Nonparametric inference of gradual changes in the jump behaviour of time-continuous processes

Michael Hoffmann*, Mathias Vetter† and Holger Dette*,

Ruhr-Universität Bochum & Christian-Albrechts-Universität zu Kiel

September 22, 2016

Abstract

In applications changes of the properties of a stochastic feature occur often gradually rather than abruptly, that is: after a constant phase for some time they slowly start to change. Efficient analysis for change points should address the specific features of such a smooth change. In this paper we discuss statistical inference for localizing and detecting gradual changes in the jump characteristic of a discretely observed Itō semimartingale. We propose a new measure of time variation for the jump behaviour of the process. The statistical uncertainty of a corresponding estimate is analyzed deriving new results on the weak convergence of a sequential empirical tail integral process and a corresponding multiplier bootstrap procedure.

Keywords and Phrases: Lévy measure; empirical processes; weak convergence, gradual changes

AMS Subject Classification: 60F17, 60G51, 62G10.

1 Introduction

Stochastic processes in continuous time are widely used in the applied sciences nowadays, as they allow for a flexible modeling of the evolution of various real-life phenomena over time. Speaking of mathematical finance, of particular interest is the family of semimartingales, which is theoretically appealing as it satisfies a certain condition on the absence of arbitrage in financial markets and yet is rich enough to reproduce stylized facts from empirical finance such as volatility clustering, leverage effects or jumps. For this reason, the development of statistical tools modeled by discretely observed Itō semimartingales has been a major topic over the last years, both regarding the estimation of crucial quantities used for model calibration purposes and with a view on tests to check whether a certain model fits the data well. For a detailed overview of the

[1]Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany. E-mail: holger.dette@rub.de, michael.hoffmann@rub.de

[2]Christian-Albrechts-Universität zu Kiel, Mathematisches Seminar, Ludewig-Meyn-Str. 4, 24118 Kiel, Germany. E-mail: vetter@math.uni-kiel.de

state of the art we refer to the recent monographs by Jacod and Protter (2012) and Aït-Sahalia and Jacod (2014).

These statistical tools typically differ highly, depending on the quantities of interest. When the focus is on the volatility, most concepts are essentially concerned with discrete observations of the continuous martingale part. In this case one is naturally close to the Gaussian framework, and so a lot of classical concepts from standard parametric statistics turn out to be powerful methods. The situation is different with a view on the jump behaviour of the process, mainly for two reasons: There is much more flexibility in the choice of the jump measure than there is regarding the diffusive part, and even if one restricts the model to certain parametric families the standard situation is the one of β-stable processes, $0 < \beta < 2$, which are quite difficult to deal with, at least in comparison to Brownian motion. To mention recent work besides the aforementioned monographs, see for example Nickl et al. (2016) and Hoffmann and Vetter (2016) on the estimation of the jump distribution function of a Lévy process or Todorov (2015) on the estimation of the jump activity index from high-frequency observations.

In the following, we are interested in the evolution of the jump behaviour over time in a completely non-parametric setting where we assume only stuctural conditions on the characteristic triplet of the underlying Itō semimartingale. To be precise, let $X = (X_t)_{t\geq 0}$ be an Itō semimartingale with a decomposition

$$X_t = X_0 + \int_0^t b_s \, ds + \int_0^t \sigma_s \, dW_s + \int_0^t \int_{\mathbb{R}} u 1_{\{|u|\leq 1\}} (\mu - \bar{\mu})(ds, du) + \int_0^t \int_{\mathbb{R}} u 1_{\{|u|>1\}} \mu(du, dz), \quad (1.1)$$

where W is a standard Brownian motion, μ is a Poisson random measure on $\mathbb{R}^+ \times \mathbb{R}$, and the predictable compensator $\bar{\mu}$ satisfies $\bar{\mu}(ds, du) = ds \, \nu_s(du)$. The main quantity of interest is the kernel ν_s which controls the number and the size of the jumps around time s.

In Bücher et al. (2016) the authors are interested in the detection of abrupt changes in the jump measure of X. Based on high-frequency observations $X_{i\Delta_n}$, $i = 0, \dots, n$, with $\Delta_n \to 0$ they construct a test for a constant ν against the alternative

$$\nu_t^{(n)} = 1_{\{t < \lfloor n\theta_0 \rfloor \Delta_n\}} \nu_1 + 1_{\{t \geq \lfloor n\theta_0 \rfloor \Delta_n\}} \nu_2.$$

Here the authors face a similar problem as in the classical situation of changes in the mean of a time series, namely that the "change point" θ_0 can only be defined relative to the length of the covered time horizon $n\Delta_n$ which needs to tend to infinity. In general, this problem cannot be avoided as there are only finitely many large jumps on every compact interval, so consistent estimators for the jump measure have to be constructed over the entire positive half-line.

There are other types of changes in the jump behaviour of a process than just abrupt ones, though. In the sequel, we will deal with gradual (smooth, continuous) changes of ν_s and discuss how and how well they can be detected. A similar problem has recently been addressed in Todorov (2016) who constructs a test for changes in the activity index. Since this index is determined by the infinitely many small jumps around zero, such a test can be constructed over a day. On the other hand, estimation of an index is obviously a simpler problem than estimation of an entire measure.

While the problem of detecting abrupt changes has been discussed intensively in a time series context [see Aue and Horváth (2013) and Jandhyala et al. (2013) for a review of the literature], detecting gradual changes is a much harder problem and the methodology is not so well developed. Most authors consider nonparametric location or parametric models with independently distributed observations and we refer to Bissell (1984), Gan (1991), Siegmund and Zhang (1994), Hušková (1999), Hušková and Steinebach (2002) and Mallik et al. (2013) among others [see also Aue and Steinebach (2002) for some results in a time series model]. Recently Vogt and Dette (2015) developed a nonparametric method to estimate a change point corresponding to a smooth change of a locally stationary time series, and the present paper is devoted to the development of nonparametric inference for gradual changes in the jump properties of a discretely observed Itō semimartingale.

In Section 2 we introduce the formal setup as well as a measure of time variation which is similar to Vogt and Dette (2015) and used to identify changes in the jump characteristic later on. Section 3 is concerned with weak convergence of an estimator for this measure, and as a consequence we also obtain weak convergence of related statistics which can be used for testing for a gradual change and for localizing the first change point. As the limiting distribution depends in a complicated way on the unknown jump characteristic, a bootstrap procedure is discussed as well to quantify the uncertainty of the analysis. Section 4 contains the formal derivation of an estimator of the change point and a test for gradual changes. Finally, all proofs are relegated to Section 5.

2 Preliminaries and a measure of gradual changes

In the sequel let $X^{(n)} = (X_t^{(n)})_{t\geq 0}$ be an Itō semimartingale of the form (1.1) with characteristic triplet $(b_s^{(n)}, \sigma_s^{(n)}, \nu_s^{(n)})$ for each $n \in \mathbb{N}$. We are interested in investigating gradual changes in the evolution of the jump behaviour and assume throughout this paper that there is a driving law behind this evolution which is common for all $n \in \mathbb{N}$. Formally, we introduce a transition kernel $g(y, dz)$ from $([0,1], \mathbb{B}([0,1]))$ into $(\mathbb{R}, \mathbb{B})$ such that

$$\nu_s^{(n)}(dz) = g\Big(\frac{s}{n\Delta_n}, dz\Big)$$

for $s \in [0, n\Delta_n]$. This transition kernel shall be an element of the set $\mathcal{G}$ to be defined below. Throughout the paper $\mathbb{B}(A)$ denotes the trace σ-algebra of a set $A \subset \mathbb{R}$ with respect to the Borel σ-algebra.

Assumption 2.1. Let $\mathcal{G}$ denote the set of all transition kernels $g(\cdot, dz)$ from $([0,1], \mathbb{B}([0,1]))$ into $(\mathbb{R}, \mathbb{B})$ such that

(1) For each $y \in [0,1]$ the measure $g(y, dz)$ does not charge $\{0\}$.

(2) The function $y \mapsto \int (1 \wedge z^2) g(y, dz)$ is bounded on the interval $[0, 1]$.

(3) If

$$\mathcal{I}(z) := \begin{cases} [z, \infty), & \text{if } z > 0 \\ (-\infty, z], & \text{if } z < 0 \end{cases}$$

denotes one-sided intervals and

$$g(y, z) := g(y, \mathcal{I}(z)) = \int_{\mathcal{I}(z)} g(y, dx); \quad (y, z) \in [0, 1] \times \mathbb{R} \setminus \{0\},$$

then for every $z \in \mathbb{R} \setminus \{0\}$ there exists a finite set $M^{(z)} = \{t_1^{(z)}, \ldots, t_{n_z}^{(z)} \mid n_z \in \mathbb{N}\} \subset [0, 1]$, such that the function $y \mapsto g(y, z)$ is continuous on $[0, 1] \setminus M^{(z)}$.

(4) For each $y \in [0, 1]$ the measure $g(y, dz)$ is absolutely continuous with respect to the Lebesgue measure with density $z \mapsto h(y, z)$, where the measurable function $h \colon ([0, 1] \times \mathbb{R}, \mathbb{B}([0, 1]) \otimes \mathbb{B}) \to (\mathbb{R}, \mathbb{B})$ is continuously differentiable with respect to $z \in \mathbb{R} \setminus \{0\}$ for fixed $y \in [0, 1]$. The function $h(y, z)$ and its derivative will be denoted by $h_y(z)$ and $h_y'(z)$, respectively. Furthermore, we assume for each $\varepsilon > 0$ that

$$\sup_{y \in [0,1]} \sup_{z \in M_\varepsilon} \Big(h_y(z) + |h_y'(z)| \Big) < \infty,$$

where $M_\varepsilon = (-\infty, -\varepsilon] \cup [\varepsilon, \infty)$.

These assumptions are all rather mild. For each fixed y, the integral $\int (1 \wedge z^2) g(y, dz)$ needs to be finite by properties of the jump compensator, so part (2) just serves as a condition on uniform boundedness over time. Part (3) essentially says that for each z only finitely many discontinuous changes of the jump measure $g(y, \cdot)$ are allowed. Finally, note that the existence of a density as in (4) is a standard condition when estimating a measure in a non-parametric framework.

In order to investigate gradual changes in the jump behaviour of the underlying process we follow Vogt and Dette (2015) and consider a measure of time variation for the jump behaviour, which is defined by

$$D(\zeta, \theta, z) := \int_0^\zeta g(y, z) dy - \frac{\zeta}{\theta} \int_0^\theta g(y, z) dy, \tag{2.1}$$

where $(\zeta, \theta, z) \in C \times \mathbb{R} \setminus \{0\}$ and

$$C := \{(\zeta, \theta) \in [0, 1]^2 \mid \zeta \leq \theta\}. \tag{2.2}$$

Here and throughout this paper we use the convention $\frac{0}{0} := 1$.

The time varying measure (2.1) will be the main theoretical tool for our inference of gradual changes in the jump behaviour of the process (1.1). Our analysis will be based on the following observation: Due to $\bar{\mu}^{(n)}(ds, du) = ds\nu_s^{(n)}(du)$ the jump behaviour corresponding to the first $\lfloor n\theta \rfloor$ observations for some $\theta \in (0, 1)$ does not vary, if and only if the kernel $g(\cdot, dz)$ is Lebesgue almost everywhere constant on the interval $[0, \theta]$. In this case we have $D(\zeta, \theta, z) \equiv 0$ for all $0 \leq \zeta \leq \theta$ and $z \in \mathbb{R} \setminus \{0\}$, since $\zeta^{-1} \int_0^\zeta g(y, z)dy$ is constant on $[0, \theta]$ for each $z \in \mathbb{R} \setminus \{0\}$. If on the other hand $D(\zeta, \theta, z) = 0$ for all $\zeta \in [0, \theta]$ and $z \in \mathbb{R} \setminus \{0\}$, then

$$\int_0^\zeta g(y, z)dy = \zeta\Big(\frac{1}{\theta}\int_0^\theta g(y, z)dy\Big) =: \zeta A(z)$$

for each $\zeta \in [0, \theta]$ and fixed $z \in \mathbb{R} \setminus \{0\}$. Therefore by the fundamental theorem of calculus and Assumption 2.1(3) for each fixed $z \in \mathbb{R} \setminus \{0\}$ we have $g(y, z) = A(z)$ for every $y \in [0, \theta] \setminus M^{(z)}$. As a consequence

$$g(y, z) = A(z) \tag{2.3}$$

holds for every $z \in \mathbb{Q} \setminus \{0\}$ and each $y \in [0, \theta]$ outside the Lebesgue null set $\bigcup_{z \in \mathbb{Q} \setminus \{0\}} M^{(z)}$. Due to Assumption 2.1(2) and Lebesgue's dominated convergence theorem $A(z)$ is left-continuous for positive $z \in \mathbb{R} \setminus \{0\}$ and right-continuous for negative $z \in \mathbb{R} \setminus \{0\}$. The same holds for $g(y, z)$ for each fixed $y \in [0, \theta]$. Consequently (2.3) holds for every $z \in \mathbb{R} \setminus \{0\}$ and each $y \in [0, \theta]$ outside the Lebesgue null set $\bigcup_{z \in \mathbb{Q} \setminus \{0\}} M^{(z)}$. Thus by the uniqueness theorem for measures the kernel $g(\cdot, dz)$ is on $[0, \theta]$ Lebesgue almost everywhere equal to the Lévy measure defined by $A(z)$.

In practice we restict ourselves to z which are bounded away from zero, as typically $g(y, z) \to \infty$ as $z \to 0$, at least if we deviate from the (simple) case of finite activity jumps. Below we discuss two standard applications of $D(\zeta, \theta, z)$ we have in mind.

(1) (*test for a gradual change*) If one defines

$$\tilde{\mathcal{D}}^{(\varepsilon)}(\theta) := \sup_{|z| \geq \varepsilon} \sup_{0 \leq \zeta \leq \theta} |D(\zeta, \theta, z)| \tag{2.4}$$

for some pre-specified constant $\varepsilon > 0$, one can characterize the existence of a change point as follows: There exists a gradual change in the behaviour of the jumps larger than ε of the process (1.1) if and only if

$$\sup_{\theta \in [0,1]} \tilde{\mathcal{D}}^{(\varepsilon)}(\theta) > 0.$$

Moreover for the analysis of gradual changes it is equivalent to consider

$$\mathcal{D}^{(\varepsilon)}(\theta) := \sup_{|z| \geq \varepsilon} \sup_{0 \leq \zeta \leq \theta' \leq \theta} |D(\zeta, \theta', z)|, \tag{2.5}$$

because the first time points where $\mathcal{D}^{(\varepsilon)}$ and $\tilde{\mathcal{D}}^{(\varepsilon)}$ deviate from zero, if existent, coincide

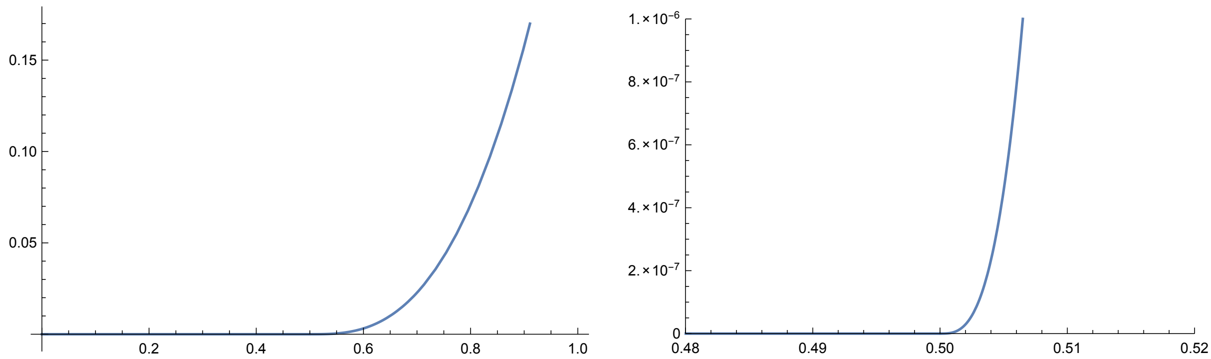

Figure 1: *The function* $\mathcal{D}^{(\varepsilon)}$ *for the transition kernel* (2.7)*, where* $\varepsilon = 1$*. The "true" change point is located at* $\theta_0 = 1/2$*.*

and this point is characteristic for a gradual change as we have seen previously. In this paper we consider $\mathcal{D}^{(\varepsilon)}$ only, since due to its monotonicity it simplifies several steps in the proofs and our notation. In Section 4.1 we construct a consistent estimator, say $\mathbb{D}_n^{(\varepsilon)}$, of $\mathcal{D}^{(\varepsilon)}$. The test for gradual changes in the behaviour of the jumps larger than ε of the process (1.1) rejects the null hypothesis for large values of $\mathbb{D}_n^{(\varepsilon)}(1)$. Quantiles for this test will be derived by a multiplier bootstrap.

(2) (*estimating the gradual change point*) In Section 4.1 we construct an estimator for the first point where the behaviour of the jumps larger than ε changes (gradually). For this purpose we also use the time varying measure (2.1) and define

$$\theta_0^{(\varepsilon)} := \inf\left\{\theta \in [0,1] \mid \mathcal{D}^{(\varepsilon)}(\theta) > 0\right\}, \tag{2.6}$$

where we set $\inf \varnothing := 1$. We call $\theta_0^{(\varepsilon)}$ the change point of the jumps larger than ε of the underlying process (1.1).

A typical example is displayed in Figure 1 where we show the function $\theta \mapsto \mathcal{D}^{(\varepsilon)}(\theta)$ defined in (2.5) for $\varepsilon = 1$, where the transition kernel is given by

$$g(y,z) = \begin{cases} 10e^{-|z|} & \text{if} \quad y \in [0, \frac{1}{2}] \\ 10\big(1 + 3(y-\frac{1}{2})^2\big)e^{-|z|} & \text{if} \quad y \in [\frac{1}{2}, 1]. \end{cases} \tag{2.7}$$

From the right panel it is clearly visible that the function $\mathcal{D}^{(\varepsilon)}$ is positive for all $\theta \in (\frac{1}{2}, 1]$, which identifies $\theta_0 = 1/2$ as the change point. Additionally we illustrate the previously introduced quantities in two further examples.

Example 2.2. (*abrupt changes*) The classical change point problem, where the jump behaviour of the underlying process is constant on two intervals, is contained in our analysis. To be precise, assume that $0 < \theta_0 < 1$ and that ν_1 and ν_2 are Lévy measures such that the transition kernel g

satisfies

$$g(y,dz) = \begin{cases} \nu_1(dz), & \text{for } y \in [0,\theta_0] \\ \nu_2(dz), & \text{for } y \in (\theta_0, 1]. \end{cases} \tag{2.8}$$

If each ν_j is absolutely continuous with respect to the Lebesgue measure, and if it has a density h_j which is continuously differentiable at any point $z_0 \neq 0$, satisfying

$$\sup_{|z|\geq\varepsilon} \{h_j(z) + |h_j'(z)|\} < \infty$$

for every $\varepsilon > 0$, then the kernel g satisfies Assumption 2.1.
For a Lévy measure ν on $(\mathbb{R}, \mathbb{B})$ and $z \in \mathbb{R} \setminus \{0\}$ let $\nu(z) := \nu(\mathcal{I}(z))$. If $g \in \mathcal{G}$ is of the form (2.8) and $\varepsilon > 0$ is chosen sufficiently small such that there exists a $\bar{z} \in \mathbb{R}$ with $|\bar{z}| \geq \varepsilon$ and

$$\nu_1(\bar{z}) \neq \nu_2(\bar{z}),$$

then we have $D(\zeta, \theta', z) = 0$ for all $(\zeta, \theta', z) \in B_\varepsilon := C \times M_\varepsilon$ with $\theta' \leq \theta_0$ and consequently $\mathcal{D}^{(\varepsilon)}(\theta) = 0$ for each $\theta \leq \theta_0$. On the other hand, if $\theta_0 < \theta' < 1$ and $\zeta \leq \theta_0$ we have

$$D(\zeta, \theta', z) = \zeta\nu_1(z) - \frac{\zeta}{\theta'}(\theta_0\nu_1(z) + (\theta' - \theta_0)\nu_2(z)) = \zeta(\nu_2(z) - \nu_1(z))\Big(\frac{\theta_0}{\theta'} - 1\Big)$$

and we obtain

$$\sup_{\zeta\leq\theta_0} \sup_{|z|\geq\varepsilon} |D(\zeta, \theta', z)| = V_\varepsilon\theta_0\Big(1 - \frac{\theta_0}{\theta'}\Big),$$

where $V_\varepsilon = \sup\limits_{|z|\geq\varepsilon} |\nu_1(z) - \nu_2(z)| > 0$. For $\theta_0 < \zeta \leq \theta'$ a similar calculation yields

$$D(\zeta, \theta', z) = \theta_0(\nu_2(z) - \nu_1(z))\Big(\frac{\zeta}{\theta'} - 1\Big)$$

which gives

$$\sup_{\theta_0<\zeta\leq\theta'} \sup_{|z|\geq\varepsilon} |D(\zeta, \theta', z)| = V_\varepsilon\theta_0\Big(1 - \frac{\theta_0}{\theta'}\Big).$$

It follows that the quantity defined (2.6) is given by $\theta_0^{(\varepsilon)} = \theta_0$, because for $\theta > \theta_0$ we have

$$\mathcal{D}^{(\varepsilon)}(\theta) = \sup_{\theta_0<\theta'\leq\theta} \max\Big\{\sup_{\zeta\leq\theta_0} \sup_{|z|\geq\varepsilon} |D(\zeta, \theta', z)|, \sup_{\theta_0<\zeta\leq\theta'} \sup_{|z|\geq\varepsilon} |D(\zeta, \theta', z)|\Big\} = V_\varepsilon\theta_0\Big(1 - \frac{\theta_0}{\theta}\Big). \tag{2.9}$$

Example 2.3. (*Locally symmetric β-stable jump behaviour*) A Lévy process is symmetric β-stable for some $0 < \beta < 2$ if and only if its Brownian part vanishes and its Lévy measure has a Lebesgue density of the form $h(z) = A/|z|^{1+\beta}$ with $A \in \mathbb{R}_+$ [see, for instance, Chapter 3 in Sato (1999)]. In this sense we say that an Itō semimartingale with decomposition (1.1) satisfying

(2.14) has locally symmetric β-stable jump behaviour, if the corresponding transition kernel g is given by

$$g(y, \mathcal{I}(z)) = g(y, z) \quad \text{and} \quad g(y, \{0\}) = 0, \quad \text{with} \quad g(y, z) = A(y)/|z|^{\beta(y)} \tag{2.10}$$

for $y \in [0,1]$ and $z \in \mathbb{R} \setminus \{0\}$. Here the functions $\beta : [0,1] \to (0,2)$ and $A : [0,1] \to (0,\infty)$ are continuous outside a finite set, A is bounded and β is bounded away from 2. In the Appendix we show that a kernel of the form (2.10) satisfies Assumption 2.1 (see Section 5.10).
Now, let $\theta_0 \in (0,1)$ with

$$A(y) = A_0 \quad \text{and} \quad \beta(y) = \beta_0 \tag{2.11}$$

for all $y \in [0, \theta_0]$ with $A_0 \in (0,\infty)$, $\beta_0 \in (0,2)$. Assume furthermore that θ_0 is contained in an open interval U on which a real analytic function $\bar{A} \colon U \to \mathbb{R}$ with

$$\bar{A}(y) = \sum_{k=0}^{\infty} a_k (y - \theta_0)^k \text{ for } y \in U \text{ with } |a_k|^{1/k} = O(1/k) \text{ as } k \to \infty \tag{2.12}$$

and an affine linear function $\bar{\beta} \colon U \to \mathbb{R}$ with $\bar{\beta}(y) = b_0 + b_1(y - \theta_0)$ exist, such that at least one of the functions $\bar{A}$, $\bar{\beta}$ is non-constant and

$$A(y) = \bar{A}(y) \quad \text{and} \quad \beta(y) = \bar{\beta}(y) \tag{2.13}$$

for all $y \in [\theta_0, 1) \cap U$. Then we also show in the Appendix that the quantity defined in (2.6) is given by $\theta_0^{(\varepsilon)} = \theta_0$ for every $\varepsilon > 0$ (see Section 5.10).

We conclude this section with the main assumption for the characteristics of an Itō semimartingale which will be used throughout this paper.

Assumption 2.4. For each $n \in \mathbb{N}$ let $X^{(n)}$ denote an Itō semimartingale of the form (1.1) with characteristics $(b_s^{(n)}, \sigma_s^{(n)}, \nu_s^{(n)})$ that satisfy

(a) There exists a $g \in \mathcal{G}$ such that

$$\nu_s^{(n)}(dz) = g\Big(\frac{s}{n\Delta_n}, dz\Big) \tag{2.14}$$

holds for all $s \in [0, n\Delta_n]$ and all $n \in \mathbb{N}$.

(b) The drift $b_s^{(n)}$ and the volatility $\sigma_s^{(n)}$ are predictable processes and satisfy

$$\sup_{n \in \mathbb{N}} \sup_{s \in \mathbb{R}_+} \Big(\mathbb{E}|b_s^{(n)}|^{\alpha} \vee \mathbb{E}|\sigma_s^{(n)}|^{p}\Big) < \infty,$$

for some $p > 2$, with $\alpha = 3p/(p+4) \in (1,3)$.

(c) The observation scheme $\{X_{i\Delta_n}^{(n)} \mid i = 0, \dots, n\}$ satisfies

$$\Delta_n \to 0, \quad n\Delta_n \to \infty, \quad \text{and} \quad n\Delta_n^{1+\tau} \to 0,$$

for $\tau = (p-2)/(p+1) \in (0,1)$.

3 Weak convergence

In order to estimate the measure of time variation introduced in (2.1) we use the sequential empirical tail integral process defined by

$$U_n(\theta, z) = \frac{1}{k_n} \sum_{j=1}^{\lfloor n\theta \rfloor} 1_{\{\Delta_j^n X^{(n)} \in \mathcal{I}(z)\}}, \tag{3.1}$$

where $\Delta_j^n X^{(n)} = X_{j\Delta_n}^{(n)} - X_{(j-1)\Delta_n}^{(n)}$, $\theta \in [0,1]$, $z \in \mathbb{R} \setminus \{0\}$ and $k_n := n\Delta_n$. The process U_n counts the number of increments that fall into $\mathcal{I}(z)$, as these are likely to be caused by a jump with the corresponding size, and will be the basic tool for estimating the measure of time variation in (2.1). The estimate is defined by

$$\mathbb{D}_n(\zeta, \theta, z) := U_n(\zeta, z) - \frac{\zeta}{\theta} U_n(\theta, z), \quad (\zeta, \theta, z) \in C \times \mathbb{R} \setminus \{0\}, \tag{3.2}$$

where the set C is defined in (2.2). The statistic $U_n(1,z)$ has been considered by Figueroa-López (2008) for observations of a Lévy process Y, so without a time-varying jump behaviour. In this case the author shows that this statistic is in fact an L^2-consistent estimator for the tail integral $\nu(\mathcal{I}(z)) = U(z)$. The following theorem provides a generalization of this statement. In particular, it provides the weak convergence of the sequential empirical tail integral

$$\mathbb{G}_n(\theta, z) := \sqrt{k_n} \Big\{ U_n(\theta, z) - \int_0^\theta g(y, z) dy \Big\}. \tag{3.3}$$

Throughout this paper we use the notation

$$A_\varepsilon = [0,1] \times M_\varepsilon$$

and $R_1 \triangle R_2$ denotes the symmetric difference of two sets R_1, R_2.

Theorem 3.1. *If Assumption 2.4 is satisfied, then we have $\mathbb{G}_n \rightsquigarrow \mathbb{G}$ in $\ell^\infty(A_\varepsilon)$ for any $\varepsilon > 0$, where $\mathbb{G}$ is a tight mean zero Gaussian process with covariance function*

$$H(\theta_1, z_1; \theta_2, z_2) = \operatorname{Cov}(\mathbb{G}(\theta_1, z_1), \mathbb{G}(\theta_2, z_2)) = \int_0^{\theta_1 \wedge \theta_2} g(y, \mathcal{I}(z_1) \cap \mathcal{I}(z_2)) dy.$$

The sample paths of $\mathbb{G}$ are almost surely uniformly continuous with respect to the semimetric

$$\rho((\theta_1, z_1); (\theta_2, z_2)) := \Big\{ \int_0^{\theta_1} g(y, \mathcal{I}(z_1) \triangle \mathcal{I}(z_2)) dy + \int_{\theta_1}^{\theta_2} g(y, \mathcal{I}(z_2)) dy \Big\}^{\frac{1}{2}},$$

defined for $\theta_1 \le \theta_2$ without loss of generality. Moreover, the space (A_ε, ρ) is totally bounded.

Recall the definition of the measure of time variation for the jump behaviour defined in (2.1) and the definition of the set C in (2.2). For $B_\varepsilon = C \times M_\varepsilon$ consider the functional $\Phi \colon \ell^\infty(A_\varepsilon) \to \ell^\infty(B_\varepsilon)$ defined by

$$\Phi(f)(\zeta, \theta, z) := f(\zeta, z) - \frac{\zeta}{\theta} f(\theta, z). \tag{3.4}$$

As $\|\Phi(f_1) - \Phi(f_2)\|_{B_\varepsilon} \le 2\|f_1 - f_2\|_{A_\varepsilon}$ the mapping Φ is Lipschitz continuous. Consequently, $\mathbb{H} := \Phi(\mathbb{G})$ is a tight mean zero Gaussian process in $\ell^\infty(B_\varepsilon)$ with covariance structure

$$\begin{aligned}
&\operatorname{Cov}(\mathbb{H}(\zeta_1, \theta_1, z_1), \mathbb{H}(\zeta_2, \theta_2, z_2)) = \\
&\quad = \int_0^{\zeta_1 \wedge \zeta_2} g(y, \mathcal{I}(z_1) \cap \mathcal{I}(z_2)) dy - \frac{\zeta_1}{\theta_1} \int_0^{\zeta_2 \wedge \theta_1} g(y, \mathcal{I}(z_1) \cap \mathcal{I}(z_2)) dy \\
&\quad\quad - \frac{\zeta_2}{\theta_2} \int_0^{\zeta_1 \wedge \theta_2} g(y, \mathcal{I}(z_1) \cap \mathcal{I}(z_2)) dy + \frac{\zeta_1 \zeta_2}{\theta_1 \theta_2} \int_0^{\theta_1 \wedge \theta_2} g(y, \mathcal{I}(z_1) \cap \mathcal{I}(z_2)) dy.
\end{aligned} \tag{3.5}$$

From the continuous mapping theorem we obtain weak convergence of the process

$$\mathbb{H}_n(\zeta, \theta, z) := \Phi(\mathbb{G}_n)(\zeta, \theta, z) = \sqrt{k_n}(\mathbb{D}_n(\zeta, \theta, z) - D(\zeta, \theta, z)). \tag{3.6}$$

Theorem 3.2. *If Assumption 2.4 is satisfied, then the process $\mathbb{H}_n$ defined in (3.6) satisfies $\mathbb{H}_n \rightsquigarrow \mathbb{H}$ in $\ell^\infty(B_\varepsilon)$ for any $\varepsilon > 0$, where $\mathbb{H}$ is a tight mean zero Gaussian process with covariance function (3.5).*

For the statistical change-point inference proposed in the following section we require the quantiles of functionals of the limiting distribution in Theorem 3.2. This distribution depends in a complicated way on the unknown underlying kernel $g \in \mathcal{G}$ and, as a consequence, corresponding quantiles are difficult to estimate.
A typical approach to problems of this type are resampling methods. One option is to use suitable estimates for drift, volatility and the unknown kernel g to draw independent samples of an Itō semimartingale. However, such a method is computationally expensive since one has to generate independent Itō semimartingales for each stage within the bootstrap algorithm. Therefore we propose an alternative bootstrap method based on multipliers. For this resampling method one only needs to generate n i.i.d. random variables with mean zero and variance one. See Inoue (2001) for a similar approach in the context of empirical processes.

To be precise let $X_1, \ldots, X_n$ and $\xi_1, \ldots, \xi_n$ denote random variables defined on probability spaces $(\Omega_X, \mathcal{A}_X, \mathbb{P}_X)$ and $(\Omega_\xi, \mathcal{A}_\xi, \mathbb{P}_\xi)$, respectively, and consider a random element $\hat{Y}_n = \hat{Y}_n(X_1, \ldots, X_n, \xi_1, \ldots, \xi_n)$ on the product space $(\Omega, \mathcal{A}, \mathbb{P}) := (\Omega_X, \mathcal{A}_X, \mathbb{P}_X) \otimes (\Omega_\xi, \mathcal{A}_\xi, \mathbb{P}_\xi)$ which maps into a metric space, say $\mathbb{D}$. Moreover, let Y be a tight, Borel measurable $\mathbb{D}$-valued random variable. Following Kosorok (2008) we call $\hat{Y}_n$ *weakly convergent to* Y *conditional on* $X_1, X_2, \ldots$ *in probability* if the following two conditions are satisfied

(a) $\sup_{f \in \mathrm{BL}_1(\mathbb{D})} |\mathbb{E}_\xi f(\hat{Y}_n) - \mathbb{E} f(Y)| \xrightarrow{\mathbb{P}^*} 0,$

(b) $\mathbb{E}_\xi f(\hat{Y}_n)^* - \mathbb{E}_\xi f(\hat{Y}_n)_* \xrightarrow{\mathbb{P}^*} 0$ for all $f \in \mathrm{BL}_1(\mathbb{D})$.

Here, $\mathbb{E}_\xi$ denotes the conditional expectation with respect to $\xi_1, \ldots, \xi_n$ given $X_1, \ldots, X_n$, whereas $\mathrm{BL}_1(\mathbb{D})$ is the space of all real-valued Lipschitz continuous functions f on $\mathbb{D}$ with sup-norm $\|f\|_\infty \leq 1$ and Lipschitz constant 1. Moreover, $f(\hat{Y}_n)^*$ and $f(\hat{Y}_n)_*$ denote a minimal measurable majorant and a maximal measurable minorant with respect to $\xi_1, \ldots, \xi_n, X_1, \ldots, X_n$, respectively. Throughout this paper we denote this type of convergence by $\hat{Y}_n \rightsquigarrow_\xi Y$.
In the following we will work with a multiplier bootstrap version of the process $\mathbb{G}_n$, that is

$$\begin{aligned} \hat{\mathbb{G}}_n &= \hat{\mathbb{G}}_n(\theta, z) = \hat{\mathbb{G}}_n(X^{(n)}_{\Delta_n}, \ldots, X^{(n)}_{n\Delta_n}, \xi_1, \ldots, \xi_n; \theta, z) \\ &:= \frac{1}{n\sqrt{k_n}} \sum_{j=1}^{\lfloor n\theta \rfloor} \sum_{i=1}^{n} \xi_j \{ \mathbb{1}_{\{\Delta_j^n X^{(n)} \in \mathcal{I}(z)\}} - \mathbb{1}_{\{\Delta_i^n X^{(n)} \in \mathcal{I}(z)\}} \} \\ &= \frac{1}{\sqrt{k_n}} \sum_{j=1}^{\lfloor n\theta \rfloor} \xi_j \{ \mathbb{1}_{\{\Delta_j^n X^{(n)} \in \mathcal{I}(z)\}} - \eta_n(z) \}, \end{aligned} \tag{3.7}$$

where $\xi_1, \ldots, \xi_n$ are independent and identically distributed random variables with mean 0 and variance 1 and $\eta_n(z) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{\Delta_i^n X^{(n)} \in \mathcal{I}(z)\}}$. The following theorem establishes conditional weak convergence of this bootstrap approximation for the sequential empirical tail integral process $\mathbb{G}_n$.

Theorem 3.3. *If Assumption 2.4 is satisfied and $(\xi_j)_{j \in \mathbb{N}}$ is a sequence of independent and identically distributed random variables with mean 0 and variance 1, defined on a distinct probability space as described above, then $\hat{\mathbb{G}}_n \rightsquigarrow_\xi \mathbb{G}$ in $(\ell^\infty(A_\varepsilon), \|\cdot\|_{A_\varepsilon})$ for any $\varepsilon > 0$, where $\mathbb{G}$ denotes the limiting process of Theorem 3.1.*

Theorem 3.3 suggests to define the following counterparts of the process $\mathbb{H}_n$ defined in (3.6)

$$\begin{aligned} \hat{\mathbb{H}}_n(\zeta, \theta, z) &:= \hat{\mathbb{H}}_n(X^{(n)}_{\Delta_n}, \ldots, X^{(n)}_{n\Delta_n}; \xi_1, \ldots, \xi_n; \zeta, \theta, z) := \hat{\mathbb{G}}_n(\zeta, z) - \frac{\zeta}{\theta} \hat{\mathbb{G}}_n(\theta, z) \\ &= \frac{1}{\sqrt{n\Delta_n}} \Big[\sum_{j=1}^{\lfloor n\zeta \rfloor} \xi_j \{ \mathbb{1}_{\{\Delta_j^n X^{(n)} \in \mathcal{I}(z)\}} - \eta_n(z) \} - \\ &\qquad - \frac{\zeta}{\theta} \sum_{j=1}^{\lfloor n\theta \rfloor} \xi_j \{ \mathbb{1}_{\{\Delta_j^n X^{(n)} \in \mathcal{I}(z)\}} - \eta_n(z) \} \Big]. \end{aligned} \tag{3.8}$$

The following result establishes consistency of $\hat{\mathbb{H}}_n$. Its proof is a consequence of Proposition 10.7 in Kosorok (2008), because we have $\hat{\mathbb{H}}_n = \Phi(\hat{\mathbb{G}}_n)$ and $\mathbb{H} = \Phi(\mathbb{G})$ with the Lipschitz continuous map Φ defined in (3.4).

Theorem 3.4. *If Assumption 2.4 is satisfied, we have $\hat{\mathbb{H}}_n \rightsquigarrow_\xi \mathbb{H}$ in $(\ell^\infty(B_\varepsilon), \|\cdot\|_{B_\varepsilon})$ for any $\varepsilon > 0$, where the process $\mathbb{H}$ is defined in Theorem 3.2.*

4 Statistical inference for gradual changes

Bücher et al. (2016) proposed test and estimation procedures in the situation of an abrupt change, that is as in the situation of Example 2.2 when it is known that the kernel $\nu_s^{(n)}$ is constant before and after the change point. Their test is based on an empirical CUSUM process

$$\mathbb{T}_n(\theta, z) = \sqrt{k_n}\Big\{U_n(\theta, z) - \frac{\lfloor n\theta \rfloor}{n} U_n(1, z)\Big\} \quad \text{for } (\theta, z) \in A_\varepsilon,$$

where U_n is defined in (3.1). It can easily be seen that in Example 2.2 with $\nu_1 = \nu_2$, which is their null hypothesis, $D(\zeta, \theta, z) \equiv 0$ for all $(\zeta, \theta, z) \in B_\varepsilon$. This implies for a transition kernel of the form (2.8)

$$|\mathbb{T}_n(\theta, z) - \mathbb{H}_n(\theta, 1, z)| = \sqrt{k_n} U_n(1, z)\Big|\frac{\lfloor n\theta \rfloor}{n} - \theta\Big| = o_{\mathbb{P}}(1)$$

uniformly in $(\theta, z) \in A_\varepsilon$. Consequently, the results of Bücher et al. (2016) follow from the statements of Section 3.

While the classical CUSUM test has advantages in the detection of abrupt change points it is less appropriate to detect gradual changes. To be precise, recall from (3.3) that $k_n^{-1/2}\mathbb{T}_n(\theta, z)$ is a consistent estimate of the quantity

$$\tau(\theta, z) := \int_0^\theta g(y, z)dy - \theta \int_0^1 g(y, z)dy.$$

Thus, the statistic $\text{argmax}_{\theta \in [0,1]} |\mathbb{T}_n(\theta, z)|$, which is the estimator for the change point in Bücher et al. (2016), estimates $\text{argmax}_{\theta \in [0,1]} |\tau(\theta, z)|$. However, if the jump behaviour changes gradually at the point θ_0 the function $|\tau(\theta, z)|$ is not necessarily maximal at the point θ_0. An example is displayed in Figure 2 where we show the function $\theta \mapsto |\tau(\theta, 1)|$ for the transition kernel in (2.7). Here the function τ is maximal for some $\theta > 1/2$. From a practical point of view this means that the argmax estimator based on the classical CUSUM statistic usually overestimates the unknown change point, if this change is not abrupt. These problems are also reflected in the power of the classical CUSUM test.

We see from this example, and especially in Figure 1, that the quantity $\mathcal{D}^{(\varepsilon)}(\theta)$ defined in (2.5) is better suited to detect gradual changes in the behaviour of the jumps larger than ε. Therefore we use the estimate $\mathbb{D}_n(\zeta, \theta, z)$ of the measure of time variation $D(\zeta, \theta, z)$ defined in (3.2) to construct a test for the existence and an estimator for the location of a gradual change point.

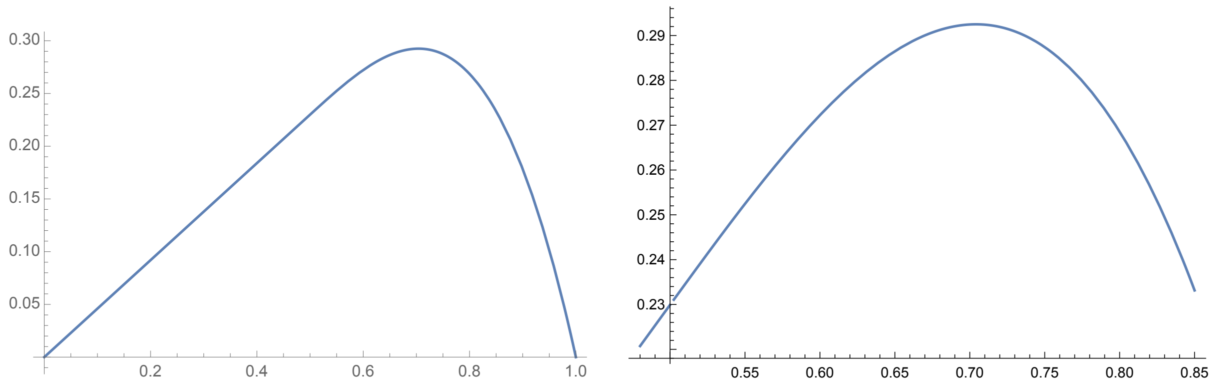

Figure 2: *The function $\theta \mapsto |\tau(\theta, 1)|$ for the transition kernel (2.7). The "true" change point is located at $\theta_0 = 1/2$.*

We begin with the problem of estimating the point of a gradual change in the jump behaviour. The discussion of tests will be referred to Section 4.2.

4.1 Localizing change points

Recall the definition

$$\mathcal{D}^{(\varepsilon)}(\theta) = \sup_{|z|\geq\varepsilon} \sup_{0\leq\zeta\leq\theta'\leq\theta} |D(\zeta, \theta', z)|$$

and the definition of the change point $\theta_0^{(\varepsilon)}$ in (2.6). By Theorem 3.2 the process $\mathbb{D}_n(\zeta, \theta, z)$ from (3.2) is a consistent estimator of $D(\zeta, \theta, z)$. Therefore we set

$$\mathbb{D}_n^{(\varepsilon)}(\theta) := \sup_{|z|\geq\varepsilon} \sup_{0\leq\zeta\leq\theta'\leq\theta} |\mathbb{D}_n(\zeta, \theta', z)|,$$

and an application of the continuous mapping theorem and Theorem 3.2 yields the following result.

Corollary 4.1. *If Assumption 2.4 is satisfied, then $k_n^{1/2}\mathbb{D}_n^{(\varepsilon)} \rightsquigarrow \mathbb{H}^{(\varepsilon)}$ in $\ell^\infty\big([0, \theta_0^{(\varepsilon)}]\big)$, where $\mathbb{H}^{(\varepsilon)}$ is the tight process in $\ell^\infty([0,1])$ defined by*

$$\mathbb{H}^{(\varepsilon)}(\theta) := \sup_{|z|\geq\varepsilon} \sup_{0\leq\zeta\leq\theta'\leq\theta} |\mathbb{H}(\zeta, \theta', z)|, \tag{4.1}$$

with the centered Gaussian process $\mathbb{H}$ defined in Theorem 3.2.

Intuitively, the estimation of $\theta_0^{(\varepsilon)}$ is more difficult, if the curve $\theta \mapsto \mathcal{D}^{(\varepsilon)}(\theta)$ is "flat" at $\theta_0^{(\varepsilon)}$. Following Vogt and Dette (2015), we describe the curvature of $\theta \mapsto \mathcal{D}^{(\varepsilon)}(\theta)$ by a local polynomial behaviour of the function $\mathcal{D}^{(\varepsilon)}(\theta)$ for values $\theta > \theta_0^{(\varepsilon)}$. More precisely, we assume throughout this section that $\theta_0^{(\varepsilon)} < 1$ and that there exist constants $\lambda, \eta, \varpi, c^{(\varepsilon)} > 0$ such that $\mathcal{D}^{(\varepsilon)}$ admits an expansion of the form

$$\mathcal{D}^{(\varepsilon)}(\theta) = c^{(\varepsilon)}\big(\theta - \theta_0^{(\varepsilon)}\big)^{\varpi} + \aleph(\theta) \tag{4.2}$$

for all $\theta \in [\theta_0^{(\varepsilon)}, \theta_0^{(\varepsilon)} + \lambda]$, where the remainder term satisfies $|\aleph(\theta)| \leq K\big(\theta - \theta_0^{(\varepsilon)}\big)^{\varpi+\eta}$ for some $K > 0$. The construction of an estimator for $\theta_0^{(\varepsilon)}$ utilizes the fact that, by Theorem 3.2, $k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(\theta) \to \infty$ in probability for any $\theta \in (\theta_0^{(\varepsilon)}, 1]$. We now consider the statistic

$$r_n^{(\varepsilon)}(\theta) := 1_{\{k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(\theta) \leq \varkappa_n\}},$$

for a deterministic sequence $\varkappa_n \to \infty$. From the previous discussion we expect

$$r_n^{(\varepsilon)}(\theta) \to \begin{cases} 1, & \text{if } \theta \leq \theta_0^{(\varepsilon)} \\ 0, & \text{if } \theta > \theta_0^{(\varepsilon)} \end{cases}$$

in probability if the threshold level $\varkappa_n$ is chosen appropriately. Consequently, we define the estimator for the change point by

$$\hat{\theta}_n^{(\varepsilon)} := \int_0^1 r_n^{(\varepsilon)}(\theta) d\theta.$$

Note that the estimate $\hat{\theta}_n^{(\varepsilon)}$ depends on the threshold $\varkappa_n$ and we make this dependence visible in our notation, i.e. $\hat{\theta}_n^{(\varepsilon)} = \hat{\theta}_n^{(\varepsilon)}(\varkappa_n)$, whenever it is necessary.

Theorem 4.2. *If Assumption 2.4 is satisfied, $\theta_0^{(\varepsilon)} < 1$, and (4.2) holds for some $\varpi > 0$, then*

$$\hat{\theta}_n^{(\varepsilon)} - \theta_0^{(\varepsilon)} = O_{\mathbb{P}}\Big(\big(\frac{\varkappa_n}{\sqrt{k_n}}\big)^{1/\varpi}\Big),$$

for any sequence $\varkappa_n \to \infty$ with $\varkappa_n/\sqrt{k_n} \to 0$.

Theorem 4.2 makes the heuristic argument of the previous paragraph more precise. A lower degree of smoothness in $\theta_0^{(\varepsilon)}$ yields a better rate of convergence of the estimator. Moreover, the slower the threshold level $\varkappa_n$ converges to infinity the better the rate of convergence. We will explain below how to choose this sequence to control the probability of over- and underestimation by using bootstrap methods. Before that we investigate the mean squared error

$$\mathrm{MSE}^{(\varepsilon)}(\varkappa_n) = \mathbb{E}\Big[\big(\hat{\theta}_n^{(\varepsilon)}(\varkappa_n) - \theta_0^{(\varepsilon)}\big)^2\Big]$$

of the estimator $\hat{\theta}_n^{(\varepsilon)}$. Recall the definition of $\mathbb{H}_n$ in (3.6) and define

$$\mathbb{H}_n^{(\varepsilon)}(\theta) := \sup_{|z|\geq\varepsilon} \sup_{0\leq\zeta\leq\theta'\leq\theta} |\mathbb{H}_n(\zeta, \theta', z)|, \quad \theta \in [0,1], \tag{4.3}$$

which measures the absolute distance between the estimator $\mathbb{D}_n^{(\varepsilon)}(\theta)$ and the true value $\mathcal{D}^{(\varepsilon)}(\theta)$. For a sequence $\alpha_n \to \infty$ with $\alpha_n = o(\varkappa_n)$ we decompose the MSE into

$$\mathrm{MSE}_1^{(\varepsilon)}(\varkappa_n, \alpha_n) \ := \ \mathbb{E}\Big[\big(\hat{\theta}_n^{(\varepsilon)} - \theta_0^{(\varepsilon)}\big)^2 1_{\big\{\mathbb{H}_n^{(\varepsilon)}(1)\leq\alpha_n\big\}}\Big],$$

$$\mathrm{MSE}_2^{(\varepsilon)}(\varkappa_n, \alpha_n) \ := \ \mathbb{E}\Big[\big(\hat{\theta}_n^{(\varepsilon)} - \theta_0^{(\varepsilon)}\big)^2 1_{\left\{\mathbb{H}_n^{(\varepsilon)}(1) > \alpha_n\right\}}\Big] \leq \mathbb{P}\big(\mathbb{H}_n^{(\varepsilon)}(1) > \alpha_n\big),$$

which can be considered as the MSE due to small and large estimation error. With these notations the following theorem gives upper and lower bounds for the mean squared error.

Theorem 4.3. *Suppose that $\theta_0^{(\varepsilon)} < 1$, Assumption 2.4 and (4.2) are satisfied. Then for any sequence $\alpha_n \to \infty$ with $\alpha_n = o(\varkappa_n)$ we have*

$$K_1\big(\frac{\varkappa_n}{\sqrt{k_n}}\big)^{2/\varpi} \leq \mathrm{MSE}_1^{(\varepsilon)}(\varkappa_n, \alpha_n) \leq K_2\big(\frac{\varkappa_n}{\sqrt{k_n}}\big)^{2/\varpi} \tag{4.4}$$
$$\mathrm{MSE}_2^{(\varepsilon)}(\varkappa_n, \alpha_n) \leq \mathbb{P}\big(\mathbb{H}_n^{(\varepsilon)}(1) > \alpha_n\big),$$

for $n \in \mathbb{N}$ sufficiently large, where the constants K_1 and K_2 can be chosen as

$$K_1 = \left(\frac{1-\varphi}{c^{(\varepsilon)}}\right)^{2/\varpi} \quad \text{and} \quad K_2 = \left(\frac{1+\varphi}{c^{(\varepsilon)}}\right)^{2/\varpi} \tag{4.5}$$

for arbitrary $0 < \varphi < 1$.

In the remaining part of this section we discuss the choice of the regularizing sequence $\varkappa_n$ for the estimator $\hat{\theta}_n^{(\varepsilon)}$. Our main goal here is to control the probability of an over- and underestimation of the change point $0 < \theta_0^{(\varepsilon)} < 1$.
For this purpose let $\hat{\theta}_n$ be a preliminary consistent estimator of $\theta_0^{(\varepsilon)}$. For example, if (4.2) holds for some $\varpi > 0$, one can take $\hat{\theta}_n = \hat{\theta}_n^{(\varepsilon)}(\varkappa_n)$ for a sequence $\varkappa_n \to \infty$ satisfying the assumptions of Theorem 4.2. In the sequel, let $B \in \mathbb{N}$ be some large number and let $(\xi^{(b)})_{b=1,\ldots,B}$ denote independent vectors of i.i.d. random variables, $\xi^{(b)} := (\xi_j^{(b)})_{j=1,\ldots,n}$, with mean zero and variance one, which are defined on a probability space distinct to the one generating the data $\{X_{i\Delta_n}^{(n)} \mid i = 0, \ldots, n\}$. We denote by $\hat{\mathbb{G}}_{n,\xi^{(b)}}$ or $\hat{\mathbb{H}}_{n,\xi^{(b)}}^{(\varepsilon)}$ the particular bootstrap statistics calculated with respect to the data and the bootstrap multipliers $\xi_1^{(b)}, \ldots, \xi_n^{(b)}$ from the b-th iteration, where

$$\hat{\mathbb{H}}_n^{(\varepsilon)}(\theta) := \sup_{|z|\geq\varepsilon} \sup_{0\leq\zeta\leq\theta'\leq\theta} |\hat{\mathbb{H}}_n(\zeta, \theta', z)| \tag{4.6}$$

for $\theta \in [0, 1]$. With these notations and for $\varepsilon > 0$, $B, n \in \mathbb{N}$ and an $0 < r \leq 1$ we define the following empirical distribution function

$$K_{n,B}^{(\varepsilon,r)}(x) = \frac{1}{B} \sum_{i=1}^{B} \mathbb{1}_{\{(\hat{\mathbb{H}}_{n,\xi^{(i)}}^{(\varepsilon)}(\hat{\theta}_n))^r \leq x\}},$$

and denote by

$$K_{n,B}^{(\varepsilon,r)-}(y) := \inf\left\{x \in \mathbb{R} \mid K_{n,B}^{(\varepsilon,r)}(x) \geq y\right\}$$

its pseudoinverse. Given a confidence level $0 < \alpha < 1$ we consider the threshold

$$\hat{\varkappa}_{n,B}^{(\varepsilon,\alpha)}(r) := K_{n,B}^{(\varepsilon,r)-}(1-\alpha). \tag{4.7}$$

This choice is optimal in the sense of the following two theorems.

Theorem 4.4. *Let $\varepsilon > 0$, $0 < \alpha < 1$ and assume that Assumption 2.4 is satisfied for some $g \in \mathcal{G}$ with $0 < \theta_0^{(\varepsilon)} < 1$. Suppose further that there exists some $\bar{z} \in M_\varepsilon$ with*

$$\int_0^{\theta_0^{(\varepsilon)}} g(y,\bar{z})dy > 0. \tag{4.8}$$

Then the probability for underestimation of the change point $\theta_0^{(\varepsilon)}$ can be controlled by

$$\limsup_{B\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big(\hat{\theta}_n^{(\varepsilon)}(\hat{\varkappa}_{n,B}^{(\varepsilon,\alpha)}(1)) < \theta_0^{(\varepsilon)}\Big) \le \alpha. \tag{4.9}$$

Theorem 4.5. *Let $\varepsilon > 0$, $0 < r < 1$. Assume that Assumption 2.4 is satisfied for some $g \in \mathcal{G}$ with $0 < \theta_0^{(\varepsilon)} < 1$ and that (4.2) holds for some $\varpi, c^{(\varepsilon)} > 0$. Furthermore suppose that there exist a constant $\rho > 0$ with $n\Delta_n^{1+\rho} \to \infty$ and a $\bar{z} \in M_\varepsilon$ satisfying (4.8). Additionally let the bootstrap multipliers be either bounded in absolute value or distributed according to $\mathcal{N}(0,1)$. Then for each $K > \big(1/c^{(\varepsilon)}\big)^{1/\varpi}$ and all sequences $(\alpha_n)_{n\in\mathbb{N}} \subset (0,1)$ with $\alpha_n \to 0$ and $(B_n)_{n\in\mathbb{N}} \subset \mathbb{N}$ with $B_n \to \infty$ such that*

1. $\alpha_n^2 B_n \to \infty$,
2. $(n\Delta_n)^{\frac{1-r}{2r}}\alpha_n \to \infty$,

we have

$$\lim_{n\to\infty} \mathbb{P}\Big(\hat{\theta}_n^{(\varepsilon)}(\hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r)) > \theta_0^{(\varepsilon)} + K\beta_n\Big) = 0, \tag{4.10}$$

where $\beta_n = (\hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r)/\sqrt{k_n})^{1/\varpi} \xrightarrow{\mathbb{P}} 0$, while $\hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r) \xrightarrow{\mathbb{P}} \infty$.

Obviously, Theorem 4.5 only gives a meaningful result if $\beta_n \xrightarrow{\mathbb{P}} 0$ can be guaranteed. Its proof shows that a sufficient condition for this property is given by

$$\mathbb{P}\Big(\hat{\mathbb{H}}_n^{(\varepsilon)}(\hat{\theta}_n) > (\sqrt{k_n}x)^{1/r}\Big) \le \mathbb{P}\Big(\hat{\mathbb{H}}_n^{(\varepsilon)}(1) > (\sqrt{k_n}x)^{1/r}\Big) = o(\alpha_n)\ . \tag{4.11}$$

Moreover, (4.11) follows from $(n\Delta_n)^{\frac{1-r}{2r}}\alpha_n \to \infty$ without any further conditions. This explains why the threshold $0 < r < 1$ needs to be introduced, and it seems that the statement of (4.11) can only be guaranteed under very restrictive assumptions in the case $r = 1$.
Finally we illustrate that the polynomial behaviour introduced in (4.2) is satisfied in the situations of Example 2.2 and Example 2.3.

Example 4.6. (1) Recall the situation of an abrupt change in the jump characteristic considered in Example 2.2. In this case it follows from (2.9) that

$$\mathcal{D}^{(\varepsilon)}(\theta) = V_\varepsilon \theta_0 \Big(1 - \frac{\theta_0}{\theta}\Big) = V_\varepsilon(\theta - \theta_0) - \frac{V_\varepsilon}{\theta}(\theta - \theta_0)^2 > 0,$$

whenever $\theta_0 < \theta \le 1$. Therefore assumption (4.2) is satisfied with $\aleph(\theta) = -\frac{V_\varepsilon}{\theta}(\theta - \theta_0)^2$. Moreover, the transition kernel given by (2.8) satisfies also assumption (4.8) if $\nu_1 \neq 0$ and $\varepsilon > 0$ is chosen small enough.

(2) In the situation discussed in Example 2.3 let $\bar{g}(y, z) = \bar{A}(y)/|z|^{\bar{\beta}(y)}$ for $y \in U$ and $z \in M_\varepsilon$. Then we have for any $\varepsilon > 0$

$$k_{0,\varepsilon} := \min \Big\{ k \in \mathbb{N} \mid \exists z \in M_\varepsilon : g_k(z) \neq 0 \Big\} < \infty,$$

where for $k \in \mathbb{N}_0$ and $z \in \mathbb{R} \setminus \{0\}$

$$g_k(z) := \Big(\frac{\partial^k \bar{g}}{\partial y^k} \Big)\Big|_{(\theta_0, z)}$$

denotes the k-th partial derivative of $\bar{g}$ with respect to y at (θ_0, z), which is a bounded function on any M_ε. Furthermore for every $\varepsilon > 0$ there is a $\lambda > 0$ such that

$$\mathcal{D}^{(\varepsilon)}(\theta) = \Big(\frac{1}{(k_{0,\varepsilon} + 1)!} \sup_{|z| \ge \varepsilon} |g_{k_{0,\varepsilon}}(z)| \Big) (\theta - \theta_0)^{k_{0,\varepsilon}+1} + \aleph(\theta) \tag{4.12}$$

on $[\theta_0, \theta_0 + \lambda]$ with $|\aleph(\theta)| \le K\big(\theta - \theta_0\big)^{k_{0,\varepsilon}+2}$ for some $K > 0$, so (4.2) is satisfied. A proof for this result can be found in the Appendix as well. Again, (4.8) holds also.

4.2 Testing for a gradual change

Bücher et al. (2016) introduced change point tests for the situation of an abrupt change as in Example 2.2, where the jump behaviour is assumed to be constant before and after the change point. Formally, they tested the hypotheses

$$\mathbf{H}_0^{(ab)}(\varepsilon) : \nu_1(z) = \nu_2(z) \ \forall |z| \ge \varepsilon \quad \text{versus} \quad \mathbf{H}_1^{(ab)}(\varepsilon) : \exists\ |z| \ge \varepsilon \text{ such that } \nu_1(z) \neq \nu_2(z). \tag{4.13}$$

In this section we want to derive test procedures for the existence of a gradual change in the data. In order to formulate suitable hypotheses for a gradual change point recall the definition of the measure of time variation for the jump behavior in (2.1) and define for $\varepsilon > 0$, $z_0 \in \mathbb{R} \setminus \{0\}$ and $\theta \in [0, 1]$ the quantities

$$\mathcal{D}^{(\varepsilon)}(\theta) := \sup_{|z| \ge \varepsilon} \sup_{0 \le \zeta \le \theta' \le \theta} |D(\zeta, \theta', z)|$$
$$\mathcal{D}_{(z_0)}(\theta) := \sup_{0 \le \zeta \le \theta' \le \theta} |D(\zeta, \theta', z_0)|.$$

We also assume that Assumption 2.4 is satisfied and we are interested in the following hypotheses

$$\mathbf{H}_0(\varepsilon):\ \mathcal{D}^{(\varepsilon)}(1)=0 \qquad \text{versus} \qquad \mathbf{H}_1(\varepsilon):\ \mathcal{D}^{(\varepsilon)}(1)>0, \tag{4.14}$$

which refer to the global behaviour of the tail integral. If one is interested in the gradual change in the tail integral for a fixed $z_0 \in \mathbb{R}\setminus\{0\}$ one could consider the hypotheses

$$\mathbf{H}_0^{(z_0)}:\ \mathcal{D}_{(z_0)}(1)=0 \qquad \text{versus} \qquad \mathbf{H}_1^{(z_0)}:\ \mathcal{D}_{(z_0)}(1)>0. \tag{4.15}$$

Remark 4.7. Note that the function D in (2.1) is uniformly continuous on every B_ε. More precisely, for any $\eta>0$ there exists a $\delta>0$ such that

$$|D(\zeta_1,\theta_1,z)-D(\zeta_2,\theta_2,z)|<\eta$$

holds for each $z\in M_\varepsilon$ and all pairs $(\zeta_1,\theta_1),(\zeta_2,\theta_2)\in C=\{(\zeta,\theta)\in[0,1]^2 \mid \zeta\le\theta\}$ with maximum distance $\|(\zeta_1,\theta_1)-(\zeta_2,\theta_2)\|_\infty<\delta$. Therefore the function $D^{(\varepsilon)}(\zeta,\theta)=\sup_{z\in M_\varepsilon}|D(\zeta,\theta,z)|$ is uniformly continuous on C and as a consequence $\mathcal{D}^{(\varepsilon)}$ is continuous on $[0,1]$. Thus the alternative $\mathbf{H}_1(\varepsilon)$ holds if and only if the point $\theta_0^{(\varepsilon)}$ defined in (2.6) satisfies $\theta_0^{(\varepsilon)}<1$.

The null hypothesis in (4.14) and (4.15) will be rejected for large values of the corresponding estimators

$$\mathbb{D}_n^{(\varepsilon)}(1) \quad \text{and} \quad \sup_{(\zeta,\theta)\in C}|\mathbb{D}_n(\zeta,\theta,z_0)|$$

for $\mathcal{D}^{(\varepsilon)}(1)$ and $\mathcal{D}_{(z_0)}(1)$, respectively. To obtain critical values we use the multiplier bootstrap introduced in the second part of Section 3. For this purpose we denote by $\xi_1^{(b)},\dots,\xi_n^{(b)}$, $b=1,\dots,B$, i.i.d. random variables with mean zero and variance one. As before, we assume that these random variables are defined on a probability space distinct to the one generating the data $\{X_{i\Delta_n}^{(n)} \mid i=0,\dots,n\}$. We denote by $\hat{\mathbb{G}}_{n,\xi^{(b)}}$ and $\hat{\mathbb{H}}_{n,\xi^{(b)}}$ the statistics in (3.7) and (3.8) calculated from $\{X_{i\Delta_n}^{(n)} \mid i=0,\dots,n\}$ and the b-th bootstrap multipliers $\xi_1^{(b)},\dots,\xi_n^{(b)}$. For given $\varepsilon>0$, $z_0\in\mathbb{R}\setminus\{0\}$ and a given level $\alpha\in(0,1)$, we propose to reject $\mathbf{H}_0(\varepsilon)$ in favor of $\mathbf{H}_1(\varepsilon)$, if

$$k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(1)\ge \hat{q}_{1-\alpha}^{(B)}\Big(\mathbb{H}_n^{(\varepsilon)}(1)\Big), \tag{4.16}$$

where $\hat{q}_{1-\alpha}^{(B)}\Big(\mathbb{H}_n^{(\varepsilon)}(1)\Big)$ denotes the $(1-\alpha)$-sample quantile of $\hat{\mathbb{H}}_{n,\xi^{(1)}}^{(\varepsilon)}(1),\dots,\hat{\mathbb{H}}_{n,\xi^{(B)}}^{(\varepsilon)}(1)$ with $\hat{\mathbb{H}}_{n,\xi^{(b)}}^{(\varepsilon)}$ defined in (4.6). Note that under the null hypothesis it follows from the definition of the process $\mathbb{H}_n$ in (3.6) that $k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(1)=\mathbb{H}_n^{(\varepsilon)}(1)$, which by Theorem 3.2 and the continuous mapping theorem converges weakly to $\mathbb{H}^{(\varepsilon)}(1)$, defined in (4.1). The bootstrap procedure mimics this behaviour.

Similarly, $\mathbf{H}_0^{(z_0)}$ is rejected in favor of $\mathbf{H}_1^{(z_0)}$ if

$$W_n^{(z_0)}:=k_n^{1/2}\sup_{(\zeta,\theta)\in C}|\mathbb{D}_n(\zeta,\theta,z_0)|\ge \hat{q}_{1-\alpha}^{(B)}(W_n^{(z_0)}), \tag{4.17}$$

where $\hat{q}_{1-\alpha}^{(B)}(W_n^{(z_0)})$ denotes the $(1-\alpha)$-sample quantile of $\hat{W}_{n,\xi^{(1)}}^{(z_0)}, \ldots, \hat{W}_{n,\xi^{(B)}}^{(z_0)}$, and

$$\hat{W}_{n,\xi^{(b)}}^{(z_0)} := \sup_{(\zeta,\theta)\in C} |\hat{\mathbb{H}}_{n,\xi^{(b)}}(\zeta,\theta,z_0)|.$$

Remark 4.8. Since $\varepsilon > 0$ has to be chosen for an application of the test (4.16), one can only detect changes in the jumps larger than ε. From a practical point of view this is not a severe restriction as in most applications only the larger jumps are of interest. If one is interested in the entire jump measure, however, its estimation is rather difficult, at least in the presence of a diffusion component, as $\Delta_n^{1/2}$ provides a natural bound to disentangle jumps from volatility. See Nickl et al. (2016) and Hoffmann and Vetter (2016) for details in case of a Lévy process.

The following two results show that the tests (4.16) and (4.17) are consistent asymptotic level α tests.

Proposition 4.9. *Under $\mathbf{H}_0(\varepsilon)$ and $\mathbf{H}_0^{(z_0)}$, respectively, the tests (4.16) and (4.17) have asymptotic level α. More precisely,*

$$\lim_{B\to\infty}\lim_{n\to\infty} \mathbb{P}\Big(k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(1) \geq \hat{q}_{1-\alpha}^{(B)}(\mathbb{H}_n^{(\varepsilon)}(1))\Big) = \alpha,$$

if there exist $|\bar{z}| \geq \varepsilon$, $\bar{\zeta} \in (0,1)$ with $\int_0^{\bar{\zeta}} g(y,\bar{z})dy > 0$, and

$$\lim_{B\to\infty}\lim_{n\to\infty} \mathbb{P}\Big(W_n^{(z_0)} \geq \hat{q}_{1-\alpha}^{(B)}(W_n^{(z_0)})\Big) = \alpha,$$

if there exists a $\bar{\zeta} \in (0,1)$ with $\int_0^{\bar{\zeta}} g(y,z_0)dy > 0$.

Proposition 4.10. *The tests (4.16) and (4.17) are consistent in the following sense. Under $\mathbf{H}_1(\varepsilon)$ we have for all $B \in \mathbb{N}$*

$$\lim_{n\to\infty} \mathbb{P}\Big(k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(1) \geq \hat{q}_{1-\alpha}^{(B)}(\mathbb{H}_n^{(\varepsilon)}(1))\Big) = 1. \tag{4.18}$$

Under $\mathbf{H}_1^{(z_0)}$, we have for all $B \in \mathbb{N}$

$$\lim_{n\to\infty} \mathbb{P}\Big(W_n^{(z_0)} \geq \hat{q}_{1-\alpha}^{(B)}(W_n^{(z_0)})\Big) = 1.$$

Remark 4.11. As illustrated in Example 2.2 the testing procedures (4.16) and (4.17) can be applied to test the hypotheses (4.13) of Bücher et al. (2016) as well (see the representation of the transition kernel in (2.8)).

Acknowledgements This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Projects A1 and C1) of the German Research Foundation (DFG). We would like to thank Martina Stein, who typed

parts of this manuscript with considerable technical expertise.

5 Proofs and technical details

The following assumptions will be used frequently in the sequel.

Assumption 5.1. For each $n \in \mathbb{N}$ let $X^{(n)}$ be an Itō semimartingale of the form (1.1) with characteristics $(b_s^{(n)}, \sigma_s^{(n)}, \nu_s^{(n)})$ and the following properties:

(a) There exists a $g \in \mathcal{G}$ such that

$$\nu_s^{(n)}(dz) = g\Big(\frac{s}{n\Delta_n}, dz\Big)$$

holds for all $s \in [0, n\Delta_n]$ and all $n \in \mathbb{N}$ as measures on $(\mathbb{R}, \mathbb{B})$.

(b) The drift $b_s^{(n)}$ and the volatility $\sigma_s^{(n)}$ are deterministic and Borel measurable functions on $\mathbb{R}_+$. Moreover, these functions are uniformly bounded in $s \in \mathbb{R}_+$ and $n \in \mathbb{N}$.

(c) The observation scheme $\{X_{i\Delta_n}^{(n)} \mid i = 0, \ldots, n\}$ satisfies $\Delta_n \to 0$ and $n\Delta_n \to \infty$.

We begin with an auxiliary result which is a generalization of Lemma 2 in Rüschendorf and Woerner (2002). Throughout this section K denotes a generic constant which typically changes from line to line and may depend on certain bounds and parameters, but not on n.

Lemma 5.2. *Let $T > 0$ and let Y be an Itō semimartingale with $Y_0 = 0$ having a representation as in (1.1) with characteristics (b_s, σ_s, ν_s), where b_s and σ_s are uniformly bounded in $\omega \in \Omega$ and $s \le T$ and ν_s is deterministic. Suppose that there are constants $0 < A, t_0 \le 1$ such that the support of the measure $\int_0^{t_0} \nu_s(dz)ds$ is contained in the set $\{z \mid |z| \le A\}$. Furthermore assume that there is a $g \in \mathcal{G}$ with $g(y, dz) = \nu_{yT}(dz)$ for all $y \in [0,1]$. Then for each $z \in \mathbb{R} \setminus \{0\}$ and $\zeta > 0$ there are $K > 0$ and $0 < t_1 \le t_0 \wedge T$, which depend only on A, z, ζ, the bound on g in Assumption 2.1(2) and the bound on b_s and σ_s, such that the transition probability is bounded by*

$$\mathbb{P}(Y_t \in \mathcal{I}(z)) \le K t^{\frac{|z|}{2A} - \zeta}$$

for all $0 \le t \le t_1$.

Proof. We will only show the inequality for $z > 0$ fixed, because otherwise we can consider the process $-Y$ which has the same properties.
The Hölder inequality and the upper Burkholder-Davis-Gundy inequality yield for $0 < t \le t_0 \wedge T$ and $m \in \mathbb{N}$:

$$\mathbb{E}\Big|\int_0^t b_s ds\Big|^m \le t^m \mathbb{E}\Big(\frac{1}{t}\int_0^t |b_s|^m ds\Big) \le K t^m$$

and

$$\mathbb{E}\Big|\int_0^t \sigma_s dW_s\Big|^m \leq K t^{m/2} \mathbb{E}\Big(\frac{1}{t}\int_0^t |\sigma_s|^2 ds\Big)^{m/2} \leq K t^{m/2}.$$

Therefore by Markov's inequality the claim follows if we can show the lemma for each Itō semimartingale with $b_s = \sigma_s \equiv 0$.
Let Y be such an Itō semimartingale and let $0 < t \leq t_0 \wedge T$. Then by Theorem II.4.15 in Jacod and Shiryaev (2002) Y is a process with independent increments with characteristic function

$$\mathbb{E}\big[\exp\{iuY_t\}\big] = \exp\Big\{\int_0^t \int (e^{iuz} - 1 - iuz)\nu_s(dz)ds\Big\} = \exp\{\Psi_t(iu)\},$$

since $t \leq t_0$ and $A \leq 1$ with

$$\Psi_t(u) := \int_0^t \int (e^{uz} - 1 - uz)\nu_s(dz)ds,$$

which exists for all such t and all $u \in \mathbb{R}$ by a Taylor expansion of the integrand and the assumption on the support of $\int_0^t \nu_s(dz)ds$ as well as item (2) in Assumption 2.1. Furthermore the first two derivatives of Ψ_t are given by

$$\Psi_t'(u) = \int_0^t \int (e^{uz} - 1)z\nu_s(dz)ds; \qquad \Psi_t''(u) = \int_0^t \int z^2 e^{uz}\nu_s(dz)ds$$

where we have exchanged differentiation and integration by the differentiation lemma of measure theory and the assumption on the support of $\int_0^t \nu_s(dz)ds$. W.l.o.g. we may assume that the measure $\int_0^t \nu_s(dz)ds$ is not zero, because otherwise $Y_t = 0$ a.s. and the assertion of the lemma is obvious. Therefore $\Psi_t''(u) > 0$ for all $u \in \mathbb{R}$ and Ψ_t' is a strictly increasing function with $\Psi_t'(0) = 0$ and $\lim_{u\to\infty} \Psi_t'(u) = B \in (0,\infty]$. Thus there is a strictly increasing differentiable inverse function $\tau_t \colon [0,B) \to \mathbb{R}_+$ with $\tau_t(0) = 0$. Moreover, it is sufficient to show the claim for all $0 < z \neq B$, because for B and $\zeta > 0$ we can find $\tilde{z} < B$ and $\tilde{\zeta} < \zeta$ with

$$\frac{B}{2A} - \zeta = \frac{\tilde{z}}{2A} - \tilde{\zeta}$$

and $\mathbb{P}(Y_t \geq B) \leq \mathbb{P}(Y_t \geq \tilde{z})$. Furthermore by Markov's inequality, the identity theorem of complex analysis and Corollary 1.50 in Hoffmann (2016) we have for arbitrary $s \geq 0$:

$$\mathbb{P}(Y_t \geq z) \leq \mathbb{E}\{\exp\{sY_t - sz\}\} = \exp\{\Psi_t(s) - sz\}. \tag{6.1}$$

First suppose that $z > B$. Then we obtain

$$\mathbb{P}(Y_t \geq z) \leq \limsup_{s\to\infty} \exp\Big\{ \int_0^s \Big(\Psi_t'(y) - z\Big)dy\Big\} \leq \lim_{s\to\infty} \exp\{(B-z)s\} = 0$$

and the claim obviously follows. Therefore for the rest of the proof we may assume $z < B$. In this case (6.1) yields (recall that τ_t is the inverse function of Ψ_t')

$$\begin{aligned}\mathbb{P}(Y_t \geq z) &\leq \exp\Big\{ \int_0^{\tau_t(z)} \Psi_t'(y)dy - z\tau_t(z)\Big\} = \exp\Big\{ \int_0^z w\tau_t'(w)dw - z\tau_t(z)\Big\} \\ &= \exp\Big\{ -\int_0^z \tau_t(w)dw\Big\}. \end{aligned} \tag{6.2}$$

By a Taylor expansion we have

$$(e^{Dy} - 1)y \leq e^{DA}Dy^2 \tag{6.3}$$

for $D > 0$ and $|y| \leq A$. Therefore if we set $D = \tau_t(w)$ in (6.3) we obtain

$$\begin{aligned} w = \Psi_t'(\tau_t(w)) &= \int_0^t \int (e^{\tau_t(w)y} - 1)y\nu_s(dy)ds \leq e^{A\tau_t(w)}\tau_t(w) \int_0^t \int y^2\nu_s(dy)ds \\ &\leq e^{A\tau_t(w)}\tau_t(w)Kt, \end{aligned} \tag{6.4}$$

for arbitrary $0 \leq w < B$ and $0 < t \leq t_0 \wedge T$, where the constant $K > 0$ depends only on the bound on g in Assumption 2.1(2). By a series expansion of the exponential function we have

$$\log(\tau_t(w)) \leq A\tau_t(w) \tag{6.5}$$

if $\tau_t(w) \geq \frac{2}{A^2}$ and this is the case if

$$\begin{aligned} \Psi_t'\Big(\frac{2}{A^2}\Big) &= \int_0^t \int (\exp\Big\{\frac{2}{A^2}y\Big\} - 1)y\nu_s(dy)ds \\ &\leq \frac{2}{A^2}e^{\frac{2}{A}} \int_0^t \int y^2\nu_s(dy)ds \leq \frac{2}{A^2}e^{\frac{2}{A}}Kt =: K_0(t) \leq w, \end{aligned} \tag{6.6}$$

where we used (6.3) again. Combining (6.4), (6.5) and (6.6) gives

$$\log\Big(\frac{w}{t}\Big) \leq \log(K) + 2A\tau_t(w) \Longleftrightarrow \frac{1}{2A}\log\Big(\frac{w}{Kt}\Big) \leq \tau_t(w) \tag{6.7}$$

for $K_0(t) \leq w < B$. Let $0 < \bar{t}_1 \leq t_0 \wedge T$ be small enough such that $K_0(t) \leq z < B$ for each

$0 \le t \le \bar{t}_1$. Then (6.7) together with (6.2) yield the estimate

$$
\begin{aligned}
\mathbb{P}(Y_t \ge z) &\le \exp\Big\{ -\int_0^z \tau_t(w)dw \Big\} \le \exp\Big\{ -\frac{1}{2A}\int_{K_0(t)}^z \log\Big(\frac{w}{Kt}\Big)dw \Big\} \\
&= \exp\Big\{ -\frac{Kt}{2A}\int_{K_0(t)/Kt}^{z/Kt} \log(u)du \Big\} = \exp\Big\{ -\frac{Kt}{2A}\Big[-u + u\log(u) \Big]_{K_0(t)/Kt}^{z/Kt} \Big\} \\
&= \exp\Big\{ -\frac{1}{2A}\Big[-z + z\log\Big(\frac{z}{Kt}\Big) + K_0(t) - K_0(t)\log\Big(\frac{K_0(t)}{Kt}\Big) \Big] \Big\} \\
&= \exp\Big\{ -\frac{1}{2A}\Big[-z + z\log\Big(\frac{z}{K}\Big) + K_0(t) - K_0(t)\log\Big(\frac{K_0(t)}{K}\Big) \Big] \Big\} \times \\
&\qquad\qquad\qquad\qquad \times \exp\Big\{ \frac{z}{2A}\log(t) - \frac{K_0(t)}{2A}\log(t) \Big\} \\
&\le \exp\Big\{ \frac{1}{2A}\Big[z - z\log\Big(\frac{z}{K}\Big) + 1 \Big] \Big\} t^{\frac{z}{2A}-\zeta} \;=\; Kt^{\frac{z}{2A}-\zeta}
\end{aligned}
$$

for each $0 < t \le t_1 \le \bar{t}_1 \le t_0 \wedge T$ with a $0 < t_1 \le \bar{t}_1$ small enough such that

$$
\Big| K_0(t) - K_0(t)\log\Big(\frac{K_0(t)}{K}\Big) \Big| \le 1
$$

and $\frac{K_0(t)}{2A} < \zeta$ for every $0 < t \le t_1$. □

The preceding result is helpful to deduce the following claim which is the main tool to establish consistency of $\mathbb{D}_n(\zeta, \theta, z)$ as an estimator for $D(\zeta, \theta, z)$, when it is applied to the Itō semimartingale $Y_s^{(j,n)} = X_{s+(j-1)\Delta_n}^{(n)} - X_{(j-1)\Delta_n}^{(n)}$.

Lemma 5.3. *Suppose that Assumption 5.1 is satisfied and let $\delta > 0$. If $X_0^{(n)} = 0$ for all $n \in \mathbb{N}$, then there exist constants $K = K(\delta) > 0$ and $0 < t_0 = t_0(\delta) \le 1$ such that*

$$
\Big| \mathbb{P}(X_t^{(n)} \in \mathcal{I}(z)) - \int_0^t \nu_s^{(n)}(\mathcal{I}(z))ds \Big| \le Kt^2
$$

holds for all $|z| \ge \delta$, $0 \le t < t_0$ and $n \in \mathbb{N}$ with $n\Delta_n \ge 1$.

Proof of Lemma 5.3. As $n\Delta_n \to \infty$ we may assume $t \le 1 \le n\Delta_n$ throughout this proof. Furthermore, let $\varepsilon < (\delta/6 \wedge 1)$ and pick a smooth cut-off function $c_\varepsilon : \mathbb{R} \to \mathbb{R}$ satisfying

$$
1_{[-\varepsilon/2,\varepsilon/2]}(u) \le c_\varepsilon(u) \le 1_{[-\varepsilon,\varepsilon]}(u).
$$

We also define the function $\bar{c}_\varepsilon$ via $\bar{c}_\varepsilon(u) = 1 - c_\varepsilon(u)$. For $n \in \mathbb{N}$ and $t \in \mathbb{R}_+$ let $M_t^{(n,\varepsilon)}$ be the measure defined by

$$
M_t^{(n,\varepsilon)}(A) = \int_0^t \int_A \bar{c}_\varepsilon(u)\nu_s^{(n)}(du)ds,
$$

for $A \in \mathbb{B}$ which has total mass

$$\lambda_t^{(n,\varepsilon)} := M_t^{(n,\varepsilon)}(\mathbb{R}) = \int_0^t \int_{\mathbb{R}} \bar{c}_\varepsilon(u)\nu_s^{(n)}(du)ds = \int_0^t \int_{\mathbb{R}} \bar{c}_\varepsilon(u) g\Big(\frac{s}{n\Delta_n}, du\Big) ds \leq Kt, \tag{6.8}$$

where K depends only on the bound on g in Assumption 2.1(2) and on ε and therefore on δ. Furthermore, let

$$d_s^{(n,\varepsilon)} := \int u \mathbb{1}_{\{|u|\leq 1\}} \bar{c}_\varepsilon(u)\nu_s^{(n)}(du).$$

By Theorem II.4.15 in Jacod and Shiryaev (2002) for each $n \in \mathbb{N}$ and $t \in \mathbb{R}_+$ with $t \leq n\Delta_n$ we can decompose $X_t^{(n)}$ in law by

$$X_t^{(n)} =_d X_t^{(n,\varepsilon)} + \widetilde{X}_t^{(n,\varepsilon)}, \tag{6.9}$$

where $X^{(n,\varepsilon)}$ and $\widetilde{X}^{(n,\varepsilon)}$ are independent Itō semimartingales starting in zero, with characteristics $(b_s^{(n,\varepsilon)}, \sigma_s^{(n)}, c_\varepsilon(u)\nu_s^{(n)}(du))$, $b_s^{(n,\varepsilon)} := b_s^{(n)} - d_s^{(n,\varepsilon)}$, and $(d_s^{(n,\varepsilon)}, 0, \bar{c}_\varepsilon(u)\nu_s^{(n)}(du))$, respectively. $\widetilde{X}^{(n,\varepsilon)}$ can be seen as a generalized compound Poisson process. To be precise, let $\hat{\mu}^{(n,\varepsilon)}$ be a Poisson random measure independent of $X^{(n,\varepsilon)}$ with predictable compensator $\hat{\nu}^{(n,\varepsilon)}(ds,du) = \bar{c}_\varepsilon(u)\nu_s^{(n)}(du)ds$ and consider the process

$$N_t^{(n,\varepsilon)} := \hat{\mu}^{(n,\varepsilon)}([0,t] \times \mathbb{R}).$$

By Theorem II.4.8 in Jacod and Shiryaev (2002) $N_t^{(n,\varepsilon)}$ is a process with independent increments and distribution

$$N_t^{(n,\varepsilon)} - N_s^{(n,\varepsilon)} =_d \text{ Poiss}\big(\lambda_t^{(n,\varepsilon)} - \lambda_s^{(n,\varepsilon)}\big), \quad (0 \leq s \leq t)$$

(here we use the convention that Poiss(0) is the Dirac measure with mass in zero). Moreover, for each $n \in \mathbb{N}$ let $((Z_j^{(n,\varepsilon)}(t))_{t\in[0,n\Delta_n]})_{j\in\mathbb{N}}$ be a sequence of independent processes, which is also independent of the Poisson random measure $\hat{\mu}^{(n,\varepsilon)}$ and of the process $X^{(n,\varepsilon)}$, such that for each $j \in \mathbb{N}$ and $t \in [0, n\Delta_n]$ its distribution is given by

$$Z_j^{(n,\varepsilon)}(t) =_d \begin{cases} M_t^{(n,\varepsilon)}/\lambda_t^{(n,\varepsilon)}, & \text{if } \lambda_t^{(n,\varepsilon)} > 0 \\ \text{Dirac}(0), & \text{if } \lambda_t^{(n,\varepsilon)} = 0. \end{cases}$$

Then we have for any $n \in \mathbb{N}$ and $0 \leq t \leq n\Delta_n$

$$\widetilde{X}_t^{(n,\varepsilon)} =_d \sum_{j=1}^{\infty} Z_j^{(n,\varepsilon)}(t) \mathbb{1}_{\{j \leq N_t^{(n,\varepsilon)}\}}, \tag{6.10}$$

because by using independence of the involved quantities we calculate the characteristic function for $w \in \mathbb{R}$ and $\lambda_t^{(n,\varepsilon)} > 0$ as follows:

$$\begin{aligned}
&\mathbb{E}\exp\Big\{iw\sum_{j=1}^{\infty} Z_j^{(n,\varepsilon)}(t)\mathbb{1}_{\{j\le N_t^{(n,\varepsilon)}\}}\Big\} = \sum_{j=0}^{\infty}\mathbb{E}\Big(\exp\Big\{iw\sum_{k=1}^{j} Z_k^{(n,\varepsilon)}(t)\Big\}\mathbb{1}_{\{N_t^{(n,\varepsilon)}=j\}}\Big)\\
&\quad= \exp\Big\{-\lambda_t^{(n,\varepsilon)}\Big\}\sum_{j=0}^{\infty}\frac{1}{j!}\big(\Phi(M_t^{(n,\varepsilon)})(w)\big)^j = \exp\Big\{\Phi(M_t^{(n,\varepsilon)})(w)-\lambda_t^{(n,\varepsilon)}\Big\}\\
&\quad= \exp\Big\{iw\int_0^t d_s^{(n,\varepsilon)}ds + \int_0^t\int\Big(e^{iwu}-1-iwu\mathbb{1}_{\{|u|\le 1\}}\Big)\bar{c}_\varepsilon(u)\nu_s^{(n)}(du)ds\Big\}\\
&\quad= \mathbb{E}[\exp(iw\tilde{X}_t^{(n,\varepsilon)})]. \qquad (6.11)
\end{aligned}$$

In the above display $\Phi(M)$ denotes the characteristic function of a finite Borel measure M. The last equality in (6.11) follows from Theorem II.4.15 in Jacod and Shiryaev (2002). Furthermore note that in the case $\lambda_t^{(n,\varepsilon)} = 0$ the distributions in (6.10) are obviously equal.

Let $z \in \mathbb{R}\setminus\{0\}$ with $|z| \ge \delta$, define $f(x) = 1_{\{x\in\mathcal{I}(z)\}}$ and recall the decomposition in (6.9) and the representation (6.10) for $t \le n\Delta_n$. As the processes $X^{(n,\varepsilon)}$ and $\widetilde{X}^{(n,\varepsilon)}$ are independent, we can calculate

$$\begin{aligned}
\mathbb{E}\Big[f\Big(X_t^{(n)}\Big)\Big] &= \sum_{j=0}^{\infty}\exp\Big\{-\lambda_t^{(n,\varepsilon)}\Big\}\Big(\lambda_t^{(n,\varepsilon)}\Big)^j\frac{1}{j!}\mathbb{E}\Big[f\Big(X_t^{(n)}\Big)\Big|N_t^{(n,\varepsilon)}=j\Big]\\
&= \exp\Big\{-\lambda_t^{(n,\varepsilon)}\Big\}\mathbb{E}\Big[f\Big(X_t^{(n,\varepsilon)}\Big)\Big]\\
&+ \exp\Big\{-\lambda_t^{(n,\varepsilon)}\Big\}\lambda_t^{(n,\varepsilon)}\mathbb{E}\Big[f\Big(X_t^{(n,\varepsilon)}+Z_1^{(n,\varepsilon)}(t)\Big)\Big]\\
&+ \sum_{j=2}^{\infty}\exp\Big\{-\lambda_t^{(n,\varepsilon)}\Big\}\Big(\lambda_t^{(n,\varepsilon)}\Big)^j\frac{1}{j!}\mathbb{E}\Big[f\Big(X_t^{(n,\varepsilon)}+\sum_{\ell=1}^{j}Z_\ell^{(n,\varepsilon)}(t)\Big)\Big]. \qquad (6.12)
\end{aligned}$$

For the first summand on the right-hand side of the last display we use Lemma 5.2 with $t_0 = 1, A = \varepsilon, T = n\Delta_n$ and $\zeta = 1$ and obtain

$$\exp\Big\{-\lambda_t^{(n,\varepsilon)}\Big\}\mathbb{E}\Big[f\Big(X_t^{(n,\varepsilon)}\Big)\Big] \le \mathbb{P}\Big(\Big|X_t^{(n,\varepsilon)}\Big| \ge \delta\Big) \le 2Kt^{\delta/2\varepsilon-\zeta} \le Kt^2 \qquad (6.13)$$

for $0 \le t \le \hat{t}_1$, where K and $\hat{t}_1$ depend only on δ, the bound for the transition kernel g in Assumption 2.1(2) and the bounds on b_s and σ_s. Note therefore that $d_s^{(n,\varepsilon)}$ is bounded for $s \le n\Delta_n$ by a bound which depends on ε and thus on δ and the previously mentioned bound on g. Also, for the third term on the right-hand side of (6.12), we have

$$\sum_{j=2}^{\infty}\exp\Big\{-\lambda_t^{(n,\varepsilon)}\Big\}\Big(\lambda_t^{(n,\varepsilon)}\Big)^j\frac{1}{j!}\mathbb{E}\Big[f\Big(X_t^{(n,\varepsilon)}+\sum_{\ell=1}^{j}Z_\ell^{(n,\varepsilon)}(t)\Big)\Big] \le \Big(\lambda_t^{(n,\varepsilon)}\Big)^2 \le Kt^2 \qquad (6.14)$$

by (6.8) since f is bounded by 1. Now if $\lambda_t^{(n,\varepsilon)} = 0$, the second term in (6.12) and $\int_0^t \nu_s^{(n)}(\mathcal{I}(z))ds$ vanish. Hence the lemma follows from (6.13) and (6.14). Thus in the following we assume

$\lambda_t^{(n,\varepsilon)} > 0$ and consider the term $\mathbb{E}\Big[f\Big(X_t^{(n,\varepsilon)} + Z_1^{(n,\varepsilon)}(t)\Big)\Big]$. For $t \leq n\Delta_n$ the distribution of $Z_1^{(n,\varepsilon)}(t)$ has the Lebesgue density

$$u \mapsto \bar{h}_t^{(n,\varepsilon)}(u) := \int_0^t \bar{c}_\varepsilon(u) h\Big(\frac{s}{n\Delta_n}, u\Big) ds / \lambda_t^{(n,\varepsilon)}.$$

As a consequence (for $t \leq n\Delta_n$), the function

$$\rho_t^{(n,\varepsilon)}(x) := \mathbb{E}\Big[f\Big(x + Z_1^{(n,\varepsilon)}(t)\Big)\Big] = \mathbb{P}\Big(x + Z_1^{(n,\varepsilon)}(t) \in \mathcal{I}(z)\Big)$$

is twice continuously differentiable and it follows

$$\sup_{x\in\mathbb{R}} \left\{ \left| \Big(\rho_t^{(n,\varepsilon)}\Big)'(x) \right| + \left| \Big(\rho_t^{(n,\varepsilon)}\Big)''(x) \right| \right\} \leq \frac{Kt}{\lambda_t^{(n,\varepsilon)}}, \tag{6.15}$$

where the constant $K > 0$ depends only on the bound in Assumption 2.1(4) for some $\varepsilon' > 0$ with $\varepsilon' \leq \varepsilon/2$ but not on n or t. Using the independence of $X^{(n,\varepsilon)}$ and $Z_1^{(n,\varepsilon)}$ it is sufficient to discuss $\mathbb{E}[\rho_t^{(n,\varepsilon)}(X_t^{(n,\varepsilon)})]$. Itô's formula (Theorem I.4.57 in Jacod and Shiryaev (2002)) gives

$$\begin{aligned}
\rho_t^{(n,\varepsilon)}(X_r^{(n,\varepsilon)}) = \rho_t^{(n,\varepsilon)}(X_0^{(n,\varepsilon)}) &+ \int_0^r (\rho_t^{(n,\varepsilon)})'(X_{s-}^{(n,\varepsilon)}) dX_s^{(n,\varepsilon)} \\
&+ \frac{1}{2}\int_0^r (\rho_t^{(n,\varepsilon)})''(X_{s-}^{(n,\varepsilon)}) d\langle X^{(n,\varepsilon),c}, X^{(n,\varepsilon),c}\rangle_s \\
&+ \sum_{0<s\leq r} \Big((\rho_t^{(n,\varepsilon)})(X_s^{(n,\varepsilon)}) - (\rho_t^{(n,\varepsilon)})(X_{s-}^{(n,\varepsilon)}) - (\rho_t^{(n,\varepsilon)})'(X_{s-}^{(n,\varepsilon)}) \Delta X_s^{(n,\varepsilon)} \Big),
\end{aligned} \tag{6.16}$$

where $t \leq n\Delta_n$, $r \geq 0$, $\langle X^{(n,\varepsilon),c}, X^{(n,\varepsilon),c}\rangle_s$ denotes the predictable quadratic variation of the continuous local martingale part of $X^{(n,\varepsilon)}$, and $\Delta X_s^{(n,\varepsilon)}$ is the jump size at time s. We now discuss each of the four summands in (6.16) separately for $r = t$: first, $u \in \mathcal{I}(z)$ implies $\bar{c}_\varepsilon(u) = 1$ by definition of ε. Thus, with $X_0^{(n,\varepsilon)} = 0$

$$\begin{aligned}
\rho_t^{(n,\varepsilon)}\Big(X_0^{(n,\varepsilon)}\Big) = \mathbb{P}\Big(Z_1^{(n,\varepsilon)}(t) \in \mathcal{I}(z)\Big) &= \frac{1}{\lambda_t^{(n,\varepsilon)}} \int 1_{\{u\in\mathcal{I}(z)\}} \bar{c}_\varepsilon(u) \int_0^t h\Big(\frac{s}{n\Delta_n}, u\Big) ds du \\
&= \frac{1}{\lambda_t^{(n,\varepsilon)}} \int 1_{\{u\in\mathcal{I}(z)\}} \int_0^t g\Big(\frac{s}{n\Delta_n}, du\Big) ds = \frac{1}{\lambda_t^{(n,\varepsilon)}} \int_0^t \nu_s^{(n)}(\mathcal{I}(z)) ds.
\end{aligned}$$

By the canonical representation of semimartingales (Theorem II.2.34 in Jacod and Shiryaev (2002)) we get the decomposition

$$X_t^{(n,\varepsilon)} = \int_0^t b_s^{(n,\varepsilon)} ds + Y_t^{(n,\varepsilon)},$$

where $Y^{(n,\varepsilon)}$ is a local martingale with characteristics $(0, \sigma_s^{(n)}, c_\varepsilon(u)\nu_s^{(n)}(du))$ which starts at zero and has bounded jumps. Consequently $Y^{(n,\varepsilon)}$ is a locally square integrable martingale and by Proposition I.4.50 b), Theorem I.4.52 and Theorem II.1.8 in the previously mentioned reference its predictable quadratic variation is given by

$$\langle Y^{(n,\varepsilon)}, Y^{(n,\varepsilon)}\rangle_t = \int_0^t \left(\sigma_s^{(n)}\right)^2 ds + \int_0^t \int u^2 c_\varepsilon(u)\nu_s^{(n)}(du)ds.$$

Thus for $t \leq n\Delta_n$ and because of the boundedness of $\left(\rho_t^{(n,\varepsilon)}\right)'$ and the construction of the stochastic integral the integral process $\left(\left(\rho_t^{(n,\varepsilon)}\right)'(X_{s-}^{(n,\varepsilon)}) \cdot Y^{(n,\varepsilon)}\right)^t$ stopped at time t is in fact a square integrable martingale because

$$\mathbb{E}\int_0^t \left(\left(\rho_t^{(n,\varepsilon)}\right)'(X_{s-}^{(n,\varepsilon)})\right)^2 d\langle Y^{(n,\varepsilon)}, Y^{(n,\varepsilon)}\rangle_s < \infty.$$

Therefore we obtain

$$\mathbb{E}\int_0^t \left(\rho_t^{(n,\varepsilon)}\right)'(X_{s-}^{(n,\varepsilon)})dY_s^{(n,\varepsilon)} = 0$$

and according to (6.15) we get a bound for the second term in (6.16):

$$\left|\mathbb{E}\Big[\int_0^t (\rho_t^{(n,\varepsilon)})'(X_{s-}^{(n,\varepsilon)})dX_s^{(n,\varepsilon)}\Big]\right| \leq \int_0^t \left|\mathbb{E}\Big[(\rho_t^{(n,\varepsilon)})'(X_{s-}^{(n,\varepsilon)})\Big]b_s^{(n,\varepsilon)}\right|ds \leq \frac{Kt^2}{\lambda_t^{(n,\varepsilon)}}, \tag{6.17}$$

where $K > 0$ depends only on δ, the bounds on the characteristics and the bounds of Assumption 2.1(2) and (4) for an appropriate $\varepsilon' > 0$. For the third term in (6.16) it is immediate to get an estimate as in (6.17). Finally, let $\mu^{(n,\varepsilon)}(ds, du)$ denote the random measure associated with the jumps of $X^{(n,\varepsilon)}$ which has the predictable compensator $\nu^{(n,\varepsilon)}(ds, du) = ds\nu_s^{(n)}(du)c_\varepsilon(u)$. Therefore Theorem II.1.8 in Jacod and Shiryaev (2002) yields for the expectation of the last term in (6.16)

$$\begin{aligned}
&\mathbb{E}\Big\{\sum_{0<s\leq t} \big(\rho_t^{(n,\varepsilon)}(X_s^{(n,\varepsilon)}) - \rho_t^{(n,\varepsilon)}(X_{s-}^{(n,\varepsilon)}) - (\rho_t^{(n,\varepsilon)})'(X_{s-}^{(n,\varepsilon)})\Delta X_s^{(n,\varepsilon)}\big)\Big\} \\
&= \mathbb{E}\Big\{\int_0^t \int \Big(\rho_t^{(n,\varepsilon)}(X_{s-}^{(n,\varepsilon)} + u) - \rho_t^{(n,\varepsilon)}(X_{s-}^{(n,\varepsilon)}) - (\rho_t^{(n,\varepsilon)})'(X_{s-}^{(n,\varepsilon)})u\Big)\mu^{(n,\varepsilon)}(ds, du)\Big\} \\
&= \mathbb{E}\Big\{\int_0^t \int \Big(\rho_t^{(n,\varepsilon)}(X_{s-}^{(n,\varepsilon)} + u) - \rho_t^{(n,\varepsilon)}(X_{s-}^{(n,\varepsilon)}) - (\rho_t^{(n,\varepsilon)})'(X_{s-}^{(n,\varepsilon)})u\Big)c_\varepsilon(u)\nu_s^{(n)}(du)ds\Big\} \\
&\leq \frac{Kt}{\lambda_t^{(n,\varepsilon)}} \int_0^t \int u^2 c_\varepsilon(u) g\Big(\frac{s}{n\Delta_n}, du\Big)ds \leq \frac{Kt^2}{\lambda_t^{(n,\varepsilon)}}.
\end{aligned} \tag{6.18}$$

Note that the integrand in the second line in (6.18) is a concatenation of a Borel measurable function on $\mathbb{R}^2$ and the obviously predictable function $(\omega, r, u) \mapsto (X_{r-}^{(n,\varepsilon)}(\omega), u)$ from $\Omega \times \mathbb{R}_+ \times \mathbb{R}$ into $\mathbb{R}^2$. Consequently, this integrand is in fact a predictable function and Theorem II.1.8 in the last named reference can be applied. The first inequality in (6.18) follows with (6.15) and a Taylor expansion of the integrand. Accordingly the constant K after the last inequality in (6.18) depends only on the quantities as claimed in the assertion of this lemma. Thus we have

$$\left| \mathbb{E}\Big[\rho_t^{(n,\varepsilon)}\Big(X_t^{(n,\varepsilon)}\Big)\Big] - \frac{1}{\lambda_t^{(n,\varepsilon)}} \int\limits_0^t \nu_s^{(n)}(\mathcal{I}(z)) \right| \leq \frac{Kt^2}{\lambda_t^{(n,\varepsilon)}},$$

which together with $|1 - \exp(-\lambda_t^{(n,\varepsilon)})| \leq Kt$ for small t as well as (6.12), (6.13) and (6.14) yields the lemma. □

5.1 Proof of Theorem 3.1.

Let $X^{(n)}$ denote a semimartingale of the form (1.1) and consider the decomposition $X_t^{(n)} = Y_t^{(n)} + Z_t^{(n)}$, where

$$Y_t^{(n)} = X_0^{(n)} + \int\limits_0^t b_s^{(n)} ds + \int\limits_0^t \sigma_s^{(n)} dW_s$$

and $Z_t^{(n)}$ is a pure jump Itō semimartingale with characteristics $(0, 0, \nu_s^{(n)})$. Furthermore we consider the process

$$\mathbb{G}_n^\circ(\theta, z) = \frac{1}{\sqrt{k_n}} \sum_{j=1}^{\lfloor n\theta \rfloor} \{ \mathbb{1}_{\{\Delta_j^n Z^{(n)} \in \mathcal{I}(z)\}} - \mathbb{P}(\Delta_j^n Z^{(n)} \in \mathcal{I}(z)) \}$$

in $\ell^\infty(A_\varepsilon)$. The proof can be devided into two steps:

$$\mathbb{G}_n^\circ \rightsquigarrow \mathbb{G} \tag{6.19}$$

$$\|\mathbb{G}_n - \mathbb{G}_n^\circ\|_{A_\varepsilon} \xrightarrow{\mathbb{P}^*} 0. \tag{6.20}$$

The assertion of Theorem 3.1 then follows from Lemma 1.10.2(i) in Van der Vaart and Wellner (1996).

(6.19) can be obtained with similar steps as in the first part of the proof of Theorem 2.3 in Bücher et al. (2016) using Theorem 11.16 in Kosorok (2008) which is a central limit theorem for triangular arrays of row-wise i.i.d. data. The main difference regards the use of Lemma 5.3 which is needed as we work in general with a time-varying kernel $\nu_s^{(n)}$.

Concerning (6.20) we have for $(\theta, z) \in A_\varepsilon$

$$|\mathbb{G}_n(\theta,z) - \mathbb{G}_n^\circ(\theta,z)| \le \sqrt{k_n}\Big|U_n(\theta,z) - U_n^\circ(\theta,z)\Big| + \sqrt{k_n}\Big|\mathbb{E}U_n^\circ(\theta,z) - \int_0^\theta g(y,z)dy\Big|, \qquad (6.21)$$

where U_n° denotes the statistic U_n based on the scheme $\{Z_{i\Delta_n}^{(n)} \mid i = 0,\dots,n\}$. For the first term in (6.21) we obtain

$$\sqrt{k_n}\Big|U_n(\theta,z) - U_n^\circ(\theta,z)\Big| = o_{\mathbb{P}}(1),$$

uniformly on A_ε, with the same arguments as in the second part of the proof of Theorem 2.3 in Bücher et al. (2016). Furthermore, along the lines of the proof of Corollary 2.5 in Bücher et al. (2016), but using Lemma 5.3 instead, one can show that the second term in (6.21) is a uniform $o(1)$ on A_ε. □

5.2 Proof of Theorem 3.3.

Recall the decomposition $X_t^{(n)} = Y_t^{(n)} + Z_t^{(n)}$ in the proof of Theorem 3.1. The idea of the proof is to show the claim of Theorem 3.3 for $\hat{\mathbb{G}}_n^\circ$, the process being defined exactly as $\hat{\mathbb{G}}_n$ in (3.7) but based on the increments $\Delta_j^n Z^{(n)}$. This can be done with Theorem 3 in Kosorok (2003), because we have i.i.d. increments of the processes $Z^{(n)}$. Furthermore, by Lemma A.1 in Bücher (2011) it is then enough to prove $\|\hat{\mathbb{G}}_n - \hat{\mathbb{G}}_n^\circ\|_{A_\varepsilon} = o_{\mathbb{P}}(1)$ in order to show Theorem 3.3. For a detailed proof we refer the reader to the proof of Theorem 3.3 in Bücher et al. (2016), which follows similar lines. □

5.3 Proof of Corollary 4.1.

As the process $\mathbb{H}$ is tight, $\mathbb{H}^{(\varepsilon)}$ is also tight and Theorem 3.2 together with the continuous mapping theorem yield $\mathbb{H}_n^{(\varepsilon)} \rightsquigarrow \mathbb{H}^{(\varepsilon)}$ in $\ell^\infty([0,1])$. The assertion now follows observing the definition of $\mathbb{H}_n$ in (3.6) and the fact that $D(\zeta,\theta,z)$ vanishes whenever $\theta \le \theta_0^{(\varepsilon)}$. □

5.4 Proof of Theorem 4.2.

The claim follows if we can prove the existence of a constant $K > 0$ such that

$$\mathbb{P}\Big(\hat{\theta}_n^{(\varepsilon)} < \theta_0^{(\varepsilon)}\Big) = o(1), \qquad (6.22)$$

$$\mathbb{P}\Big(\hat{\theta}_n^{(\varepsilon)} > \theta_0^{(\varepsilon)} + K\beta_n\Big) = o(1), \qquad (6.23)$$

where $\beta_n = (\varkappa_n/\sqrt{k_n})^{1/\varpi}$. In order to verify (6.22) we calculate as follows

$$\mathbb{P}\Big(\hat{\theta}_n^{(\varepsilon)} < \theta_0^{(\varepsilon)}\Big) \le \mathbb{P}\Big(k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(\theta) > \varkappa_n \text{ for some } \theta < \theta_0^{(\varepsilon)}\Big) \qquad (6.24)$$
$$\le \mathbb{P}\Big(k_n^{1/2}\mathcal{D}^{(\varepsilon)}(\theta) + \mathbb{H}_n^{(\varepsilon)}(\theta) > \varkappa_n \text{ for some } \theta < \theta_0^{(\varepsilon)}\Big) \le \mathbb{P}\Big(\mathbb{H}_n^{(\varepsilon)}(1) > \varkappa_n\Big) = o(1),$$

where $\mathbb{H}_n^{(\varepsilon)}(1)$ is defined in (4.3). The third estimate is a consequence of the fact that $\mathcal{D}^{(\varepsilon)}(\theta) = 0$ whenever $\theta < \theta_0^{(\varepsilon)}$ and the final convergence follows because a weakly converging sequence in $(\mathbb{R}, \mathbb{B})$ is asymptotically tight.
For a proof of (6.23) we note that $k_n^{1/2}\mathcal{D}^{(\varepsilon)}(\theta) - \mathbb{H}_n^{(\varepsilon)}(\theta) \leq k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(\theta)$ and we obtain

$$\begin{aligned}\mathbb{P}\Big(\hat{\theta}_n^{(\varepsilon)} > \theta_0^{(\varepsilon)} + K\beta_n\Big) &\leq \mathbb{P}\Big(k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(\theta) \leq \varkappa_n \text{ for some } \theta > \theta_0^{(\varepsilon)} + K\beta_n\Big) \\ &\leq \mathbb{P}\Big(k_n^{1/2}\mathcal{D}^{(\varepsilon)}(\theta) - \mathbb{H}_n^{(\varepsilon)}(\theta) \leq \varkappa_n \text{ for some } \theta > \theta_0^{(\varepsilon)} + K\beta_n\Big). \end{aligned} \tag{6.25}$$

Now it follows from (4.2) that for sufficiently large $n \in \mathbb{N}$

$$\inf_{\theta \in [\theta_0^{(\varepsilon)} + K\beta_n, 1]} \mathcal{D}^{(\varepsilon)}(\theta) = \mathcal{D}^{(\varepsilon)}(\theta_0^{(\varepsilon)} + K\beta_n) \geq \frac{1}{2} c^{(\varepsilon)}(K\beta_n)^{\varpi}. \tag{6.26}$$

Therefore with (6.25) and by the definition of β_n we get for large $n \in \mathbb{N}$ and $K > 0$ large enough

$$\begin{aligned}\mathbb{P}\Big(\hat{\theta}_n^{(\varepsilon)} > \theta_0^{(\varepsilon)} + K\beta_n\Big) &\leq \mathbb{P}\Big(\frac{1}{2}\sqrt{k_n} c^{(\varepsilon)}(K\beta_n)^{\varpi} - \mathbb{H}_n^{(\varepsilon)}(1) \leq \varkappa_n\Big) \\ &\leq \mathbb{P}\Big(\frac{1}{2}\sqrt{k_n} c^{(\varepsilon)}(K\beta_n)^{\varpi} - \mathbb{H}_n^{(\varepsilon)}(1) \leq \varkappa_n, \mathbb{H}_n^{(\varepsilon)}(1) \leq \alpha_n\Big) + \mathbb{P}\Big(\mathbb{H}_n^{(\varepsilon)}(1) > \alpha_n\Big) = o(1),\end{aligned}$$

where $\alpha_n \to \infty$ is a sequence with $\alpha_n / \varkappa_n \to 0$, using asymptotic tightness again. □

5.5 Proof of Theorem 4.3.

For a proof of (4.4) note that

$$(\hat{\theta}_n^{(\varepsilon)} - \theta_0^{(\varepsilon)})^2 = \Big\{ \int_{\theta_0^{(\varepsilon)}}^1 1_{\{k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(\theta) \leq \varkappa_n\}} d\theta - \int_0^{\theta_0^{(\varepsilon)}} \Big(1 - 1_{\{k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(\theta) \leq \varkappa_n\}}\Big) d\theta \Big\}^2$$

and furthermore we have for any $\theta \in [0, 1]$

$$k_n^{1/2}\mathcal{D}^{(\varepsilon)}(\theta) - \mathbb{H}_n^{(\varepsilon)}(1) \leq k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(\theta) \leq k_n^{1/2}\mathcal{D}^{(\varepsilon)}(\theta) + \mathbb{H}_n^{(\varepsilon)}(1). \tag{6.27}$$

Thus, if $\mathbb{H}_n^{(\varepsilon)}(1) \leq \alpha_n$, we get for sufficiently large $n \in \mathbb{N}$

$$\begin{aligned} 0 &\leq \int_0^{\theta_0^{(\varepsilon)}} \Big(1 - 1_{\{k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(\theta) \leq \varkappa_n\}}\Big) d\theta \leq \int_0^{\theta_0^{(\varepsilon)}} \Big(1 - 1_{\{k_n^{1/2}\mathcal{D}^{(\varepsilon)}(\theta) + \alpha_n \leq \varkappa_n\}}\Big) d\theta \\ &= \int_0^{\theta_0^{(\varepsilon)}} \Big(1 - 1_{\{\alpha_n \leq \varkappa_n\}}\Big) d\theta = 0, \end{aligned}$$

because $\mathcal{D}^{(\varepsilon)}(\theta) = 0$ for $\theta \le \theta_0^{(\varepsilon)}$. Hence, for n sufficiently large,

$$\mathrm{MSE}_1^{(\varepsilon)}(\varkappa_n, \alpha_n) = \mathbb{E}\Big[\Big\{ \int\limits_{\theta_0^{(\varepsilon)}}^{1} 1_{\left\{k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(\theta) \le \varkappa_n\right\}} d\theta \Big\}^2 1_{\left\{\mathbb{H}_n^{(\varepsilon)}(1) \le \alpha_n\right\}}\Big]. \tag{6.28}$$

In the following let $0 < \varphi < 1$ be arbitrary, let K_1, K_2 be as in (4.5) and define

$$K_1^* := \Big(\frac{1-\varphi/2}{c^{(\varepsilon)}}\Big)^{1/\varpi} \quad \text{and} \quad K_2^* := \Big(\frac{1+\varphi}{c^{(\varepsilon)}}\Big)^{1/\varpi}. \tag{6.29}$$

As in (6.26) we obtain from (4.2)

$$\max_{\theta \in [\theta_0^{(\varepsilon)}, \theta_0^{(\varepsilon)} + K_1^*\beta_n]} \mathcal{D}^{(\varepsilon)}(\theta) = \mathcal{D}^{(\varepsilon)}(\theta_0^{(\varepsilon)} + K_1^*\beta_n) \le \frac{1}{1-\varphi/3} c^{(\varepsilon)}(K_1^*\beta_n)^{\varpi} \tag{6.30}$$

and

$$\inf_{\theta \in [\theta_0^{(\varepsilon)} + K_2^*\beta_n, 1]} \mathcal{D}^{(\varepsilon)}(\theta) = \mathcal{D}^{(\varepsilon)}(\theta_0^{(\varepsilon)} + K_2^*\beta_n) \ge \frac{1}{1+\varphi/2} c^{(\varepsilon)}(K_2^*\beta_n)^{\varpi} \tag{6.31}$$

for $n \in \mathbb{N}$ large enough. Now (6.27) and (6.28) yield

$$\begin{aligned} \mathrm{MSE}_1^{(\varepsilon)}(\varkappa_n, \alpha_n) &\le \Big[\int\limits_{\theta_0^{(\varepsilon)}}^{1} 1_{\left\{\sqrt{k_n}\mathcal{D}^{(\varepsilon)}(\theta) \le \varkappa_n + \alpha_n\right\}} d\theta \Big]^2 \\ &= \Big[\int\limits_{\theta_0^{(\varepsilon)}}^{\theta_0^{(\varepsilon)} + K_2^*\beta_n} 1_{\left\{\sqrt{k_n}\mathcal{D}^{(\varepsilon)}(\theta) \le \varkappa_n + \alpha_n\right\}} d\theta \Big]^2 \le (K_2^*)^2 \beta_n^2 = K_2 \beta_n^2 \end{aligned} \tag{6.32}$$

for a sufficiently large $n \in \mathbb{N}$ which is the desired bound. Note that the first equation in the second line of (6.32) follows from (6.31), because for $\theta \in [\theta_0^{(\varepsilon)} + K_2^*\beta_n, 1]$ we have

$$\sqrt{k_n}\mathcal{D}^{(\varepsilon)}(\theta) \le \varkappa_n + \alpha_n \Longrightarrow \frac{1}{1+\varphi/2} c^{(\varepsilon)}(K_2^*)^{\varpi} \varkappa_n \le \varkappa_n + \alpha_n$$

which cannot hold for large $n \in \mathbb{N}$ due to (6.29).
In order to get a lower bound recall (6.28) and use (6.27) to obtain for $n \in \mathbb{N}$ sufficiently large

$$\begin{aligned} \mathrm{MSE}_1^{(\varepsilon)}(\varkappa_n, \alpha_n) &\ge \mathbb{P}\big(\mathbb{H}_n^{(\varepsilon)}(1) \le \alpha_n\big) \Big(\int\limits_{\theta_0^{(\varepsilon)}}^{\theta_0^{(\varepsilon)} + K_1^*\beta_n} 1_{\left\{\sqrt{k_n}\mathcal{D}^{(\varepsilon)}(\theta) \le \varkappa_n - \alpha_n\right\}} d\theta \Big)^2 \\ &= \mathbb{P}\big(\mathbb{H}_n^{(\varepsilon)}(1) \le \alpha_n\big)(K_1^*)^2 \beta_n^2, \end{aligned} \tag{6.33}$$

where the equality follows from the implication (see (6.30))

$$\frac{1}{1-\varphi/3}c^{(\varepsilon)}(K_1^*)^{\varpi}\varkappa_n \leq \varkappa_n - \alpha_n \Longrightarrow \sqrt{k_n}\mathcal{D}^{(\varepsilon)}(\theta) \leq \varkappa_n - \alpha_n \quad \text{for all } \theta \in [\theta_0^{(\varepsilon)}, \theta_0^{(\varepsilon)} + K_1^*\beta_n].$$

The left-hand side in the previous display always holds for large $n \in \mathbb{N}$ by the choice of K_1^* in (6.29). Using asymptotical tightness we also have

$$\mathbb{P}\big(\mathbb{H}_n^{(\varepsilon)}(1) \leq \alpha_n\big) \geq \Big(\frac{1-\varphi}{1-\varphi/2}\Big)^{2/\varpi} = K_1/(K_1^*)^2$$

for a large n, which together with (6.33) yields $\mathrm{MSE}_1^{(\varepsilon)}(\varkappa_n, \alpha_n) \geq K_1\beta_n^2$. □

5.6 Proof of Theorem 4.4.

Similarly to (6.24) we get

$$\mathbb{P}\Big(\hat{\theta}_n^{(\varepsilon)}(\hat{\varkappa}_{n,B}^{(\varepsilon,\alpha)}(1)) < \theta_0^{(\varepsilon)}\Big) \leq \mathbb{P}\Big(\mathbb{H}_n^{(\varepsilon)}(\theta_0^{(\varepsilon)}) \geq \hat{\varkappa}_{n,B}^{(\varepsilon,\alpha)}(1)\Big). \tag{6.34}$$

Recall $\mathbb{H}^{(\varepsilon)}(\theta)$ from (4.1). It holds that $\mathbb{H}^{(\varepsilon)}(\theta_0^{(\varepsilon)}) \geq |\mathbb{H}(\theta_0^{(\varepsilon)}/2, \theta_0^{(\varepsilon)}, \bar{z})|$, with $\bar{z}$ from (4.8), and by (3.5) we have

$$\mathrm{Var}(\mathbb{H}(\theta_0^{(\varepsilon)}/2, \theta_0^{(\varepsilon)}, \bar{z})) = \frac{1}{4}\int_0^{\theta_0^{(\varepsilon)}} g(y, \bar{z})dy > 0. \tag{6.35}$$

Thus $\mathbb{H}^{(\varepsilon)}(\theta_0^{(\varepsilon)})$ is a supremum of a non-vanishing Gaussian process with mean zero. Due to Corollary 1.3 and Remark 4.1 in Gaenssler et al. (2007) $\mathbb{H}^{(\varepsilon)}(\theta_0^{(\varepsilon)})$ then has a continuous distribution function. As a consequence (4.9) follows from (6.34) and Proposition F.1 in the supplement to Bücher and Kojadinovic (2016) as soon as we can show

$$\Big(\mathbb{H}_n^{(\varepsilon)}(\theta_0^{(\varepsilon)}), \hat{\mathbb{H}}_{n,\xi^{(1)}}^{(\varepsilon)}(\hat{\theta}_n), \ldots, \hat{\mathbb{H}}_{n,\xi^{(B)}}^{(\varepsilon)}(\hat{\theta}_n)\Big) \rightsquigarrow \Big(\mathbb{H}^{(\varepsilon)}(\theta_0^{(\varepsilon)}), \mathbb{H}_{(1)}^{(\varepsilon)}(\theta_0^{(\varepsilon)}), \ldots, \mathbb{H}_{(B)}^{(\varepsilon)}(\theta_0^{(\varepsilon)})\Big) \tag{6.36}$$

for any fixed $B \in \mathbb{N}$, where $\mathbb{H}_{(1)}^{(\varepsilon)}(\theta_0^{(\varepsilon)}), \ldots, \mathbb{H}_{(B)}^{(\varepsilon)}(\theta_0^{(\varepsilon)})$ are independent copies of $\mathbb{H}^{(\varepsilon)}(\theta_0^{(\varepsilon)})$.
In order to establish (6.36) we first show that the sample paths of $\mathbb{H}^{(\varepsilon)}$ are uniformly continuous on $[0,1]$ with respect to the Euclidean distance. By Theorem 3.1 and Assumption 2.1 the sample paths of the process $\mathbb{G}$ in $\ell^\infty(A_\varepsilon)$ satisfy $\mathbb{G}(0,z) = 0$ for all $z \in M_\varepsilon$ and they are uniformly continuous with respect to the Euclidean distance on A_ε. Thus the uniform continuity of the sample paths of $\mathbb{H}^{(\varepsilon)}$ holds if we can show that for a bounded and uniformly continuous function $f\colon A_\varepsilon \to \mathbb{R}$ with $f(0,z) = 0$ for all $z \in M_\varepsilon$ the function $H\colon [0,1] \to \mathbb{R}$ defined via

$$H(\theta) := \sup_{\theta' \in [0,\theta]} \sup_{\zeta \in [0,\theta']} \sup_{z \in M_\varepsilon} |f(\zeta, z) - \frac{\zeta}{\theta'} f(\theta', z)|$$

is uniformly continuous on $[0,1]$. But since a continuous function on a compact metric space is uniformly continuous it suffices to show continuity of the function

$$F(\theta) := \sup_{\zeta \in [0,\theta]} \sup_{z \in M_\varepsilon} |f(\zeta, z) - \frac{\zeta}{\theta} f(\theta, z)|$$

in every $\theta_0 \in [0,1]$. The continuity of F in $\theta_0 = 0$ is obvious by the property $f(0,z) = 0$ for all $z \in M_\varepsilon$ of the function f and therefore only the case $0 < \theta_0 \leq 1$ remains. Let U be a neighbourhood of θ_0 in $[0,1]$ which is bounded away from 0. Then it is immediate to see that the function $h\colon B_\varepsilon \to \mathbb{R}$ defined by

$$h(\zeta, \theta, z) := f(\zeta, z) - \frac{\zeta}{\theta} f(\theta, z)$$

is uniformly continuous on $B_\varepsilon \cap ([0,1] \times U \times M_\varepsilon)$.
Let $\eta > 0$ be arbitrary and choose $\delta > 0$ such that $|h(\zeta_1, \theta_1, z_1) - h(\zeta_2, \theta_2, z_2)| < \eta/2$ for all $(\zeta_1, \theta_1, z_1), (\zeta_2, \theta_2, z_2) \in B_\varepsilon$ with maximum distance $\|(\zeta_1, \theta_1, z_1)^T - (\zeta_2, \theta_2, z_2)^T\|_\infty \leq \delta$. Furthermore, let $\theta \in [0,1]$ with $|\theta - \theta_0| < \delta$. Then there exists $(\zeta_1, \theta_0, z_1) \in B_\varepsilon$ with $F(\theta_0) - \eta < |h(\zeta_1, \theta_0, z_1)| - \eta/2$ and we can choose a $\zeta_2 \leq \theta$ such that $\|(\zeta_1, \theta_0, z_1)^T - (\zeta_2, \theta, z_1)^T\|_\infty \leq \delta$ which gives $F(\theta_0) - \eta < |h(\zeta_2, \theta, z_1)| \leq F(\theta)$. In an analogous manner we see that also $F(\theta) - \eta < F(\theta_0)$ for each $\theta \in [0,1]$ with $|\theta - \theta_0| < \delta$ and therefore F is continuous in θ_0.
For arbitrary $\eta > 0$ we first want to show

$$\mathbb{P}\Big(\Big\| \Big(\mathbb{H}_n^{(\varepsilon)}(\theta_0^{(\varepsilon)}), \hat{\mathbb{H}}_{n,\xi^{(1)}}^{(\varepsilon)}(\hat{\theta}_n), \ldots, \hat{\mathbb{H}}_{n,\xi^{(B)}}^{(\varepsilon)}(\hat{\theta}_n)\Big)^T - \\ - \Big(\mathbb{H}_n^{(\varepsilon)}(\theta_0^{(\varepsilon)}), \hat{\mathbb{H}}_{n,\xi^{(1)}}^{(\varepsilon)}(\theta_0^{(\varepsilon)}), \ldots, \hat{\mathbb{H}}_{n,\xi^{(B)}}^{(\varepsilon)}(\theta_0^{(\varepsilon)})\Big)^T \Big\|_\infty > \eta\Big) \to 0. \quad (6.37)$$

By Proposition 10.7 in Kosorok (2008) and Theorem 3.4 we have $\hat{\mathbb{H}}_{n,\xi^{(i)}}^{(\varepsilon)} \rightsquigarrow_\xi \mathbb{H}^{(\varepsilon)}$ in $\ell^\infty([0,1])$ for all $i = 1, \ldots, B$, which yields $\hat{\mathbb{H}}_{n,\xi^{(i)}}^{(\varepsilon)} \rightsquigarrow \mathbb{H}^{(\varepsilon)}$ for all $i = 1, \ldots, B$ with the same reasoning as in the proof of Theorem 2.9.6 in Van der Vaart and Wellner (1996). Theorem 1.5.7 and its addendum therein show that $\hat{\mathbb{H}}_{n,\xi^{(i)}}^{(\varepsilon)}$ is asymptotically uniformly ρ-equicontinuous in probability for each i, where ρ denotes the Euclidean metric on $[0,1]$ because the sample paths of $\mathbb{H}^{(\varepsilon)}$ are uniformly continuous with respect to ρ and $([0,1], \rho)$ is totally bounded.
Therefore, for any $\gamma > 0$ we can choose a $\delta > 0$ such that

$$\max_{i=1,\ldots,B} \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{\rho(\theta_1,\theta_2)<\delta} \Big|\hat{\mathbb{H}}_{n,\xi^{(i)}}^{(\varepsilon)}(\theta_1) - \hat{\mathbb{H}}_{n,\xi^{(i)}}^{(\varepsilon)}(\theta_2)\Big| > \eta\Big) < \gamma/(2B),$$

which yields

$$\mathbb{P}\Big(\Big\| \Big(\mathbb{H}_n^{(\varepsilon)}(\theta_0^{(\varepsilon)}), \hat{\mathbb{H}}_{n,\xi^{(1)}}^{(\varepsilon)}(\hat{\theta}_n), \ldots, \hat{\mathbb{H}}_{n,\xi^{(B)}}^{(\varepsilon)}(\hat{\theta}_n)\Big)^T - \\ - \Big(\mathbb{H}_n^{(\varepsilon)}(\theta_0^{(\varepsilon)}), \hat{\mathbb{H}}_{n,\xi^{(1)}}^{(\varepsilon)}(\theta_0^{(\varepsilon)}), \ldots, \hat{\mathbb{H}}_{n,\xi^{(B)}}^{(\varepsilon)}(\theta_0^{(\varepsilon)})\Big)^T \Big\|_\infty > \eta\Big) \\ \leq \mathbb{P}\Big(\Big|\hat{\mathbb{H}}_{n,\xi^{(i)}}^{(\varepsilon)}(\hat{\theta}_n) - \hat{\mathbb{H}}_{n,\xi^{(i)}}^{(\varepsilon)}(\theta_0^{(\varepsilon)})\Big| > \eta \text{ for at least one } i = 1, \ldots, B \text{ and } |\hat{\theta}_n - \theta_0^{(\varepsilon)}| < \delta\Big) +$$

$$+\mathbb{P}\Big(|\hat\theta_n-\theta_0^{(\varepsilon)}|\geq\delta\Big)$$

$$\leq\mathbb{P}\Big(|\hat\theta_n-\theta_0^{(\varepsilon)}|\geq\delta\Big)+\sum_{i=1}^{B}\mathbb{P}\Big(\sup_{\rho(\theta_1,\theta_2)<\delta}\Big|\hat{\mathbb{H}}^{(\varepsilon)}_{n,\xi^{(i)}}(\theta_1)-\hat{\mathbb{H}}^{(\varepsilon)}_{n,\xi^{(i)}}(\theta_2)\Big|>\eta\Big)<\gamma$$

for $n\in\mathbb{N}$ large enough, using consistency of the preliminary estimator.
Thus, now that we have established (6.37), by Lemma 1.10.2(i) in Van der Vaart and Wellner (1996) we obtain (6.36) if we can show

$$\Big(\mathbb{H}^{(\varepsilon)}_n(\theta_0^{(\varepsilon)}),\hat{\mathbb{H}}^{(\varepsilon)}_{n,\xi^{(1)}}(\theta_0^{(\varepsilon)}),\ldots,\hat{\mathbb{H}}^{(\varepsilon)}_{n,\xi^{(B)}}(\theta_0^{(\varepsilon)})\Big)\rightsquigarrow\Big(\mathbb{H}^{(\varepsilon)}(\theta_0^{(\varepsilon)}),\mathbb{H}^{(\varepsilon)}_{(1)}(\theta_0^{(\varepsilon)}),\ldots,\mathbb{H}^{(\varepsilon)}_{(B)}(\theta_0^{(\varepsilon)})\Big).$$

But this is an immediate consequence of the continuous mapping theorem and

$$(\mathbb{G}_n,\hat{\mathbb{G}}_{n,\xi^{(1)}},\ldots,\hat{\mathbb{G}}_{n,\xi^{(B)}})\rightsquigarrow(\mathbb{G},\mathbb{G}^{(1)},\ldots,\mathbb{G}^{(B)}) \tag{6.38}$$

in $(\ell^\infty(A_\varepsilon))^{B+1}$ for all $B\in\mathbb{N}$, where $\mathbb{G}^{(1)},\ldots,\mathbb{G}^{(B)}$ are independent copies of $\mathbb{G}$, since $\mathbb{H}^{(\varepsilon)}_n(\theta_0^{(\varepsilon)})$, $\hat{\mathbb{H}}^{(\varepsilon)}_{n,\xi^{(i)}}(\theta_0^{(\varepsilon)})$ and $\mathbb{H}^{(\varepsilon)}(\theta_0^{(\varepsilon)})$ are the images of the same continuous functional applied to $\mathbb{G}_n$, $\hat{\mathbb{G}}_{n,\xi^{(i)}}$ and $\mathbb{G}$, respectively. (6.38) follows as in Proposition 6.2 in Bücher et al. (2016). □

5.7 Proof of Theorem 4.5.

We start with a proof of $\beta_n\xrightarrow{\mathbb{P}}0$ which is equivalent to $\hat\varkappa^{(\varepsilon,\alpha_n)}_{n,B_n}(r)/\sqrt{k_n}\xrightarrow{\mathbb{P}}0$. Therefore we have to show

$$\mathbb{P}(\hat\varkappa^{(\varepsilon,\alpha_n)}_{n,B_n}(r)/\sqrt{k_n}\leq x)=\mathbb{P}\Big(\frac{1}{B_n}\sum_{i=1}^{B_n}\mathbb{1}_{\{\hat{\mathbb{H}}^{(\varepsilon)}_{n,\xi^{(i)}}(\hat\theta_n)\leq(\sqrt{k_n}x)^{1/r}\}}\geq1-\alpha_n\Big)\to1, \tag{6.39}$$

for arbitrary $x>0$, by the definition of $\hat\varkappa^{(\varepsilon,\alpha_n)}_{n,B_n}(r)$ in (4.7). Since the

$$\mathbb{1}_{\{\hat{\mathbb{H}}^{(\varepsilon)}_{n,\xi^{(i)}}(\hat\theta_n)\leq(\sqrt{k_n}x)^{1/r}\}}-\mathbb{P}_\xi\Big(\hat{\mathbb{H}}^{(\varepsilon)}_n(\hat\theta_n)\leq(\sqrt{k_n}x)^{1/r}\Big),\quad i=1,\ldots,B_n,$$

are pairwise uncorrelated with mean zero and bounded by 1, we have

$$\mathbb{P}\Big(\Big|\frac{1}{B_n}\sum_{i=1}^{B_n}\mathbb{1}_{\{\hat{\mathbb{H}}^{(\varepsilon)}_{n,\xi^{(i)}}(\hat\theta_n)\leq(\sqrt{k_n}x)^{1/r}\}}-\mathbb{P}_\xi\Big(\hat{\mathbb{H}}^{(\varepsilon)}_n(\hat\theta_n)\leq(\sqrt{k_n}x)^{1/r}\Big)\Big|>\alpha_n/2\Big)\leq4\alpha_n^{-2}B_n^{-1}\to0. \tag{6.40}$$

Therefore, in order to prove (6.39), it suffices to verify

$$\mathbb{P}\Big(\mathbb{P}_\xi\Big(\hat{\mathbb{H}}^{(\varepsilon)}_n(\hat\theta_n)\leq(\sqrt{k_n}x)^{1/r}\Big)<1-\alpha_n/2\Big)\leq\frac{2}{\alpha_n}\mathbb{P}\Big(\hat{\mathbb{H}}^{(\varepsilon)}_n(\hat\theta_n)>(\sqrt{k_n}x)^{1/r}\Big)$$
$$\leq\frac{2}{\alpha_n}\mathbb{P}\Big(2\sup_{\theta\in[0,1]}\sup_{|z|\geq\varepsilon}|\hat{\mathbb{G}}_n(\theta,z)|>(\sqrt{k_n}x)^{1/r}\Big)\to0, \tag{6.41}$$

where the first inequality in the above display follows with the Markov inequality and the last inequality in (6.41) is a consequence of the fact that $\hat{\mathbb{H}}_n^{(\varepsilon)}(\hat{\theta}_n) \le \hat{\mathbb{H}}_n^{(\varepsilon)}(1) \le 2 \sup_{\theta\in[0,1]} \sup_{|z|\ge\varepsilon} |\hat{\mathbb{G}}_n(\theta,z)|$. Furthermore, by the definition of $\hat{\mathbb{G}}_n$ in (3.7) we have

$$\mathbb{E}\Big\{\sup_{\theta\in[0,1]} \sup_{|z|\ge\varepsilon} |\hat{\mathbb{G}}_n(\theta,z)|\Big\} \le \frac{1}{n\sqrt{k_n}} \sum_{j=1}^n \sum_{i=1}^n (\mathbb{P}(|\Delta_j^n X^{(n)}| \ge \varepsilon) + \mathbb{P}(|\Delta_i^n X^{(n)}| \ge \varepsilon)), \tag{6.42}$$

because of $\mathbb{E}|\xi_j| \le 1$ for every $j = 1,\dots,n$. Recall the decomposition $X^{(n)} = Y^{(n)} + Z^{(n)}$ in the proof of Theorem 3.1 and let $v_n = \Delta_n^{\tau/2} \to 0$ with τ from Assumption 2.4. Then we have for $i = 1,\dots,n$ and $n \in \mathbb{N}$ large enough

$$\mathbb{P}(|\Delta_i^n X^{(n)}| \ge \varepsilon) \le \mathbb{P}(|\Delta_i^n Y^{(n)}| \ge v_n) + \mathbb{P}(|\Delta_i^n Z^{(n)}| \ge \varepsilon/2) \le \mathbb{P}(|\Delta_i^n Y^{(n)}| \ge v_n) + K\Delta_n, \tag{6.43}$$

where the last inequality follows using Lemma 5.3. By Hölder's inequality, the Burkholder-Davis-Gundy inequalities (see for instance page 39 in Jacod and Protter, 2012) and Fubini's theorem we have with $p > 2$, $1 < \alpha < 3$ from Assumption 2.4, for each $1 \le j \le n$,

$$\mathbb{E}\Big|\int_{(j-1)\Delta_n}^{j\Delta_n} b_s^{(n)} ds\Big|^\alpha \le \Delta_n^\alpha \mathbb{E}\Big(\frac{1}{\Delta_n}\int_{(j-1)\Delta_n}^{j\Delta_n} |b_s^{(n)}|^\alpha ds\Big) \le K\Delta_n^\alpha \tag{6.44}$$

and

$$\begin{aligned}\mathbb{E}\Big|\int_{(j-1)\Delta_n}^{j\Delta_n} \sigma_s^{(n)} dW_s\Big|^p &\le K\Delta_n^{p/2}\mathbb{E}\Big(\frac{1}{\Delta_n}\int_{(j-1)\Delta_n}^{j\Delta_n} |\sigma_s^{(n)}|^2 ds\Big)^{p/2} \\ &\le K\Delta_n^{p/2}\mathbb{E}\Big(\frac{1}{\Delta_n}\int_{(j-1)\Delta_n}^{j\Delta_n} |\sigma_s^{(n)}|^p ds\Big) \le K\Delta_n^{p/2}. \end{aligned}\tag{6.45}$$

Together with (6.43) and the Markov inequality these estimates yield

$$\mathbb{P}(|\Delta_i^n X^{(n)}| \ge \varepsilon) \le K\Delta_n^{p/2-p\tau/2} + K\Delta_n^{\alpha-\alpha\tau/2} + K\Delta_n = K\Delta_n^{\frac{2p+p}{2p+2}} + K\Delta_n \le K\Delta_n. \tag{6.46}$$

Therefore due to (6.41), (6.42), (6.46) and the Markov inequality we obtain

$$\mathbb{P}\Big(\mathbb{P}_\xi\Big(\hat{\mathbb{H}}_n^{(\varepsilon)}(\hat{\theta}_n) \le (\sqrt{k_n}x)^{1/r}\Big) < 1-\alpha_n/2\Big) \le K\frac{n^2\Delta_n}{\alpha_n n\sqrt{k_n}(\sqrt{k_n})^{1/r}} = K\Big((n\Delta_n)^{\frac{1-r}{2r}}\alpha_n\Big)^{-1} \to 0,$$

by the assumptions on the involved sequences. Thus we conclude $\beta_n \xrightarrow{\mathbb{P}} 0$.
Next we show $\hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r) \xrightarrow{\mathbb{P}} \infty$, which is equivalent to

$$\mathbb{P}(\hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r) \le x) = \mathbb{P}\Big(\frac{1}{B_n}\sum_{i=1}^{B_n} \mathbb{1}_{\{\hat{\mathbb{H}}_{n,\xi^{(i)}}^{(\varepsilon)}(\hat{\theta}_n) \le x^{1/r}\}} \ge 1-\alpha_n\Big) \to 0,$$

for each $x > 0$. With the same considerations as for (6.40) it is sufficient to show

$$\mathbb{P}\Big(\mathbb{P}_\xi\Big(\hat{\mathbb{H}}_n^{(\varepsilon)}(\hat{\theta}_n) > x^{1/r}\Big) \le 2\alpha_n\Big) \to 0.$$

By continuity of the function $\zeta \mapsto \int_0^\zeta g(y,\bar{z})dy$ for $\bar{z}$ from (4.8) we can find $\bar{\zeta} < \bar{\theta} < \theta_0^{(\varepsilon)}$ with

$$\int_0^{\bar{\zeta}} g(y,\bar{z})dy > 0 \tag{6.47}$$

and because of

$$\hat{\mathbb{H}}_n(\bar{\zeta},\bar{\theta},\bar{z}) \le \hat{\mathbb{H}}_n^{(\varepsilon)}(\hat{\theta}_n) \Longrightarrow \mathbb{P}_\xi\Big(\hat{\mathbb{H}}_n(\bar{\zeta},\bar{\theta},\bar{z}) > x^{1/r}\Big) \le \mathbb{P}_\xi\Big(\hat{\mathbb{H}}_n^{(\varepsilon)}(\hat{\theta}_n) > x^{1/r}\Big)$$

on the set $\{\bar{\theta} < \hat{\theta}_n\}$ and the consistency of the preliminary estimate it further suffices to prove

$$\mathbb{P}\Big(\mathbb{P}_\xi\Big(\hat{\mathbb{H}}_n^{(\varepsilon)}(\hat{\theta}_n) > x^{1/r}\Big) \le 2\alpha_n \text{ and } \bar{\theta} < \hat{\theta}_n\Big) \le \mathbb{P}\Big(\mathbb{P}_\xi\Big(\hat{\mathbb{H}}_n(\bar{\zeta},\bar{\theta},\bar{z}) > x^{1/r}\Big) \le 2\alpha_n\Big) \to 0. \tag{6.48}$$

In order to show (6.48) we want to use a Berry-Esseen type result. Recall

$$\hat{\mathbb{H}}_n(\bar{\zeta},\bar{\theta},\bar{z}) = \frac{1}{\sqrt{n\Delta_n}} \sum_{j=1}^n B_j \xi_j$$

from (3.8) with $B_j = \Big(\mathbb{1}_{\{j \le \lfloor n\bar{\zeta}\rfloor\}} - \frac{\bar{\zeta}}{\bar{\theta}}\mathbb{1}_{\{j \le \lfloor n\bar{\theta}\rfloor\}}\Big)A_j$, where

$$A_j = \mathbb{1}_{\{\Delta_j^n X^{(n)} \in \mathcal{I}(\bar{z})\}} - \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\Delta_i^n X^{(n)} \in \mathcal{I}(\bar{z})\}}.$$

By the assumptions on the multiplier sequence it is immediate to see that

$$\bar{W}_n^2 := \mathbb{E}_\xi(\hat{\mathbb{H}}_n(\bar{\zeta},\bar{\theta},\bar{z}))^2 = \frac{1}{n\Delta_n}\sum_{j=1}^n B_j^2.$$

Thus Theorem 2.1 in Chen and Shao (2001) yields

$$\sup_{x\in\mathbb{R}} \Big|\mathbb{P}_\xi\Big(\hat{\mathbb{H}}_n(\bar{\zeta},\bar{\theta},\bar{z}) > x\Big) - (1-\Phi(x/\bar{W}_n))\Big| \le K\Big\{\sum_{i=1}^n \mathbb{E}_\xi U_i^2 \mathbb{1}_{\{|U_i|>1\}} + \sum_{i=1}^n \mathbb{E}_\xi |U_i|^3 \mathbb{1}_{\{|U_i|\le 1\}}\Big\}, \tag{6.49}$$

with $U_i = \frac{B_i\xi_i}{\sqrt{n\Delta_n}\bar{W}_n}$ and where Φ denotes the standard normal distribution function. Before we proceed further in the proof of (6.48), we first show

$$\frac{1}{\bar{W}_n^2} = \frac{n\Delta_n}{\sum_{j=1}^n B_j^2} = O_{\mathbb{P}}(1), \tag{6.50}$$

which is

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big(n\Delta_n > M\sum_{j=1}^n B_j^2\Big) = 0.$$

Let $M > 0$. Then a straightforward calculation gives

$$\begin{aligned}
\mathbb{P}\Big(n\Delta_n > M\sum_{j=1}^n B_j^2\Big) &\le \mathbb{P}\Big(n\Delta_n > M'\sum_{j=1}^{\lfloor n\bar\zeta\rfloor} A_j^2\Big) \\
&= \mathbb{P}\Big(n\Delta_n > M'\frac{1}{n^2}\sum_{j=1}^{\lfloor n\bar\zeta\rfloor}\sum_{i=1}^n\sum_{k=1}^n \Big(\mathbb{1}_{\{\Delta_j^n X^{(n)}\in\mathcal{I}(\bar z)\}} + \mathbb{1}_{\{\Delta_i^n X^{(n)}\in\mathcal{I}(\bar z)\}}\mathbb{1}_{\{\Delta_k^n X^{(n)}\in\mathcal{I}(\bar z)\}} \\
&\qquad - \mathbb{1}_{\{\Delta_i^n X^{(n)}\in\mathcal{I}(\bar z)\}}\mathbb{1}_{\{\Delta_j^n X^{(n)}\in\mathcal{I}(\bar z)\}} - \mathbb{1}_{\{\Delta_j^n X^{(n)}\in\mathcal{I}(\bar z)\}}\mathbb{1}_{\{\Delta_k^n X^{(n)}\in\mathcal{I}(\bar z)\}}\Big)\Big),
\end{aligned} \tag{6.51}$$

with $M' = M(1-\bar\zeta/\bar\theta)^2$. Now consider again the decomposition $X^{(n)} = Y^{(n)} + Z^{(n)}$ of the underlying Itō semimartingale as in the proof of Theorem 3.1 and the sequence $v_n = \Delta_n^{\tau/2} \to 0$ with τ from Assumption 2.4. With (6.44), (6.45) and Lemma 5.3 it is immediate to see that for $1 \le i, k \le n$

$$\mathbb{E}\Big|\mathbb{1}_{\{\Delta_i^n X^{(n)}\in\mathcal{I}(\bar z)\}}\mathbb{1}_{\{\Delta_k^n X^{(n)}\in\mathcal{I}(\bar z)\}} - \mathbb{1}_{\{\Delta_i^n Z^{(n)}\in\mathcal{I}(\bar z)\}}\mathbb{1}_{\{\Delta_k^n Z^{(n)}\in\mathcal{I}(\bar z)\}}\Big| = o(\Delta_n).$$

Setting

$$\begin{aligned}
D_n := \frac{1}{n^3\Delta_n}\sum_{j=1}^{\lfloor n\bar\zeta\rfloor}\sum_{i=1}^n\sum_{k=1}^n \Big(&\mathbb{1}_{\{\Delta_i^n X^{(n)}\in\mathcal{I}(\bar z)\}}\mathbb{1}_{\{\Delta_k^n X^{(n)}\in\mathcal{I}(\bar z)\}} - \mathbb{1}_{\{\Delta_i^n X^{(n)}\in\mathcal{I}(\bar z)\}}\mathbb{1}_{\{\Delta_j^n X^{(n)}\in\mathcal{I}(\bar z)\}} - \\
&- \mathbb{1}_{\{\Delta_j^n X^{(n)}\in\mathcal{I}(\bar z)\}}\mathbb{1}_{\{\Delta_k^n X^{(n)}\in\mathcal{I}(\bar z)\}}\Big)
\end{aligned}$$

it is easy to deduce $\mathbb{E}|D_n| = o(1)$, using Lemma 5.3 again as well as independence of the increments of $Z^{(n)}$. Combining this result with (6.51) we have

$$\begin{aligned}
&\mathbb{P}\Big(n\Delta_n > M\sum_{j=1}^n B_j^2\Big) \le \\
&\qquad \le \mathbb{P}(|D_n| > 1/M') + \mathbb{P}\Big(1/M' > D_n + \frac{1}{n\Delta_n}\sum_{j=1}^{\lfloor n\bar\zeta\rfloor}\mathbb{1}_{\{\Delta_j^n X^{(n)}\in\mathcal{I}(\bar z)\}} \text{ and } |D_n| \le 1/M'\Big) \\
&\qquad \le \mathbb{P}(|D_n| > 1/M') + \mathbb{P}\Big(2/M' > \frac{1}{n\Delta_n}\sum_{j=1}^{\lfloor n\bar\zeta\rfloor}\mathbb{1}_{\{\Delta_j^n X^{(n)}\in\mathcal{I}(\bar z)\}}\Big)
\end{aligned}$$

for $M > 0$. Thus with (6.47) we obtain (6.50), because by Theorem 3.1 we have

$$\frac{1}{n\Delta_n}\sum_{j=1}^{\lfloor n\bar{\zeta}\rfloor} \mathbb{1}_{\{\Delta_j^n X^{(n)}\in\mathcal{I}(\bar{z})\}} = \int_0^{\bar{\zeta}} g(y,\bar{z})dy + o_{\mathbb{P}}(1).$$

Recall that our main objective is to show (6.48) and thus we consider the Berry-Esseen bound on the right-hand side of (6.49). For the first summand we distinguish two cases according to the assumptions on the multiplier sequence.

Let us discuss the case of bounded multipliers first. For $M > 0$ we have

$$|U_i| \le \frac{\sqrt{M}K}{\sqrt{n\Delta_n}}$$

for all $i = 1,\dots,n$ on the set $\{1/\bar{W}_n^2 \le M\}$, since $|B_i|$ is bounded by 1. As a consequence

$$\sum_{i=1}^n \mathbb{E}_\xi U_i^2 \mathbb{1}_{\{|U_i|>1\}} = 0 \tag{6.52}$$

for large $n\in\mathbb{N}$ on the set $\{1/\bar{W}_n^2 \le M\}$.

In the situation of normal multipliers, recall that there exist constants $K_1, K_2 > 0$ such that for $\xi \sim \mathcal{N}(0,1)$ and $y > 0$ large enough we have

$$\mathbb{E}_\xi \xi^2 \mathbb{1}_{\{|\xi|>y\}} = \frac{2}{\sqrt{2\pi}}\int_y^\infty z^2 e^{-z^2/2}dz \le K\mathbb{P}(\mathcal{N}(0,2)>y) \le K_1\exp(-K_2y^2). \tag{6.53}$$

Thus we can calculate for $n\in\mathbb{N}$ large enough on the set $\{1/\bar{W}_n^2 \le M\}$

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}_\xi U_i^2 \mathbb{1}_{\{|U_i|>1\}} &= \sum_{i=1}^n \Big(\sum_{j=1}^n B_j^2\Big)^{-1} B_i^2 \mathbb{E}_\xi \xi_i^2 \mathbb{1}_{\{|\xi_i|>(\sum_{j=1}^n B_j^2)^{1/2}/|B_i|\}} \\ &\le \sum_{i=1}^n \Big(\sum_{j=1}^n B_j^2\Big)^{-1} \mathbb{E}_\xi \xi_i^2 \mathbb{1}_{\{|\xi_i|>(\sum_{j=1}^n B_j^2)^{1/2}\}} \\ &\le \frac{M}{n\Delta_n}\sum_{i=1}^n \mathbb{E}_\xi \xi_i^2 \mathbb{1}_{\{|\xi_i|>(n\Delta_n/M)^{1/2}\}} \le \frac{K_1}{\Delta_n}\exp(-K_2 n\Delta_n),\end{aligned}$$

where K_1 and K_2 depend on M. The first inequality in the above display uses $|B_i| \le 1$ again and the last one follows with (6.53). Now let $\rho > 0$ with $n\Delta_n^{1+\rho}\to\infty$ and define $\bar{p} := 1/\rho$. Then, for $n \ge N(M)\in\mathbb{N}$ on the set $\{1/\bar{W}_n^2 \le M\}$, using $\exp(-K_2 n\Delta_n) \le (n\Delta_n)^{-\bar{p}}$, we conclude

$$\sum_{i=1}^n \mathbb{E}_\xi U_i^2 \mathbb{1}_{\{|U_i|>1\}} \le K_1\Delta_n^{-1}(n\Delta_n)^{-\bar{p}} = K_1(n\Delta_n^{1+\rho})^{-\bar{p}}. \tag{6.54}$$

We now consider the second term on the right-hand side of (6.49), for which

$$\sum_{i=1}^{n} \mathbb{E}_\xi |U_i|^3 \mathbb{1}_{\{|U_i|\le 1\}} \le \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} B_j^2 \Big)^{-3/2} |B_i|^3 \mathbb{E}_\xi |\xi_i|^3 \le \frac{K}{(n\Delta_n)^{3/2}} \sum_{i=1}^{n} |B_i|$$

holds on $\{1/\bar{W}_n^2 \le M\}$, using $|B_i| \le 1$ again. With (6.46) we see that

$$\mathbb{E}\Big(\sum_{i=1}^{n} |B_i| \Big) \le \mathbb{E}\Big(\sum_{i=1}^{n} |A_i| \Big) \le 2n \max_{i=1,\dots,n} \mathbb{P}(|\Delta_i^n X^{(n)}| \ge \varepsilon) \le Kn\Delta_n.$$

Consequently,

$$\begin{aligned}\mathbb{P}\Big(1/\bar{W}_n^2 \le M \text{ and } K \sum_{i=1}^{n} \mathbb{E}_\xi |U_i|^3 \mathbb{1}_{\{|U_i|\le 1\}} > (n\Delta_n)^{-1/4}\Big) &\le \mathbb{P}\Big(\frac{K}{(n\Delta_n)^{3/2}} \sum_{i=1}^{n} |B_i| > (n\Delta_n)^{-1/4} \Big) \\ &\le K(n\Delta_n)^{-1/4} \qquad (6.55)\end{aligned}$$

follows. Thus from (6.52), (6.54) and (6.55) we see that with $K > 0$ from (6.49) for each $M > 0$ there exists a $K_3 > 0$ such that

$$\begin{aligned}\mathbb{P}\Big(1/\bar{W}_n^2 \le M \text{ and } K\Big\{ \sum_{i=1}^{n} \mathbb{E}_\xi U_i^2 \mathbb{1}_{\{|U_i|>1\}} + \sum_{i=1}^{n} \mathbb{E}_\xi |U_i|^3 \mathbb{1}_{\{|U_i|\le 1\}} \Big\} \\ > K_3((n\Delta_n)^{-1/4} + (n\Delta_n^{1+\rho})^{-\bar{p}})\Big) \to 0. \quad (6.56)\end{aligned}$$

Now we can show (6.48). Let $\eta > 0$ and according to (6.50) choose an $M > 0$ with $\mathbb{P}(1/\bar{W}_n^2 > M) < \eta/2$ for all $n \in \mathbb{N}$. For this $M > 0$ choose a $K_3 > 0$ such that the probability in (6.56) is smaller than $\eta/2$ for large n. Then for $n \in \mathbb{N}$ large enough we have

$$\begin{aligned}&\mathbb{P}\Big(\mathbb{P}_\xi\Big(\hat{\mathbb{H}}_n(\bar{\zeta},\bar{\theta},\bar{z}) > x^{1/r}\Big) \le 2\alpha_n\Big) < \\ &\quad \mathbb{P}\Big((1-\Phi(x^{1/r}/\bar{W}_n)) \le 2\alpha_n + K_3((n\Delta_n)^{-1/4} + (n\Delta_n^{1+\rho})^{-\bar{p}}) \text{ and } 1/\bar{W}_n^2 \le M\Big) + \eta = \eta,\end{aligned}$$

using (6.49) and the fact, that if $1/\bar{W}_n^2 \le M$ there exists a $c > 0$ with $(1-\Phi(x^{1/r}/\bar{W}_n)) > c$. Thus we have shown $\hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r) \xrightarrow{\mathbb{P}} \infty$ and we are only left with proving (4.10). Let

$$K = \Big((1+\varphi)/c^{(\varepsilon)}\Big)^{1/\varpi} > \Big(1/c^{(\varepsilon)}\Big)^{1/\varpi}$$

for some $\varphi > 0$. Then

$$\begin{aligned}\mathbb{P}\Big(\hat{\theta}_n^{(\varepsilon)}(\hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r)) > \theta_0^{(\varepsilon)} + K\beta_n\Big) &\le \mathbb{P}\Big(\sqrt{n\Delta_n}\mathbb{D}_n^{(\varepsilon)}(\theta) \le \hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r) \text{ for some } \theta > \theta_0^{(\varepsilon)} + K\beta_n\Big) \\ &\le \mathbb{P}\Big(\sqrt{n\Delta_n}\mathcal{D}^{(\varepsilon)}(\theta) - \mathbb{H}_n^{(\varepsilon)}(1) \le \hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r) \text{ for some } \theta > \theta_0^{(\varepsilon)} + K\beta_n\Big).\end{aligned}$$

By (4.2) there exists a $y_0 > 0$ with

$$\inf_{\theta \in [\theta_0^{(\varepsilon)} + Ky_1, 1]} \mathcal{D}^{(\varepsilon)}(\theta) = \mathcal{D}^{(\varepsilon)}(\theta_0^{(\varepsilon)} + Ky_1) \geq (c^{(\varepsilon)}/(1+\varphi/2))(Ky_1)^{\varpi}$$

for all $0 \leq y_1 \leq y_0$. Distinguishing the cases $\{\beta_n > y_0\}$ and $\{\beta_n \leq y_0\}$ we get due to $\beta_n \xrightarrow{\mathbb{P}} 0$

$$\begin{aligned}
&\mathbb{P}\Big(\hat{\theta}_n^{(\varepsilon)}(\hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r)) > \theta_0^{(\varepsilon)} + K\beta_n\Big) \\
\leq &\mathbb{P}\Big(\sqrt{n\Delta_n}(c^{(\varepsilon)}/(1+\varphi/2))(K\beta_n)^{\varpi} - \mathbb{H}_n^{(\varepsilon)}(1) \leq \hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r)\Big) + o(1) \leq P_n^{(1)} + P_n^{(2)} + o(1)
\end{aligned}$$

with

$$\begin{aligned}
P_n^{(1)} &= \mathbb{P}\Big(\sqrt{n\Delta_n}(c^{(\varepsilon)}/(1+\varphi/2))(K\beta_n)^{\varpi} - \mathbb{H}_n^{(\varepsilon)}(1) \leq \hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r) \text{ and } \mathbb{H}_n^{(\varepsilon)}(1) \leq b_n\Big), \\
P_n^{(2)} &= \mathbb{P}\Big(\mathbb{H}_n^{(\varepsilon)}(1) > b_n\Big),
\end{aligned}$$

where $b_n := \sqrt{\hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r)}$. Due to the choice $K = \Big((1+\varphi)/c^{(\varepsilon)}\Big)^{1/\varpi}$ and the definition of β_n it is clear that $P_n^{(1)} = o(1)$, because $\hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r) \xrightarrow{\mathbb{P}} \infty$.
Concerning $P_n^{(2)}$ let H_n be the distribution function of $\mathbb{H}_n^{(\varepsilon)}(1)$ and let H be the distribution function of $\mathbb{H}^{(\varepsilon)}(1)$. Then as we have seen in (6.35) in the proof of Theorem 4.4 the function H is continuous and by Theorem 3.2 and the continuous mapping theorem H_n converges pointwise to H. Thus for $\eta > 0$ choose an $x > 0$ with $1 - H(x) < \eta/2$ and conclude

$$P_n^{(2)} \leq \mathbb{P}(b_n \leq x) + 1 - H_n(x) \leq \mathbb{P}(b_n \leq x) + 1 - H(x) + |H_n(x) - H(x)| < \eta,$$

for $n \in \mathbb{N}$ large enough, because of $\hat{\varkappa}_{n,B_n}^{(\varepsilon,\alpha_n)}(r) \xrightarrow{\mathbb{P}} \infty$. □

5.8 Proof of Proposition 4.9.

Under the null hypothesis $\mathbf{H}_0(\varepsilon)$ we have $k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(1) = \mathbb{H}_n^{(\varepsilon)}(1)$. Furthermore,

$$\begin{aligned}
\operatorname{Var}(\mathbb{H}(\bar{\zeta}, 1, \bar{z})) &= \int_0^{\bar{\zeta}} g(y,\bar{z})dy - 2\bar{\zeta}\int_0^{\bar{\zeta}} g(y,\bar{z})dy + \bar{\zeta}^2 \int_0^1 g(y,\bar{z})dy \\
&= (1-\bar{\zeta})^2 \int_0^{\bar{\zeta}} g(y,\bar{z})dy + \bar{\zeta}^2 \int_{\bar{\zeta}}^1 g(y,\bar{z})dy > 0.
\end{aligned}$$

Therefore, as in the proof of Theorem 4.4, $\mathbb{H}^{(\varepsilon)}(1)$ has a continuous cdf and

$$\Big(\mathbb{H}_n^{(\varepsilon)}(1), \hat{\mathbb{H}}_{n,\xi^{(1)}}^{(\varepsilon)}(1), \ldots, \hat{\mathbb{H}}_{n,\xi^{(B)}}^{(\varepsilon)}(1)\Big) \rightsquigarrow \Big(\mathbb{H}^{(\varepsilon)}(1), \mathbb{H}_{(1)}^{(\varepsilon)}(1), \ldots, \mathbb{H}_{(B)}^{(\varepsilon)}(1)\Big).$$

holds in $(\mathbb{R}^{B+1}, \mathbb{B}^{B+1})$ for every $B \in \mathbb{N}$, where $\mathbb{H}^{(\varepsilon)}_{(1)}(1), \ldots, \mathbb{H}^{(\varepsilon)}_{(B)}(1)$ are independent copies of $\mathbb{H}^{(\varepsilon)}(1)$. As a consequnce the assertion follows with Proposition F.1 in the supplement to Bücher and Kojadinovic (2016). The result for the test (4.17) follows in the same way. □

5.9 Proof of Proposition 4.10.

If $\mathbf{H}_1(\varepsilon)$ holds, then (4.18) is a simple consequence of $\lim_{n\to\infty} \mathbb{P}(k_n^{1/2}\mathbb{D}_n^{(\varepsilon)}(1) \geq K) = 1$ for all $K > 0$ and $\lim_{K\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big(\hat{\mathbb{H}}^{(\varepsilon)}_{n,\xi^{(b)}}(1) > K\Big) = 0$ which follow from Theorem 3.2 and Theorem 3.4 by similar arguments as in previous proofs. The second claim can be shown in the same way. □

5.10 Proof of the results in Example 2.3 and Example 4.6(2).

(1) First we show that a kernel as in (2.10) belongs to the set $\mathcal{G}$. Using the uniqueness theorem for measures we see that $g(y, dz)$ is the measure with Lebesgue density $h_y(z) = A(y)\beta(y)/|z|^{1+\beta(y)}$ for each $y \in [0, 1]$ and $z \neq 0$. This function is continuously differentiable with derivative $h'_y(z) = -\operatorname{sgn}(z)A(y)\beta(y)(1 + \beta(y))/|z|^{2+\beta(y)}$, and we obtain

$$\sup_{y\in[0,1]} \sup_{|z|\geq\varepsilon} \Big(h_y(z) + |h'_y(z)|\Big) < \infty$$

for any $\varepsilon > 0$ such that Assumption 2.1(4) is satisfied. Assumption 2.1(3) is obvious, and by definition it is also clear that $g(y, dz)$ does not charge $\{0\}$ for any $y \in [0, 1]$, Finally, a simple calculation using symmetry of the integrand yields

$$\begin{aligned}\sup_{y\in[0,1]} \Big(\int (1 \wedge z^2) g(y, dz)\Big) &= \sup_{y\in[0,1]} \Big(2A(y)\beta(y)\Big\{\int_0^1 z^{1-\beta(y)}dz + \int_1^\infty z^{-1-\beta(y)}dz\Big\}\Big) \\ &= \sup_{y\in[0,1]} \Big(2A(y)\beta(y)\Big\{\frac{1}{2-\beta(y)} + \frac{1}{\beta(y)}\Big\}\Big) < \infty,\end{aligned}$$

by the assumptions on A and β. Thus also Assumption 2.1(2) is valid.

(2) Now we show that if additionally (2.11), (2.12) and (2.13) are satisfied, $k_{0,\varepsilon} < \infty$ holds for every $\varepsilon > 0$ and $g_k(z)$ is a bounded function on M_ε as stated in Example 4.6(2). By shrinking the interval U if necessary we may assume without loss of generality that the functions $\bar{A}, \bar{\beta}$ are bounded away from 0 on U. But then it is well known from complex analysis that there is a domain $U \subset U^* \subset \mathbb{C}$ with holomorphic functions $A^*, \beta^* \colon U^* \to \mathbb{C}^+ := \{u \in \mathbb{C} \mid \operatorname{Re}(u) > 0\}$ such that $\bar{A}, \bar{\beta}$ are the restrictions of A^* and β^* to U. Therefore for any $z \in \mathbb{R} \setminus \{0\}$ the function $g^*(y, z) = A^*(y) \exp\{-\beta^*(y) \log(|z|)\}$ is holomorphic in $y \in U^*$ as a concatenation of holomorphic functions and thus its restriction $\bar{g}(y, z)$ to $y \in U$ is real analytic. Consequently, by shrinking U again if necessary, we have the power series

expansion

$$\bar{g}(y,z) = \sum_{k=0}^{\infty} \frac{g_k(z)}{k!}(y-\theta_0)^k, \tag{6.57}$$

for every $y \in U$ and $z \in \mathbb{R} \setminus \{0\}$. If $k_{0,\varepsilon} = \infty$ for some $\varepsilon > 0$, then for any $k \in \mathbb{N}$ and $z \in M_\varepsilon$ we have $g_k(z) = 0$ and for every $z \in M_\varepsilon$ there is a $C(z) \in (0,\infty)$ with

$$\bar{g}(y,z) = C(z) \Longleftrightarrow \log(\bar{A}(y)) = \log(C(z)) + \bar{\beta}(y)\log(|z|)$$

for every $y \in U$. Taking the derivative with respect to y for a fixed z yields $(\log(\bar{A}(y)))' = \log(|z|)\bar{\beta}'(y)$ for each $y \in U$ and $z \in M_\varepsilon$. But since it is assumed that at least one of the functions $\bar{A}$ and $\bar{\beta}$ is non-constant, there is a $y_0 \in U$ such that one derivative (and therefore both) are different from zero. Varying z for this $y_0 \in U$ yields a contradiction.

In order to show that for each $k \in \mathbb{N}_0$ the function $g_k(z)$ is bounded in $z \in M_\varepsilon$, we use

$$\frac{\partial^\ell}{\partial y^\ell}\Big(\frac{1}{|z|^{\bar{\beta}(y)}}\Big)(\theta_0) = \frac{\partial^\ell}{\partial y^\ell}(\exp(-\bar{\beta}(y)\log(|z|)))(\theta_0) = (-1)^\ell \frac{\log^\ell(|z|)}{|z|^{\beta_0}} b_1^\ell,$$

for $\ell \in \mathbb{N}$. Using the generalization of the product formula for higher derivatives

$$\begin{aligned}\sup_{z\in M_\varepsilon}|g_k(z)| &= \sup_{z\in M_\varepsilon}\Big|\sum_{\ell=0}^{k}\binom{k}{\ell}\bar{A}^{(\ell)}(\theta_0)\frac{\partial^{(k-\ell)}}{\partial y^{(k-\ell)}}\Big(\frac{1}{|z|^{\bar{\beta}(y)}}\Big)(\theta_0)\Big| \\ &\le (K(k+1))^{k+1}\sum_{\ell=0}^{k}\binom{k}{\ell}|a_\ell|\ell!|b_1|^{k-\ell} \le (K(k+1))^{k+1}\sum_{\ell=0}^{k}\binom{k}{\ell}M^\ell|b_1|^{k-\ell} \\ &= (K(k+1))^{k+1}(M+|b_1|)^k \le (K(k+1))^{k+1} \end{aligned} \tag{6.58}$$

follows. The first inequality in (6.58) holds because for $\ell \in \mathbb{N}$ the continuously differentiable function $f_\ell(z) = \log^\ell(z)/z^{\beta_0}$ on (ε,∞) satisfies $\lim_{z\to\infty} f_\ell(z) = 0$ and its derivative can change the sign only in $z = 1$ and $z = \exp\{\ell/\beta_0\}$. Therefore we obtain

$$\sup_{z\in M_\varepsilon}\frac{|\log^\ell(|z|)|}{|z|^{\beta_0}} = \max\Big\{\frac{|\log^\ell(\varepsilon)|}{\varepsilon^{\beta_0}}, \Big(\frac{\ell}{\beta_0}\Big)^\ell e^{-\ell}\Big\} \le (K(\ell+1))^{\ell+1},$$

for some suitable $K > 0$ which does not depend on ℓ. The second inequality in (6.58) is a consequence of (2.12).

(3) Finally, we show the expansion (4.12), and it is immediate to see that it suffices to verify it for $\tilde{\mathcal{D}}^{(\varepsilon)}$ from (2.4).

A power series can be integrated term by term within its radius of convergence. Therefore, (6.57) gives for $\theta_0 \le \theta \in U$

$$\int_{\theta_0}^{\theta} g(y,z)dy = \sum_{k=0}^{\infty} \frac{g_k(z)}{(k+1)!}(\theta-\theta_0)^{k+1},$$

which yields

$$D(\zeta,\theta,z) = \begin{cases} -\frac{\zeta}{\theta}\sum_{k=1}^{\infty}\frac{g_k(z)}{(k+1)!}(\theta-\theta_0)^{k+1}, & \text{if } \zeta \le \theta_0 \\ \frac{1}{\theta}\sum_{k=1}^{\infty}\frac{g_k(z)}{(k+1)!}[\theta(\zeta-\theta_0)^{k+1}-\zeta(\theta-\theta_0)^{k+1}], & \text{if } \theta_0 < \zeta \le \theta. \end{cases} \tag{6.59}$$

For any set T and functions $g, h\colon T \to \mathbb{R}$ we have $|\sup_{t\in T}|g(t)| - \sup_{t\in T}|h(t)|| \le \sup_{t\in T}|g(t)-h(t)|$. Together with (6.59) this yields for $\theta_0 < \theta \in U$

$$\sup_{\zeta\in[0,\theta_0]}\sup_{z\in M_\varepsilon}|D(\zeta,\theta,z)| = \sup_{\zeta\in[0,\theta_0]}\sup_{z\in M_\varepsilon}\Big|\frac{\zeta}{\theta}\frac{g_{k_{0,\varepsilon}}(z)}{(k_{0,\varepsilon}+1)!}(\theta-\theta_0)^{k_{0,\varepsilon}+1}\Big| + O((\theta-\theta_0)^{k_{0,\varepsilon}+2}), \tag{6.60}$$

and

$$\sup_{\zeta\in(\theta_0,\theta]}\sup_{z\in M_\varepsilon}|D(\zeta,\theta,z)| = \sup_{\zeta\in(\theta_0,\theta]}\sup_{z\in M_\varepsilon}\Big|\frac{1}{\theta}\frac{g_{k_{0,\varepsilon}}(z)}{(k_{0,\varepsilon}+1)!}[\theta(\zeta-\theta_0)^{k_{0,\varepsilon}+1}-\zeta(\theta-\theta_0)^{k_{0,\varepsilon}+1}]\Big| + O((\theta-\theta_0)^{k_{0,\varepsilon}+2}), \tag{6.61}$$

for $\theta \downarrow \theta_0$, as soon as we can show

$$\sup_{\zeta\in(\theta_0,\theta]}\sup_{z\in M_\varepsilon}\Big|\frac{1}{\theta}\sum_{k=k_{0,\varepsilon}+1}^{\infty}\frac{g_k(z)}{(k+1)!}[\theta(\zeta-\theta_0)^{k+1}-\zeta(\theta-\theta_0)^{k+1}]\Big| = O((\theta-\theta_0)^{k_{0,\varepsilon}+2}) \tag{6.62}$$

and

$$\sup_{\zeta\in[0,\theta_0]}\sup_{z\in M_\varepsilon}\Big|\frac{\zeta}{\theta}\sum_{k=k_{0,\varepsilon}+1}^{\infty}\frac{g_k(z)}{(k+1)!}(\theta-\theta_0)^{k+1}\Big| = O((\theta-\theta_0)^{k_{0,\varepsilon}+2}). \tag{6.63}$$

To prove (6.62) note that for $k \in \mathbb{N}$ and $\theta_0 < \zeta \le \theta$

$$\begin{aligned}&|\theta(\zeta-\theta_0)^{k+1}-\zeta(\theta-\theta_0)^{k+1}| \le \theta|[(\zeta-\theta_0)^{k+1}-(\theta-\theta_0)^{k+1}]| + (\theta-\zeta)(\theta-\theta_0)^{k+1}\\ =&\theta\Big|\sum_{j=0}^{k}(\theta-\theta_0)^{k-j}(\zeta-\theta)(\zeta-\theta_0)^j\Big| + (\theta-\zeta)(\theta-\theta_0)^{k+1} \le 2\theta(k+1)(\theta-\theta_0)^{k+1}\end{aligned}$$

holds, and by (6.58) we have $\limsup_{k\to\infty}(\bar{g}_k/k!)^{1/k} < \infty$ for $\bar{g}_k = \sup_{z\in M_\varepsilon}|g_k(z)|$. Consequently,

$$\sup_{\zeta\in(\theta_0,\theta]}\sup_{z\in M_\varepsilon}\Big|\frac{1}{\theta}\sum_{k=k_{0,\varepsilon}+1}^{\infty}\frac{g_k(z)}{(k+1)!}[\theta(\zeta-\theta_0)^{k+1}-\zeta(\theta-\theta_0)^{k+1}]\Big| \le \sum_{k=k_{0,\varepsilon}+1}^{\infty}\frac{2\bar{g}_k}{k!}(\theta-\theta_0)^{k+1}$$
$$= (\theta-\theta_0)^{k_{0,\varepsilon}+2}\sum_{k=0}^{\infty}\frac{2\bar{g}_{k+k_{0,\varepsilon}+1}}{(k+k_{0,\varepsilon}+1)!}(\theta-\theta_0)^k = O((\theta-\theta_0)^{k_{0,\varepsilon}+2})$$

for $\theta\downarrow\theta_0$, because the latter power series has a positive radius of convergence around θ_0. For the same reason, in order to prove (6.63) we use

$$\sup_{\zeta\in[0,\theta_0]}\sup_{z\in M_\varepsilon}\Big|\frac{\zeta}{\theta}\sum_{k=k_{0,\varepsilon}+1}^{\infty}\frac{g_k(z)}{(k+1)!}(\theta-\theta_0)^{k+1}\Big| \le \frac{\theta_0}{\theta}(\theta-\theta_0)^{k_{0,\varepsilon}+2}\sum_{k=0}^{\infty}\frac{\bar{g}_{k+k_{0,\varepsilon}+1}}{(k+k_{0,\varepsilon}+2)!}(\theta-\theta_0)^k$$
$$= O((\theta-\theta_0)^{k_{0,\varepsilon}+2})$$

as $\theta\downarrow\theta_0$. Now, because of

$$\tilde{\mathcal{D}}^{(\varepsilon)}(\theta) = \max\Big\{\sup_{\zeta\in[0,\theta_0]}\sup_{z\in M_\varepsilon}|D(\zeta,\theta,z)|,\ \sup_{\zeta\in(\theta_0,\theta]}\sup_{z\in M_\varepsilon}|D(\zeta,\theta,z)|\Big\}$$

(6.60) and (6.61) yield the desired expansion (4.12) for $\tilde{\mathcal{D}}^{(\varepsilon)}$, if we can show for $\theta\downarrow\theta_0$

$$\sup_{\zeta\in[0,\theta_0]}\sup_{z\in M_\varepsilon}\Big|\frac{\zeta}{\theta}\frac{g_{k_{0,\varepsilon}}(z)}{(k_{0,\varepsilon}+1)!}(\theta-\theta_0)^{k_{0,\varepsilon}+1}\Big| = \frac{\bar{g}_{k_{0,\varepsilon}}}{(k_{0,\varepsilon}+1)!}(\theta-\theta_0)^{k_{0,\varepsilon}+1}+O((\theta-\theta_0)^{k_{0,\varepsilon}+2}),$$
$$\sup_{\zeta\in(\theta_0,\theta]}\sup_{z\in M_\varepsilon}\Big|\frac{1}{\theta}\frac{g_{k_{0,\varepsilon}}(z)}{(k_{0,\varepsilon}+1)!}[\theta(\zeta-\theta_0)^{k_{0,\varepsilon}+1}-\zeta(\theta-\theta_0)^{k_{0,\varepsilon}+1}]\Big|$$
$$= \frac{\bar{g}_{k_{0,\varepsilon}}}{(k_{0,\varepsilon}+1)!}(\theta-\theta_0)^{k_{0,\varepsilon}+1}+O((\theta-\theta_0)^{k_{0,\varepsilon}+2}).$$

The first assertion is obvious, since $|\theta_0/\theta-1|\le K(\theta-\theta_0)$ by $0<\theta_0<\theta$. In order to prove the latter claim, consider for $0<\theta_0<\theta<1$ and $k\in\mathbb{N}$ the function $f_k:[\theta_0,\theta]\to\mathbb{R}$ with $f_k(\zeta) = \theta(\zeta-\theta_0)^{k+1}-\zeta(\theta-\theta_0)^{k+1}$. Its derivative is given by $f_k'(\zeta) = \theta(k+1)(\zeta-\theta_0)^k-(\theta-\theta_0)^{k+1}$ and it has a unique root at ζ_0 with

$$\theta_0 < \zeta_0 = \theta_0 + \frac{(\theta-\theta_0)^{1+1/k}}{(\theta(k+1))^{1/k}} < \theta.$$

Thus because of $f_k(\theta_0)<0$, $f_k(\theta)=0$, $f_k'(\zeta)<0$ for $\zeta<\zeta_0$ and $f_k'(\zeta)>0$ for $\zeta>\zeta_0$ we obtain the result, since for $\theta\downarrow\theta_0$

$$\frac{1}{\theta}\sup_{\zeta\in(\theta_0,\theta]}|f_{k_{0,\varepsilon}}(\zeta)| = \frac{1}{\theta}|f_{k_{0,\varepsilon}}(\zeta_0)|$$
$$= \Big|\frac{\theta_0}{\theta}+\frac{(\theta-\theta_0)^{1+1/k_{0,\varepsilon}}}{\theta^{1+1/k_{0,\varepsilon}}(k_{0,\varepsilon}+1)^{1/k_{0,\varepsilon}}}-\frac{(\theta-\theta_0)^{1+1/k_{0,\varepsilon}}}{\theta^{1+1/k_{0,\varepsilon}}(k_{0,\varepsilon}+1)^{1+1/k_{0,\varepsilon}}}\Big|(\theta-\theta_0)^{k_{0,\varepsilon}+1}$$

$$= \frac{\theta_0}{\theta}(\theta-\theta_0)^{k_{0,\varepsilon}+1} + O((\theta-\theta_0)^{k_{0,\varepsilon}+2+1/k_{0,\varepsilon}}) = (\theta-\theta_0)^{k_{0,\varepsilon}+1} + O((\theta-\theta_0)^{k_{0,\varepsilon}+2}).$$

□

References

Aït-Sahalia, Y. and Jacod, J. (2014). *High-Frecuency Financial Econometrics.* Princeton University Press.

Aue, A. and Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16.

Aue, A. and Steinebach, J. (2002). A note on estimating the change-point of a gradually changing stochastic process. *Statistics & Probability Letters*, 56:177–191.

Bissell, A. F. (1984). The performance of control charts and cusums under linear trend. *Applied Statistics*, 33:145–151.

Bücher, A. (2011). *Statistical Inference for Copulas and Extremes.* PhD thesis, Ruhr-Universität Bochum.

Bücher, A., Hoffmann, M., Vetter, M., and Dette, H. (2016+). Nonparametric tests for detecting breaks in the jump behaviour of a time-continuous process. *Bernoulli, to appear.*

Bücher, A. and Kojadinovic, I. (2016). A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing. *Bernoulli*, 22(2):927–968.

Chen, L. and Shao, Q.-M. (2001). A non-uniform Berry-Esseen bound via Stein's method. *Probability Theory and Related Fields*, 120:236–254.

Figueroa-López, J. (2008). Small-time moment asymptotics for Lévy processes. *Statistics and Probability Letters*, 78:3355–3365.

Gaenssler, P., Molnár, P., and Rost, D. (2007). On continuity and strict increase of the CDF for the sup-functional of a Gaussian process with applications to statistics. *Results in Mathematics*, 51:51–60.

Gan, F. F. (1991). Ewma control chart under linear drift. *Journal of Statistical Computation and Simulation*, 38:181–200.

Hoffmann, M. (2016). *Stochastische Integration. Eine Einführung in die Finanzmathematik.* Springer.

Hoffmann, M. and Vetter, M. (2016+). Weak convergence of the empirical truncated distribution function of the Lévy measure of an Itō semimartingale. *Stochastic Processes and their Applications, to appear.*

Hušková, M. (1999). Gradual changes versus abrupt changes. *Journal of Statistical Planning and Inference*, 76:109–125.

Hušková, M. and Steinebach, J. (2002). Asymptotic tests for gradual changes. *Statistics & Decisions*, 20:137–151.

Inoue, A. (2001). Testing for distributional change in time series. *Econometric Theory*, 17(1):156–187.

Jacod, J. and Protter, P. (2012). *Discretization of Processes.* Springer.

Jacod, J. and Shiryaev, A. (2002). *Limit Theorems for Stochastic Processes.* Springer.

Jandhyala, V., Fotopoulos, S., MacNeill, I., and Liu, P. (2013). Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, forthcoming, doi: 10.1111/jtsa12035.

Kosorok, M. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *Journal of Multivariate Analysis*, 84:299–318.

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer Series in Statistics. Springer.

Mallik, A., Banerjee, M., and Sen, B. (2013). Asymptotics for p-value based threshold estimation in regression settings. *Preprint*.

Nickl, R., Reiß, M., Söhl, J., and Trabs, M. (2016). High-frequency Donsker theorems for Lévy measures. *Probability Theory and Related Fields*, 164(1-2):61–108.

Rüschendorf, L. and Woerner, J. H. C. (2002). Expansion of transition distributions of Lévy processes in small time. *Bernoulli*, 8(1):81–96.

Sato, K.-I. (1999). *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press.

Siegmund, D. O. and Zhang, H. (1994). *Confidence regions in broken line regression*, volume 23 of *Lecture Notes–Monograph Series*, pages 292–316. Institute of Mathematical Statistics.

Todorov, V. (2015). Jump activity estimation for pure-jump semimartingales via self-normalized statistics. *The Annals of Statistics*, 43(4):1831–1864.

Todorov, V. (2016). Testing for time-varying jump activity for pure jump semimartingales. *The Annals of Statistics, to appear.*

Van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes.* Springer.

Vogt, M. and Dette, H. (2015). Detecting gradual changes in locally stationary processes. *The Annals of Statistics*, 43(2):713–740.