, 226–248.
De Sousa, B., Michailidis, G., 2004. A diagnostic plot for estimating the tail index of a distribution. Journal of Computational and Graphical Statistics 13, 974–995.
Drees, J., Kaufmann, E., 1998. Selecting the optimal sample fraction in univariate extreme value estimation. Stochastic Processes and Applications 75, 149–172.
Fan, J., 1991. On the optimal rates of convergence for nonparametric deconvolution problems. Annals of Statistics 19, 1257–1272.
Feuerverger, A., McDunnough, P., 1981. On the efficiency of empirical characteristic function procedures. Journal of Royal Statistical Society Series B 43, 20–27.
Groeneboom, P., Lopuhaa, H.P., De Wolf, P.P., 2003. Kernel-type estimators for the extreme value index. Annals of Statistics 31, 1956–1995.
Guillou, A., Hall, P., 2001. A diagnostic for selecting the threshold in extreme value analysis. Journal of Royal Statistical Society Series B 63, 293–305.
Hall, P., 1982. On some simple estimates of an exponent of regular variation. Journal of Royal Statistical Society Series B 44, 37–42.
Hall, P., Lahiri, S.N., 2008. Estimation of distributions, moments and quantiles in deconvolutions. Annals of Statistics 36, 2110–2134.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. Annals of Statistics 3, 1163–1174.
Jedidi, W., 2001. Local asymptotic normality of statistical models associated with discrete observations of Lévy processes. Manuscript.
Johannes, M., 2004. The economic and statistical role of jumps in continuous-time interest rate models. Journal of Finance 59, 227–260.
Li, T., Vuong, Q., 1998. Nonparametric estimation of the measurement error model using multiple indicators. Journal of Multivariate Analysis 65, 139–165.
Neumann, M.H., Reiss, M., 2009. Nonparametric estimation for Lévy processes from low-frequency observations. Bernoulli 15, 223–248.
Rachev, S.T., 2003. Hand Book of Heavy Tailed Distributions in Finance. Elsevier, Amsterdam.
Resnick, S., Stărică, C., 1997. Smoothing the Hill estimator. Advance in Applied Probability 29, 271–293.
Rubin, H., Tucker, H.G., 1959. Estimating the parameters of a differential process. Annals of Mathematical Statistics 30, 641–658.
Sato, K.-I., 1999. Lévy Processes and Infinitely Divisible Distributions. Cambridge University Press, Cambridge.
Singleton, K.J., 2001. Estimation of affine asset pricing models using the empirical characteristic function. Journal of Econometrics 102, 111–141.
Woerner, J.H.C., 2004. Estimating the skewness in discretely observed Lévy processes. Econometric Theory 20, 927–942.
Woerner, J.H.C., 2006. Purely discontinuous Lévy processes and power variation: Inference for integrated volatility and the scale parameter. In: Shiryaev, A.N., Grosshino, M.R., Oliveira, P., Esquivel, M. (Eds.), Stochastic Finance. Springer, Berlin, pp. 264–276.