The Annals of Statistics
2017, Vol. 45, No. 3, 1254–1283
DOI: 10.1214/16-AOS1484

SEMIMARTINGALE DETECTION AND GOODNESS-OF-FIT TESTS

BY ADAM D. BULL[1]

University of Cambridge

In quantitative finance, we often fit a parametric semimartingale model to asset prices. To ensure our model is correct, we must then perform goodness-of-fit tests. In this paper, we give a new goodness-of-fit test for volatility-like processes, which is easily applied to a variety of semimartingale models. In each case, we reduce the problem to the detection of a semimartingale observed under noise. In this setting, we then describe a wavelet-thresholding test, which obtains adaptive and near-optimal detection rates.

1. Introduction.

In quantitative finance, we often model asset prices as semimartingales; in other words, we assume prices are given by a sum of drift, diffusion and jump processes. As these models can be difficult to fit to data, we often restrict our attention to a parametric class, of which many have been suggested by practitioners. To verify our choice of parametric class, we must then perform goodness-of-fit tests.

As semimartingale models can be quite complex, there are many potential tests to perform. In the following, we will be interested in testing whether models accurately describe processes such as the volatility, covolatility, vol-of-vol or leverage. We will further be looking for tests which can be shown to obtain good rates of detection against a variety of alternatives.

While many goodness-of-fit tests exist in the literature, fewer have been shown to obtain good detection rates. Those tests which do achieve good rates are generally designed for one type of semimartingale model, and one way of measuring performance.

In the following, we will therefore describe a new goodness-of-fit test for volatility-like processes in semimartingales. Our test can easily be applied to a wide range of models, including stochastic volatility, jumps and microstructure noise, and obtains good detection rates against both local and nonparametric alternatives.

Our method involves reducing any goodness-of-fit test to one of semimartingale detection: given a series of observations, is the series white noise, or does it contain a hidden semimartingale? We will show how this problem can be solved efficiently, obtaining adaptive and near-optimal detection rates.

Received September 2015; revised June 2016.
[1]Supported by EPSRC Grant EP/K000993/1.
MSC2010 subject classifications. Primary 62M02; secondary 60J60, 60J75, 62G10, 65T60.
Key words and phrases. Diffusion, goodness-of-fit, jump process, semimartingale, wavelets.

We now describe in more detail the problems we consider, as well as relevant previous work. Our goal will be to test the goodness-of-fit of a parametric semimartingale model. Many such models have been described, including simple models such as Black–Scholes or Cox–Ingersoll–Ross; Lévy models such as the generalised hyperbolic or CGMY processes; and stochastic volatility models such as the Heston or Bates models. [For definitions, see Cont and Tankov (2004), Papapantoleon (2008).]

In the simplest case, where our observations are known to come from a stationary or ergodic diffusion process, a great many authors have described goodness-of-fit tests. We briefly mention some initial work [Aït-Sahalia (1996), Corradi and White (1999), Kleinow (2002)] as well as more recent discussion [Chen, Zheng and Pan (2015), González-Manteiga and Crujeiras (2013), Papanicolaou and Giesecke (2014)].

In a financial setting, however, even if our model is stationary, we may need to test it against non-stationary alternatives. When observations can come from a non-stationary diffusion, goodness-of-fit tests have been described using the integrated volatility [Corradi and White (1999)], estimated residuals [Lee (2006), Lee and Wee (2008), Nguyen (2010)] and marginal density [Aït-Sahalia and Park (2012)]. Goodness-of-fit tests also exist for regressions between diffusions [Mykland and Zhang (2006)].

In the following, we will be interested in goodness-of-fit tests which not only detect non-stationary alternatives, but also achieve good detection rates. In this setting, Dette and von Lieres und Wilkau (2003) propose a test which can detect misspecification of the volatility at a rate $n^{-1/4}$ in L^2 norm [see also Dette, Podolskij and Vetter (2006), Papanicolaou and Giesecke (2014), Podolskij and Ziggel (2008)].

A similar test proposed by Dette and Podolskij (2008) detects alternatives in a fixed direction at the faster rate $n^{-1/2}$, although the authors do not give rates in L^p. This test can also be applied to more complex models, including stochastic volatility [Vetter (2012)] and microstructure noise [Vetter and Dette (2012)].

In some volatility testing problems, previous work has described tests which achieve optimal detection rates against nonparametric alternatives [Bibinger, Jirak and Vetter (2015), Reiß, Todorov and Tauchen (2014)]. However, these tests are specific to the problems considered, and do not assess the goodness-of-fit of general models.

In the following, we will therefore describe a new method of goodness-of-fit testing for volatility-like processes. We will show how our approach applies to a wide variety of semimartingale models, including those with jumps, stochastic volatility and microstructure noise. In each case, we will obtain adaptive detection rates, with near-optimal behaviour not only against alternatives in a fixed direction, but also against nonparametric alternatives.

To construct our tests, we will reduce each goodness-of-fit problem to one of semimartingale detection: we will construct a series of observations Z_i, which under the null hypothesis are approximately white noise, and then test whether the Z_i contain a hidden semimartingale S_t.

For example, suppose we have a semimartingale:

$$dX_t = b_t\,dt + \sqrt{\mu_t}\,dB_t,$$

where B_t is a Brownian motion, b_t and μ_t are predictable processes, and we make observations X_{t_i}, $i = 0, \dots, n$, where the times $t_i := i/n$. Further suppose we have a model $\mu(t, X_t)$ for the volatility, and wish to test the hypotheses:

$$H_0 : \mu_t = \mu(t, X_t) \quad \text{versus} \quad H_1 : \mu_t \text{ unrestricted}.$$

To estimate μ_t, we define the realised volatility estimates

$$Y_i := n(X_{t_{i+1}} - X_{t_i})^2, \qquad i = 0, \dots, n-1.$$

Since the scaled increments $\sqrt{n}(X_{t_{i+1}} - X_{t_i})$ are approximately $N(0, \mu_{t_i})$, the observations Y_i have approximate mean μ_{t_i} and variance $2\mu_{t_i}^2$. Under H_0, we thus have that the normalised observations:

$$Z_i := \big(Y_i - \mu(t_i, X_{t_i})\big)/\sigma(t_i, X_{t_i}), \qquad \sigma^2 := 2\mu^2,$$

are approximately white noise.

Under H_1, we instead obtain

$$Z_i = S_{t_i} + \varepsilon_i, \tag{1}$$

where the semimartingale

$$S_t := \big(\mu_t - \mu(t, X_t)\big)/\sigma(t, X_t),$$

and the approximately-centred noises

$$\varepsilon_i := (Y_i - \mu_{t_i})/\sigma(t_i, X_{t_i}).$$

To test our hypotheses, we must therefore test whether the series Z_i is approximately white noise, or contains a hidden semimartingale S_t.

If the noises ε_i were independent standard Gaussian, independent of S_t, we could consider this a standard detection problem in nonparametric regression. Conditioning on S_t, we could take the semimartingale as fixed, and then apply the methods of Ingster and Suslina (2003), for example.

Under suitable assumptions on the process S_t, its sample paths would be almost $\frac{1}{2}$-smooth, and we would thus be able to detect a signal S_t at rate $n^{-1/4}$ in supremum norm, up to log terms. Alternatively, if we wished to detect signals $S_t \propto e_t$, for a fixed direction e_t, we could do so at a rate $n^{-1/2}$.

In general, however, the signal S_t may depend on past values of the noises ε_i, and vice versa. We will thus not be able to appeal directly to results in nonparametric regression, and will instead need to use arguments developed specifically for the semimartingale setting.

In the following, we will show that testing problems like (1) can be solved with detection rates similar to those of nonparametric regression. We will further show that many semimartingale goodness-of-fit tests can be described in a form like (1), including models with stochastic volatility, jumps or microstructure noise.

Our approach will be similar to wavelet thresholding [Donoho et al. (1995), Hoffmann, Munk and Schmidt-Hieber (2012)]; essentially, we will reject the null whenever a suitable wavelet-thresholding estimate of S_t is nonzero. While this method is known to work well in the standard nonparametric setting, we will need to prove new results to apply it to settings like (1).

Our proofs will use a Gaussian coupling derived from Skorokhod embeddings. We note that as our results must apply in a general semimartingale setting, we will not be able to use faster-converging couplings, such as the KMT approximation. We will show, however, that under reasonable moment bounds, a Skorokhod embedding will suffice to achieve the desired detection rates.

Indeed, with this construction we will show our tests detect semimartingales S_t at a rate $n^{-1/4}$ in supremum norm, up to log terms, even when S_t contains finite-variation jumps. Furthermore, our tests will simultaneously detect simpler signals at faster rates; for example, we will be able to detect signals S_t in a fixed direction e_t at a rate $n^{-1/2}$ up to logs, without knowledge of the direction e_t.

We will finally show that in each case, the rates obtained are near-optimal. Applying our tests to problems like (1), we will thus be able to construct goodness-of-fit tests for a wide variety of semimartingale models, obtaining adaptive and near-optimal detection rates.

The paper will be organised as follows. In Section 2, we give a rigorous description of the problems we consider, and discuss examples. In Section 3, we then construct our tests, and state our theoretical results. In Section 4, we then give empirical results, and in Section 5, proofs.

2. Semimartingale detection problems.

We now describe our concept of a semimartingale detection problem. Our setting will include volatility goodness-of-fit problems like (1), as well as many other semimartingale goodness-of-fit tests.

We begin with some examples of the problems we will consider. In each case, we will describe a semimartingale model with a volatility-like process μ_t. We will wish to test the null hypothesis that μ_t is given by some known function $\mu(\theta_0, t, X_t)$, for an unknown paramter $\theta_0 \in \Theta$, and an estimable covariate process $X_t \in \mathbb{R}^q$; our alternative hypothesis will be that μ_t is not given by μ.

To test our hypothesis, we will construct $\mathcal{F}_{t_{i+1}}$-measurable observations Y_i, and a variance function σ^2. Under the null, and conditional on $\mathcal{F}_{t_i}$, the observations Y_i will have approximate mean and variance $\mu(\theta_0, t_i, X_{t_i})$ and $\sigma^2(\theta_0, t_i, X_{t_i})$. To estimate these means and variances, we will further construct estimates $\widehat{\theta}$ and $\widehat{X}_i$ of the parameters θ_0 and covariates X_{t_i}.

We will then be able to estimate the difference between the observations Y_i and their means μ, scaled according to their variances σ^2; we will reject the null hypothesis when the size of these scaled differences are large. In Section 3, we describe in detail how we perform such tests, as well as giving theoretical results on their performance.

For now, we proceed with some examples of semimartingale goodness-of-fit problems in this form. Let B_t and B'_t be independent Brownian motions, $\lambda(dx, dt)$ be an independent Poisson random measure with intensity $dx\,dt$, b_t and b'_t be predictable locally-bounded processes, and $f_t(x)$ be a predictable function with $\int_{\mathbb{R}} 1 \wedge |f_t(x)|^\beta\,dx$ locally bounded, for some $\beta \in [0, 1)$. Further define times $t'_i := i/n^2$.

We then have the following examples.

Local volatility. We wish to test a model μ for μ_t in the process

$$dX_t = b_t\,dt + \sqrt{\mu_t}\,dB_t, \tag{2}$$

making observations X_{t_i}, $i = 0, \dots, n$. We set $\widehat{X}_i := X_{t_i}$, and estimate μ_{t_i} by the realised volatility [Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002)],

$$Y_i := n(X_{t_{i+1}} - X_{t_i})^2.$$

We then define the variance function $\sigma^2 := 2\mu^2$.

Jumps. We wish to test a model μ for μ_t in the process

$$dX_t = b_t\,dt + \sqrt{\mu_t}\,dB_t + \int_{\mathbb{R}} f_t(x)\lambda(dx, dt),$$

making observations X_{t_i}, $i = 0, \dots, n$. We set $\widehat{X}_i := X_{t_i}$, and estimate μ_{t_i} by the truncated realised volatility [Jacod and Reiss (2014), Mancini (2009)],

$$Y_i = g_n\big(\sqrt{n}(X_{t_{i+1}} - X_{t_i})\big), \qquad g_n(x) = x^2 1_{x^2 < \alpha_n},$$

for any sequence $\alpha_n > 0$ satisfying

$$\log(n) = o(\alpha_n), \qquad \alpha_n = o(n^\kappa) \qquad \text{for all } \kappa > 0. \tag{3}$$

We then define the variance function $\sigma^2 := 2\mu^2$.

Microstructure noise. We wish to test a model μ for μ_t in the process

$$dX_{1,t} = b_t\,dt + \sqrt{\mu_t}\,dB_t.$$

We make observations

$$\widetilde{X}_{1,i} := X_{1,t'_i} + \varepsilon_i, \qquad i = 0, \dots, n^2,$$

where the noises ε_i are measurable in the filtrations $\mathcal{F}^+_{t'_i} := \bigcap_{s > t'_i} \mathcal{F}_s$, and satisfy

$$\mathbb{E}[\varepsilon_i | \mathcal{F}_{t'_i}] = 0,$$

$$\mathbb{E}[\varepsilon_i^2 | \mathcal{F}_{t'_i}] = X_{2,t'_i},$$

$$\mathbb{E}\big[|\varepsilon_i|^\kappa | \mathcal{F}_{t'_i}\big] \le C,$$

for an Itô semimartingale $X_{2,t}$ with locally-bounded characteristics, and constants $\kappa > 8$, $C > 0$. We estimate X_{t_j} and μ_{t_j} by their pre-averaged counterparts [Jacod et al. (2009), Reiß (2011)],

$$\widehat{X}_{1,j} := n^{-1} \sum_{i=0}^{n-1} \widetilde{X}_{1,nj+i},$$

$$\widehat{X}_{2,j} := (2n)^{-1} \sum_{i=0}^{n-1} (\widetilde{X}_{1,nj+i+1} - \widetilde{X}_{1,nj+i})^2,$$

$$Y_j := \pi^2 \left(2n^{-1} \left(\sum_{i=0}^{n-1} \cos\left(\pi \left(i + \frac{1}{2}\right)/n\right) \widetilde{X}_{1,nj+i} \right)^2 - \widehat{X}_{2,j} \right).$$

We then define the variance function $\sigma^2 := 2(\mu + \pi^2 X_{2,t})^2$.

Stochastic volatility. We wish to test a model μ for μ_t in the processes

$$dX_{1,t} = b_t \, dt + \sqrt{X_{2,t}} \, dB_t,$$
$$dX_{2,t} = b_t' \, dt + \sqrt{\mu_t} \, dB_t',$$

making observations $X_{1,t_i'}$, $i = 0, \dots, n^2$. We define volatility estimates

$$\widetilde{X}_{2,i} := n^2 (X_{1,t_{i+1}'} - X_{1,t_i'})^2, \qquad i = 0, \dots, n^2 - 1,$$

which we use to estimate X_{t_j} and μ_{t_j} [Barndorff-Nielsen and Veraart (2009), Vetter (2012)],

$$\widehat{X}_{1,j} := X_{1,t_j},$$

$$\widehat{X}_{2,j} := n^{-1} \sum_{i=0}^{n-1} \widetilde{X}_{2,nj+i},$$

$$Y_j := 2\pi^2 \left(n^{-1} \left(\sum_{i=0}^{n-1} \cos\left(\pi \left(i + \frac{1}{2}\right)/n\right) \widetilde{X}_{2,nj+i} \right)^2 - \widehat{X}_{2,j}^2 \right).$$

We then define the variance function $\sigma^2 := 2(\mu + 2\pi^2 X_{2,t}^2)^2$.

Others. Many other models, for example, including covolatility or leverage, or combining any of the above features, can be described similarly. For simplicity, we assume in the following that the times t_i are deterministic and uniform; however, models with uneven or random times that are suitably dense and predictable can be addressed in a similar fashion.

To concisely describe these examples, and others, we will state a set of assumptions on the observations Y_i, mean and variance functions μ and σ^2, parameters θ, covariates X_t and estimates $\widehat{X}_i$. It will be possible to show that the above models all lie within our assumptions, and we may thus work within these assumptions with some generality.

To begin, we define some notation. Let $\|\cdot\|$ denote any finite-dimensional vector norm; write $a = O(b)$ if $\|a\| \leq C\|b\|$, for some universal constant C; and write $a = O_p(b)$ if for each $\varepsilon > 0$, the random variables a and b satisfy $\mathbb{P}(\|a\| > C_\varepsilon\|b\|) \leq \varepsilon$, for universal constants C_ε.

We stress here that the implied constants C and C_ε are universal; in statements such as $a = O(1)$, we require the supremum $\sup\|a\|$ over all such a to be bounded. Given a function $f : X \to \mathbb{R}$, we also define the supremum norm $\|f\|_\infty := \sup_{x\in X}|f(x)|$.

Our assumptions are then as follows.

ASSUMPTION 1. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,1]}, \mathbb{P})$ be a filtered probability space, with adapted unobserved mean, variance and covariate processes $\mu_t \in \mathbb{R}$, $\sigma_t^2 \geq 0$, and $X_t \in \mathbb{R}^q$, respectively. For $0 \leq t \leq t + h \leq 1$, letting W_t denote either of the processes μ_t or X_t, we have

$$\begin{aligned} W_t &= O(1), \\ \mathbb{E}[W_{t+h} - W_t|\mathcal{F}_t] &= O(h), \\ \mathbb{E}\big[\|W_{t+h} - W_t\|^2|\mathcal{F}_t\big] &= O(h). \end{aligned} \tag{4}$$

For $i = 0, \dots, n-1$, we have $\mathcal{F}_{t_{i+1}}$-measurable estimates $\widehat{X}_i$ of X_{t_i}, satisfying

$$\begin{aligned} \mathbb{E}\big[\|\widehat{X}_i - X_{t_i}\|^2|\mathcal{F}_{t_i}\big] &= O(n^{-1}), \\ \mathbb{E}\big[\|\widehat{X}_i - X_{t_i}\|^4|\mathcal{F}_{t_i}\big] &= O(n^{-1}). \end{aligned} \tag{5}$$

We also have $\mathcal{F}_{t_{i+1}}$-measurable observations Y_i, satisfying

$$\begin{aligned} \mathbb{E}[Y_i|\mathcal{F}_{t_i}] &= \mu_{t_i} + O(n^{-1/2}), \\ \mathbb{V}\mathrm{ar}[Y_i|\mathcal{F}_{t_i}] &= \sigma_{t_i}^2 + O(n^{-1/4}), \\ \mathbb{E}\big[|Y_i|^{4+\varepsilon}|\mathcal{F}_{t_i}\big] &= O(1), \end{aligned} \tag{6}$$

for a constant $\varepsilon > 0$.

Under the null hypothesis H_0, we suppose our observations Y_i are described by a parametric model,

$$\mu_t = \mu(\theta_0, t, X_t), \qquad \sigma_t^2 = \sigma^2(\theta_0, t, X_t),$$

for known functions $\mu, \sigma^2 : \Theta \times [0,1] \times \mathbb{R}^q \to \mathbb{R}$, and an unknown parameter $\theta_0 \in \Theta$. We suppose that $\Theta \subseteq \mathbb{R}^p$ is closed, and σ^2 is positive. We also suppose the functions μ and σ^2 are locally Lipschitz in θ, continuously differentiable in t and twice continuously differentiable in X. Finally, we suppose we have a good estimate $\widehat{\theta}$ of θ_0, satisfying

$$\widehat{\theta} - \theta_0 = O_p(n^{-1/2}).$$

Under the alternative hypothesis H_1, we instead allow μ_t, σ_t unrestricted, and require only that $\widehat{\theta} = O_p(1)$.

To ensure the examples given above lie within Assumption 1, we must require that the parameter space $\Theta \subseteq \mathbb{R}^p$ be closed, and the model function μ be locally Lipschitz in θ, continuously differentiable in t, and twice continuously differentiable in X_t. These conditions should all be satisfied for most common models.

We must further require the semimartingales X_t to be bounded, and have bounded characteristics. In general, this assumption may not hold directly; however, we can assume it without loss of generality using standard localisation arguments.

In the supplemental material [Bull (2016b)], we then check that the above examples satisfy our conditions on the processes μ_t, σ_t and X_t; estimates $\widehat{X}_i$; and observations Y_i. Most of these conditions follow from standard results on stochastic processes where necessary, higher-moment bounds can be proved using our Lemma 1 below.

To satisfy Assumption 1, it remains to choose an estimate $\widehat{\theta}$ of θ_0, having error $O_p(n^{-1/2})$ under H_0, and being $O_p(1)$ under H_1. While our results are agnostic as to the choice of $\widehat{\theta}$, a simple choice is given by the least-squares estimate

$$\widehat{\theta} := \operatorname*{arg\,min}_{\theta\in\Theta} \sum_{i=0}^{n-1} \bigl(Y_i - \mu(\theta, t_i, \widehat{X}_i)\bigr)^2, \tag{7}$$

which can be found by numerical optimisation. Under standard regularity assumptions for nonlinear regression, this estimate $\widehat{\theta}$ can be shown to satisfy our conditions arguing, for example, as in Section 5 of Vetter and Dette (2012).

Finally, we note that in the microstructure noise and stochastic volatility models, we need to make $n^2 + 1$ observations of the underlying process X_t to construct the n estimates Y_i. We may thus expect to achieve the square-root of any convergence rates given below; such behaviour, however, is common to all approaches to these problems in the literature.

We have thus shown that many different semimartingale goodness-of-fit problems can be described by our Assumption 1. Next, we will describe our solutions to these problems.

3. Wavelet detection tests. To state our tests for the problems given by Assumption 1, we first consider the signal function:

$$S_t(\theta) := \bigl(\mu_t - \mu(\theta, t, X_t)\bigr)/\sigma(\theta, t, X_t).$$

This function measures the distance of the model mean μ from the true mean μ_t, weighted by the model variance σ^2. Under H_0, we have

$$S_t(\widehat{\theta}) \approx S_t(\theta_0) = 0,$$

while under H_1, we can in general expect $|S_t(\widehat{\theta})|$ to be large. We may thus reject H_0 whenever an estimate of $S_t(\widehat{\theta})$ is significantly different from zero.

To estimate the signal $S_t(\theta)$, we will use wavelet methods. Let φ and ψ be the Haar scaling function and wavelet,

$$\varphi := 1_{[0,1)}, \qquad \psi := 1_{[0,1/2)} - 1_{[1/2,1)},$$

and for $j = 0, 1, \dots, k = 0, \dots, 2^j - 1$, define the Haar basis functions

$$\varphi_{j,k}(t) := 2^{j/2}\varphi(2^j t - k), \qquad \psi_{j,k}(t) := 2^{j/2}\psi(2^j t - k).$$

We can then describe $S_t(\theta)$ in terms of its scaling and wavelet coefficients

$$\alpha_{j,k}(\theta) := \int_0^1 \varphi_{j,k}(t) S_t(\theta)\, dt, \qquad \beta_{j,k}(\theta) := \int_0^1 \psi_{j,k}(t) S_t(\theta)\, dt.$$

To estimate these coefficients, we first pick a resolution level $J \in \mathbb{N}_0$, so that 2^J is of order $n^{1/2}$. We then estimate the scaling coefficients $\alpha_{J,k}(\theta)$ by

$$\widehat{\alpha}_{J,k}(\theta) := n^{-1} \sum_{i=0}^{n-1} \varphi_{J,k}(t_i) Z_i(\theta),$$

where the normalised observations

$$Z_i(\theta) := \bigl(Y_i - \mu(\theta, t_i, \widehat{X}_i)\bigr)/\sigma(\theta, t_i, \widehat{X}_i).$$

We note that for fixed θ, these estimates can be computed in linear time, as each observation Y_i contributes to only one coefficient $\widehat{\alpha}_{J,k}(\theta)$.

To estimate the coefficients $\alpha_{0,0}(\theta)$ and $\beta_{j,k}(\theta)$, $0 \le j < J$, we then perform a fast wavelet transform, obtaining estimates

$$\widehat{\alpha}_{0,0}(\theta) := \sum_l \widehat{\alpha}_{J,l}(\theta) \int_0^1 \varphi_{J,l}\varphi_{0,0}, \qquad \widehat{\beta}_{j,k}(\theta) := \sum_l \widehat{\alpha}_{J,l}(\theta) \int_0^1 \varphi_{J,l}\psi_{j,k}.$$

We note that efficient implementations of this transformation, running in linear time, are widely available.

To test our hypotheses, we will take the maximum size of these estimated coefficients, producing test statistics

$$\widehat{T}(\theta) := \max_{0 \le j < J, k} \bigl|\widehat{\alpha}_{0,0}(\theta)\bigr|, \bigl|\widehat{\beta}_{j,k}(\theta)\bigr|.$$

We will show that under H_0, $\widehat{T}(\widehat{\theta})$ is asymptotically Gumbel distributed, while under H_1, $\widehat{T}(\widehat{\theta})$ will tend to be greater.

THEOREM 1. *Let Assumption 1 hold*:

(i) *Under* H_0,

$$a_{2^J}^{-1}\big(n^{1/2}\widehat{T}(\widehat{\theta}) - b_{2^J}\big) \xrightarrow{d} G$$

uniformly, *where the constants*

$$a_m := \big(2\log(m)\big)^{-1/2},$$

$$b_m := a_m^{-1} - \frac{1}{2}a_m \log\big(\pi \log(m)\big),$$

and G denotes the standard Gumbel distribution.

(ii) *Under* H_1,

$$\widehat{T}(\widehat{\theta}) - T(\widehat{\theta}) = O_p\big(n^{-1/2}\log(n)^{1/2}\big)$$

uniformly, *where*

$$T(\theta) := \max_{0\le j<J,k} \big|\alpha_{0,0}(\theta)\big|, \big|\beta_{j,k}(\theta)\big|.$$

We thus obtain that under H_0, $\widehat{T}(\widehat{\theta})$ concentrates around zero at a rate $n^{-1/2}\log(n)^{1/2}$. Under H_1, it concentrates at the same rate around the quantity $T(\widehat{\theta})$, which measures the size of the signal $S_t(\widehat{\theta})$. We can use this result to construct tests of our hypotheses, and prove bounds on their performance; we first note that for some of our bounds, we will require the following assumption.

ASSUMPTION 2. The processes μ_t and X_t are Itô semimartingales,

$$\mu_t = \int_0^t \bigg(b_s^{\mu}\,ds + \big(c_s^{\mu}\big)^T\,dB_s + \int_{\mathbb{R}} f_s^{\mu}(x)\lambda(dx, ds)\bigg),$$

$$X_{i,t} = \int_0^t \bigg(b_{i,s}^{X}\,ds + \big(c_{i,s}^{X}\big)^T\,dB_s + \int_{\mathbb{R}} f_{i,s}^{X}(x)\lambda(dx, ds)\bigg),$$

for a Brownian motion $B_s \in \mathbb{R}^{q+1}$, independent Poisson random measure $\lambda(dx, ds)$ having compensator $dx\,ds$, predictable processes b_s^{μ}, $b_{i,s}^{X}$, c_s^{μ}, $c_{i,s}^{X} = O(1)$, and predictable functions $f_s^{\mu}(x)$, $f_{i,s}^{X}(x)$ satisfying $\int_{\mathbb{R}} 1 \wedge |f_s(x)|\,dx = O(1)$.

Under Assumption 2, we thus have that μ_t and X_t are Itô semimartingales, with bounded characteristics and finite-variation jumps. This assumption holds for many common financial models, if necessary after a suitable localisation step. Using this condition, we are now ready to describe our tests, and bound their performance.

THEOREM 2. *Let Assumption* 1 *hold*, *and for* $\alpha \in (0, 1)$, *define the Gumbel quantile*

$$q_{n,\alpha} := -a_{2^J}\log\big(-\log(1-\alpha)\big) + b_{2^J},$$

and critical region

$$C_{n,\alpha} := \{n^{1/2}\widehat{T}(\widehat{\theta}) > q_{n,\alpha}\}.$$

(i) *Under* H_0, *we have* $\mathbb{P}[C_{n,\alpha}] \to \alpha$ *uniformly.*
(ii) *Under* H_1, *let* $M_n > 0$ *be a fixed sequence with* $M_n \to \infty$. *If* E_n *is one of the events*:
  (a) $\{\|S(\widehat{\theta})\|_\infty \geq M_n n^{-1/4}\log(n)^{1/2}\}$, *given also Assumption* 2; *or*
  (b) $\{\max_{0\leq j\leq J,k} 2^{j/2}|\int_{2^{-j}k}^{2^{-j}(k+1)} S_t(\widehat{\theta})\,dt| \geq M_n n^{-1/2}\log(n)^{1/2}\}$;
  we have $\mathbb{P}[E_n \setminus C_{n,\alpha}] \to 0$ *uniformly.*

We thus obtain that the test which rejects H_0 on the event $C_{n,\alpha}$ is of asymptotic size α, and under Assumption 2, can detect signals $S_t(\widehat{\theta})$ at the rate $n^{-1/4}\log(n)^{1/2}$ in supremum norm. We further have that, even without Assumption 2, our test can detect a signal whenever the size of its mean over a dyadic interval is large.

In particular, if $S_t(\widehat{\theta}) \propto e_t$ for some nonzero deterministic process e_t, then e_t must have nonzero integral over some dyadic interval $2^{-j}[k, k+1)$. We deduce that our test can detect signals in the fixed direction e_t at the rate $n^{-1/2}\log(n)^{1/2}$, without prior knowledge of e_t.

We can further show that these detection rates are near-optimal.

THEOREM 3. *Let Assumption* 1 *hold, and* $\delta_n > 0$ *be a fixed sequence with* $\delta_n \to 0$. *If* E_n *is one of the events*:

(i) $\{\|S(\widehat{\theta})\|_\infty \geq \delta_n n^{-1/4}\}$, *given also Assumption* 2; *or*
(ii) $\{\max_k 2^{j_n/2}|\int_{2^{-j_n}k}^{2^{-j_n}(k+1)} S_t(\widehat{\theta})\,dt| \geq \delta_n n^{-1/2}\}$, *for some* $j_n = 0, \dots, J$;

then no sequence of critical regions C_n *can satisfy*

$$\limsup_n \mathbb{P}[C_n] < 1$$

uniformly over H_0, *and*

$$\mathbb{P}[E_n \setminus C_n] \to 0$$

uniformly over H_1.

We thus conclude that our goodness-of-fit tests achieve the near-optimal detection rate of $n^{-1/4}\log(n)^{1/2}$ against general nonparametric alternatives, in a wide variety of semimartingale models. This result is already a significant improvement over previous work; we note that similar methods do not establish near-optimality for the procedures of Dette and von Lieres und Wilkau (2003), for example, where the corresponding lower bound would be $n^{-1/3}$.

Furthermore, we have shown that our method simultaneously provides near-optimal detection rates against alternatives which are easier to detect, including the case where the signal $S_t(\widehat{\theta})$ lies in a fixed direction e_t. We may thus achieve good detection rates in a fully nonparametric setting, without sacrificing performance against fixed alternatives.

4. Finite-sample tests. We next consider the empirical performance of our tests. As convergence to the Gumbel distribution can be quite slow, in the following, we will consider a bootstrap version of our tests, which will be more accurate in finite samples.

The general procedure is as follows. First, we estimate the parameters θ from the data, using some estimate $\widehat{\theta}$. Next, we simulate many sets of observations $Y_i^{(j)}$ from the null hypothesis, with parameters chosen by $\widehat{\theta}$. Any components of the null hypothesis not described by θ, such as drift or jump processes, are set to zero.

For each set of simulated observations $Y_i^{(j)}$, we then compute a parameter estimate $\widehat{\theta}^{(j)}$, and statistic $\widehat{T}^{(j)}(\widehat{\theta}^{(j)})$. Finally, we reject the null hypothesis if the original statistic $\widehat{T}(\widehat{\theta})$ is larger than the $(1-\alpha)$-quantile of the simulated statistics $\widehat{T}^{(j)}(\widehat{\theta}^{(j)})$.

We now perform some simple Monte Carlo experiments on these tests. We will compare our tests to those of Dette and von Lieres und Wilkau (2003), Dette, Podolskij and Vetter (2006) and Dette and Podolskij (2008), using the same methodology as Dette and Podolskij. As in that paper, we will generate Monte Carlo observations in the local volatility setting (2). We will then use our tests to evaluate the goodness-of-fit of various parametric models for the volatility.

In each case, we consider receiving $n = 100$, 200 or 500 observations, and constructing confidence tests at the $\alpha = 5\%$ or 10% level. We then generate 1000 realisations of simulated data, compare our statistic against 1000 bootstrap samples in each realisation, and report the proportion of runs in which the null hypothesis is rejected.

In our tests, we set the resolution level $J := \lfloor \log_2(n)/2 \rfloor$, and use the least-squares parameter estimates $\widehat{\theta}$ given by (7). As the models we consider will be linear in the parameters θ, we will be able to compute these estimates in closed form, as linear regressions.

Table 1 then gives the observed rejection probabilities of our tests in two models: a constant volatility model, where $\mu(x,t,\theta) = \theta$; and a proportional volatility model, where $\mu(x,t,\theta) = \theta x^2$. In each case, we give results for our tests under a variety of null and alternative hypotheses.

We note the hypotheses tested are the same as in Tables 1–4 of Dette and Podolskij (2008), as well as Table 3 of Dette and von Lieres und Wilkau (2003), and Tables 3.1 and 3.4 of Dette, Podolskij and Vetter (2006). We may thus directly compare the performance of our tests to those given in previous work.

We find that in both models, our tests have good coverage under the null hypothesis, and reliably reject under the alternative hypothesis. The power of our tests is competitive with previous work under the constant volatility model, and generally improves upon previous work under the proportional volatility model.

TABLE 1
Observed rejection probabilities for bootstrap test

n	100		200		500	
α	5%	10%	5%	10%	5%	10%
Constant volatility, null, $\mu_t = 1$						
$b_t = 0$	0.048	0.105	0.056	0.101	0.035	0.089
$b_t = 2$	0.055	0.114	0.057	0.103	0.044	0.084
$b_t = X_t$	0.056	0.101	0.041	0.093	0.037	0.092
$b_t = 2 - X_t$	0.048	0.095	0.052	0.105	0.051	0.100
$b_t = tX_t$	0.038	0.094	0.060	0.101	0.063	0.111
Constant volatility, alternative, $b_t = X_t$						
$\sqrt{\mu_t} = 1 + X_t$	0.777	0.840	0.898	0.932	0.976	0.985
$\sqrt{\mu_t} = 1 + \sin 5X_t$	0.964	0.977	0.997	0.999	1.000	1.000
$\sqrt{\mu_t} = 1 + X_t \exp t$	0.954	0.975	0.987	0.994	0.999	0.999
$\sqrt{\mu_t} = 1 + X_t \sin 5t$	0.851	0.908	0.970	0.982	0.994	0.995
$\sqrt{\mu_t} = 1 + tX_t$	0.742	0.796	0.883	0.914	0.951	0.972
Proportional volatility, null, $\mu_t = X_t^2$						
$b_t = 0$	0.062	0.119	0.044	0.090	0.043	0.087
$b_t = 2$	0.073	0.120	0.056	0.106	0.043	0.081
$b_t = X_t$	0.070	0.115	0.055	0.100	0.043	0.098
$b_t = 2 - X_t$	0.053	0.085	0.055	0.100	0.034	0.081
$b_t = tX_t$	0.070	0.106	0.062	0.123	0.045	0.106
Proportional volatility, alternative, $b_t = 2 - X_t$						
$\mu_t = 1 + X_t^2$	0.602	0.673	0.700	0.766	0.844	0.884
$\mu_t = 1$	0.832	0.871	0.927	0.951	0.979	0.991
$\mu_t = 5\lvert X_t\rvert^{3/2}$	0.580	0.669	0.672	0.760	0.854	0.902
$\mu_t = 5\lvert X_t\rvert$	0.896	0.932	0.963	0.974	0.995	0.998
$\mu_t = (1 + X_t)^2$	0.831	0.878	0.894	0.929	0.964	0.979

We conclude that our tests not only achieve good theoretical detection rates, but also provide strong finite-sample performance. They may thus be recommended for many different goodness-of-fit problems, whether previously discussed in the literature, or newly described by our more general assumptions.

5. Proofs. We now give proofs of our results. In Section 5.1, we will state some technical results, in Section 5.2 give our main proofs and in the supplemental material [Bull (2016b)] prove our technical results.

5.1. *Technical results.* We first state the technical results we will require. Our main technical result will be a central limit theorem for martingale difference sequences, bounding the exponential moments of the distance from Gaussian.

LEMMA 1. *Let* $(\Omega, \mathcal{F}, (\mathcal{F}_j)_{j=0}^n, \mathbb{P})$ *be a filtered probability space, and let* X_i, $i = 0, \dots, n-1$, *be* $\mathcal{F}_{i+1}$*-measurable real random variables. Suppose that for* *some* $\kappa \geq 1$,

$$\mathbb{E}[X_i | \mathcal{F}_i] = 0,$$

$$\sum_{i=0}^{n-1} \mathbb{E}\big[|X_i|^{4\kappa} | \mathcal{F}_i\big] = O\big(n^{1-2\kappa}\big).$$

(i) *If also*

$$\mathbb{E}\Bigg[\Bigg|\sum_{i=0}^{n-1} \mathbb{E}\big[X_i^2 | \mathcal{F}_i\big] - 1\Bigg|^{2\kappa} \Bigg| \mathcal{F}_0\Bigg] = O\big(n^{-\kappa}\big),$$

then on a suitably-extended probability space, we have real random variables ξ, η *and* M, *independent of* $\mathcal{F}$ *given* $\mathcal{F}_n$, *such that*

$$\sum_{i=0}^{n-1} X_i = \xi + \eta;$$

ξ *is standard Gaussian given* $\mathcal{F}_0$; *we have*

$$\mathbb{E}\big[|\eta|^{4\kappa} | \mathcal{F}_0\big] = O\big(n^{-\kappa}\big);$$

for $u \in \mathbb{R}$,

$$\mathbb{E}\bigg[\exp\bigg(u\eta - \frac{1}{2}u^2 M\bigg)\bigg|\mathcal{F}_0\bigg] \leq 1;$$

and $M \geq 0$ *satisfies*

$$\mathbb{E}\big[M^{2\kappa} | \mathcal{F}_0\big] = O\big(n^{-\kappa}\big). \tag{8}$$

(ii) *For random variables* $c_i = O(1)$, *let* $\upsilon_c := \sum_{i=0}^{n-1} c_i X_i$. *Then on a suitably-extended probability space, we have a constant* $A = O(1)$ *and real random variable* M, *independent of* $\mathcal{F}$ *given* $\mathcal{F}_n$, *such that*

$$\sup_c \mathbb{E}\big[|\upsilon_c|^{4\kappa} | \mathcal{F}_0\big] = O(1);$$

for $u \in \mathbb{R}$,

$$\sup_c \mathbb{E}\bigg[\exp\bigg(u\upsilon_c - \frac{1}{2}u^2(A + M)\bigg)\bigg|\mathcal{F}_0\bigg] \leq 1;$$

and $M \geq 0$ *satisfies* (8).

We will also need the following result on combining exponential moment bounds.

LEMMA 2. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, with real random variables* $(X_i)_{i=0}^{n-1}$ *and* M. *Suppose that for* $u \in \mathbb{R}$,

$$\mathbb{E}\left[\exp\left(uX_i - \frac{1}{2}u^2 M\right)\right] = O(1),$$

and $M = O_p(r_n)$ *for some rate* $r_n > 0$. *Then*

$$\max_i |X_i| = O_p\big(r_n^{1/2} \log(n)^{1/2}\big).$$

Our next technical result will bound the moments of our observations Y_i, and their normalisations $Z_i(\theta)$. The result will be stated using the Hölder spaces C^s, defined as follows. Given a function $f : X \to \mathbb{R}$, for suitable $X \subseteq \mathbb{R}^d$, we define the 1-Hölder norm

$$\|f\|_{C^1} := \|f\|_\infty \vee \sup_{x,y \in X} \big|f(x) - f(y)\big| / \|x - y\|,$$

and the 2-Hölder norm

$$\|f\|_{C^2} := \begin{cases} \|f\|_\infty \vee \max_{i=1}^{d} \big\|(\nabla f)_i\big\|_{C^1}, & f \text{ is differentiable,} \\ \infty, & \text{otherwise.} \end{cases}$$

We also say f is C^s if $\|f\|_{C^s} < \infty$.

LEMMA 3. *Under* H_0 *or* H_1, *suppose the* $\widehat{X}_i = O(1)$, *and* Θ *is bounded*:

(i) *For fixed* i *and* Y_i, *the variables* $Z_i(\theta)$ *are* C^1 *functions of* θ *and* $\widehat{X}_i$, *with Hölder norm* $O(1 + |Y_i|)$.

(ii) *The variables* $S_t(\theta)$ *are* C^1 *functions of* θ, t, μ_t *and* X_t, *and for fixed* θ *and* t, *also* C^2 *functions of* μ_t *and* X_t, *both with Hölder norm* $O(1)$.

(iii) *For* $\theta \in \Theta$, *we have*

$$\mathbb{E}\big[Z_i(\theta) | \mathcal{F}_{t_i}\big] = S_{t_i}(\theta) + O\big(n^{-1/2}\big),$$

$$\mathbb{E}\big[\big|Z_i(\theta)\big|^{4+\varepsilon} | \mathcal{F}_{t_i}\big] = O(1),$$

and under H_0, *also*

$$\mathbb{E}\big[Z_i(\theta_0)^2 | \mathcal{F}_{t_i}\big] = 1 + O\big(n^{-1/4}\big).$$

(iv) *Define times*

$$s_k := \lceil n 2^{-J} k \rceil / n, \qquad k = 0, \dots, 2^J. \tag{9}$$

Then

$$\max_k n^{-1/2} \sum_{i=ns_k}^{ns_{k+1}-1} Y_i^2 = O_p(1).$$

Finally, we will need a result controlling the behaviour of the processes $S_t(\theta)$ under Assumption 2.

LEMMA 4. *Under H_1, suppose Θ is bounded, let $\underline{\Theta}_n \subseteq \Theta$ be a sequence of finite sets, of size $O(n^\kappa)$ for some $\kappa \geq 0$, and let $\delta_n = O(n^{-1/2})$. Given Assumption 2, we have*

$$S_t(\theta) = \widetilde{S}_t(\theta) + \overline{S}_t(\theta),$$

where the processes $\widetilde{S}_t(\theta)$ and $\overline{S}_t(\theta)$ are as follows:

(i) *We have*

$$\sup_{\theta \in \underline{\Theta}_n, |s-t| \leq \delta_n} \left|\widetilde{S}_s(\theta) - \widetilde{S}_t(\theta)\right| = O_p\left(n^{-1/4} \log(n)^{1/2}\right).$$

(ii) *In $L^2([0, 1])$, let $P_J f$ denote the orthogonal projection of f onto the subspace spanned by the scaling functions $\varphi_{J,k}$, and define the remainder $R_J f := f - P_J f$. Then*

$$\sup_{\theta \in \underline{\Theta}_n} \|R_J \widetilde{S}(\theta)\|_\infty = O_p\left(n^{-1/4} \log(n)^{1/2}\right).$$

(iii) *We have a random variable $N \in \mathbb{N}$, and random times $0 = \tau_0 < \cdots < \tau_N = 1$, such that the processes $\overline{S}_t(\theta)$, $\theta \in \underline{\Theta}_n$, are constant on intervals $[\tau_i, \tau_{i+1})$, $[\tau_{N-1}, \tau_N]$, and*

$$\mathbb{P}\Big[\min_i (\tau_{i+1} - \tau_i) < \delta_n\Big] \to 0.$$

5.2. *Main proofs*. We may now proceed with our main proofs. We first prove Theorem 1, beginning with a lemma controlling the variance of our estimated scaling coefficients $\widehat{\alpha}_{J,k}(\theta)$.

LEMMA 5. *For $k = 0, \ldots, 2^J - 1$, $\theta \in \Theta$, define scaling-coefficient variance terms*

$$\widetilde{\alpha}_{J,k}(\theta) := n^{-1} \sum_{i=0}^{n-1} \varphi_{J,k}(t_i)\big(Z_i(\theta) - \mathbb{E}[Z_i(\theta)|\mathcal{F}_{t_i}]\big).$$

(i) *Under H_0, suppose the $\widehat{X}_i = O(1)$. Then on a suitably-extended probability space, we have a filtration $(\mathcal{G}_k)_{k=0}^{2^J}$, and $\mathcal{G}_{k+1}$-measurable real random variables ξ_k, η_k, M_k, such that*

$$n^{1/2} \widetilde{\alpha}_{J,k}(\theta_0) = \xi_k + \eta_k;$$

the variables ξ_k are standard Gaussian given $\mathcal{G}_k$;

$$\mathbb{E}\left[\exp\left(u\eta_k - \frac{1}{2}u^2 M_k\right)\Big|\mathcal{G}_k\right] \leq 1;$$

and the variables $M_k \geq 0$ satisfy

$$\mathbb{E}\big[M_k^{2+\varepsilon/2}|\mathcal{G}_k\big] = O\big(n^{-(1/2+\varepsilon/8)}\big). \tag{10}$$

(ii) *Under H_1, suppose Θ is bounded, and the $\widehat{X}_i = X_{t_i}$. We then have constants $A_k = O(1)$, and on a suitably-extended probability space, a filtration $(\mathcal{G}_k)_{k=0}^{2^J}$ and real random variables M_k, such that*

$$\sup_{\theta\in\Theta} \mathbb{E}\bigg[\exp\bigg(un^{1/2}\widetilde{\alpha}_{J,k}(\theta) - \frac{1}{2}u^2(A_k + M_k)\bigg)\bigg|\mathcal{G}_k\bigg] \leq 1;$$

the variables $M_k \geq 0$ satisfy (10); *and the $\widetilde{\alpha}_{J,k}(\theta)$ and M_k are $\mathcal{G}_{k+1}$-measurable.*

PROOF. We first prove part (i), and argue by induction on k. Let $\mathcal{G}_0 = \mathcal{F}_0$, and suppose that for $i = 0, \dots, k-1$ we have constructed, on an extended probability space, σ-algebras $\mathcal{G}_{i+1}$, and random variables ξ_i, η_i, M_i satisfying our conditions. We suppose also that $\mathcal{G}_k$ has been chosen to be independent of $\mathcal{F}$ given $\mathcal{F}_{s_k}$, where the times s_k are given by (9); we note this condition is trivially satisfied for $\mathcal{G}_0$.

We can then write

$$n^{1/2}\widetilde{\alpha}_{J,k}(\theta_0) = \sum_{i=ns_k}^{ns_{k+1}-1} \zeta_i,$$

where the $m := n(s_{k+1} - s_k)$ summands

$$\zeta_i := n^{-1/2}2^{J/2}\big(Z_i(\theta_0) - \mathbb{E}\big[Z_i(\theta_0)|\mathcal{F}_{t_i}\big]\big).$$

To compute the moments of the ζ_i, we may apply Lemma 3(iii), noting that since we are only interested in $\theta = \theta_0$, we may assume Θ is bounded. We thus have

$$\mathbb{E}[\zeta_i|\mathcal{F}_{t_i}, \mathcal{G}_k] = 0,$$

$$\sum_{i=ns_k}^{ns_{k+1}-1} \mathbb{E}\big[\zeta_i^2|\mathcal{F}_{t_i}, \mathcal{G}_k\big] = 1 + O\big(m^{-1/2}\big),$$

$$\sum_{i=ns_k}^{ns_{k+1}-1} \mathbb{E}\big[|\zeta_i|^{4+\varepsilon}|\mathcal{F}_{t_i}, \mathcal{G}_k\big] = O\big(m^{-(1+\varepsilon/2)}\big),$$

using also that the ζ_i are independent of $\mathcal{G}_k$ given $\mathcal{F}_{t_i}$.

We may therefore apply Lemma 1(i) to the variables $n^{1/2}\widetilde{\alpha}_{J,k}(\theta_0)$. On a further-extended probability space, we obtain random variables ξ_k, η_k, M_k satisfying the conditions of part (i), independent of $\mathcal{F}$ given $\mathcal{G}_k$ and $\mathcal{F}_{s_{k+1}}$. Defining $\mathcal{G}_{k+1}$ to be the σ-algebra generated by $\mathcal{G}_k$, $\mathcal{F}_{s_{k+1}}$, ξ_k, η_k and M_k, we deduce that $\mathcal{G}_{k+1}$ satisfies the conditions of our inductive hypothesis. By induction, we conclude that part (i) of our result holds.

To prove part (ii), we argue similarly, noting that the random variables

$$n^{1/2}\widetilde{\alpha}_{J,k}(\theta) = \sum_{i=ns_k}^{ns_{k+1}-1} c_i(\theta)\widetilde{\zeta}_i,$$

where the $\mathcal{F}_{t_{i+1}}$-measurable summands

$$\widetilde{\zeta}_i := n^{-1/2}2^{J/2}\big(Y_i - \mathbb{E}[Y_i|\mathcal{F}_{t_i}]\big),$$

and the $\mathcal{F}_{t_i}$-measurable coefficients

$$c_i(\theta) := 1/\sigma(\theta, t_i, X_{t_i}).$$

As the function σ is continuous and positive, and θ and X_t are bounded, we have the variables $c_i(\theta) = O(1)$. We may thus apply Lemma 1(ii), producing random variables A_k, M_k satisfying the conditions of part (ii). The result then follows as before. $\square$

We now prove a lemma bounding the variance of our estimated scaling and wavelet coefficients $\widehat{\alpha}_{0,0}(\theta)$, $\widehat{\beta}_{j,k}(\theta)$.

LEMMA 6. *Suppose the $\widehat{X}_i = O(1)$, and for $j = 0, \dots, J-1$, $k = 0, \dots, 2^j - 1$ and $\theta \in \Theta$, define the wavelet-coefficient variance terms*

$$\widetilde{\beta}_{j,k}(\theta) := n^{-1}\sum_{i=0}^{n-1} \psi_{j,k}(t_i)\big(Z_i(\theta) - \mathbb{E}\big[Z_i(\theta)|\mathcal{F}_{t_i}\big]\big).$$

Similarly, define scaling-coefficient variance terms $\widetilde{\alpha}_{0,0}(\theta)$ using $\varphi_{0,0}$.

(i) *Under H_0, suppose $\widehat{\theta} - \theta_0 = O(n^{-1/2})$. Then on a suitably-extended probability space, we have real random variables $\widetilde{\xi}_{j,k}, \widetilde{\eta}_{j,k}, \widetilde{\upsilon}_{j,k}$ such that*

$$n^{1/2}\widetilde{\alpha}_{0,0}(\widehat{\theta}) = \widetilde{\xi}_{-1,0} + \widetilde{\eta}_{-1,0} + \widetilde{\upsilon}_{-1,0};$$
$$n^{1/2}\widetilde{\beta}_{j,k}(\widehat{\theta}) = \widetilde{\xi}_{j,k} + \widetilde{\eta}_{j,k} + \widetilde{\upsilon}_{j,k};$$

the $\widetilde{\xi}_{j,k}$ are independent standard Gaussian; and for some $\varepsilon' > 0$,

$$\max_{j,k} |\widetilde{\eta}_{j,k}| = O_p\big(n^{-\varepsilon'}\big), \qquad \max_{j,k} 2^{j/2}|\widetilde{\upsilon}_{j,k}| = O_p(1).$$

(ii) *Under H_1, suppose Θ is bounded. Then*

$$\sup_{j,k,\theta\in\Theta} \big|\widetilde{\alpha}_{0,0}(\theta)\big|, \big|\widetilde{\beta}_{j,k}(\theta)\big| = O_p\big(n^{-1/2}\log(n)^{1/2}\big).$$

PROOF. We will consider the wavelet-coefficient variance terms $\widetilde{\beta}_{j,k}(\theta)$; we note we may include scaling-coefficient variance terms $\widetilde{\alpha}_{0,0}(\theta)$ similarly. To prove part (i), we then apply Lemma 5(i). We obtain a filtration $\mathcal{G}_l$, and variables M_l, ξ_l and η_l as in the statement of the lemma. Since

$$\widetilde{\beta}_{j,k}(\theta) = \sum_l b_{j,k,l} \widetilde{\alpha}_{J,l}(\theta),$$

where the coefficients

$$b_{j,k,l} := \int_0^1 \psi_{j,k} \varphi_{J,l},$$

we have

$$n^{1/2} \widetilde{\beta}_{j,k}(\widehat{\theta}) = \widetilde{\xi}_{j,k} + \widetilde{\eta}_{j,k} + \widetilde{\upsilon}_{j,k},$$

for terms

$$\widetilde{\xi}_{j,k} := \sum_l b_{j,k,l} \xi_l, \qquad \widetilde{\eta}_{j,k} := \sum_l b_{j,k,l} \eta_l$$

and

$$\widetilde{\upsilon}_{j,k} := n^{1/2} \big(\widetilde{\beta}_{j,k}(\widehat{\theta}) - \widetilde{\beta}_{j,k}(\theta_0) \big).$$

We first describe the terms $\widetilde{\xi}_{j,k}$. Since the ξ_l are jointly centred Gaussian, so are the $\widetilde{\xi}_{j,k}$. Furthermore, we have

$$\begin{aligned} \mathbb{C}\mathrm{ov}[\widetilde{\xi}_{j,k} \widetilde{\xi}_{j',k'}] &= \sum_l b_{j,k,l} b_{j',k',l} \\ &= \int_0^1 \Big(\sum_l b_{j,k,l} \varphi_{J,l} \Big) \Big(\sum_{l'} b_{j',k',l'} \varphi_{J,l'} \Big) \\ &= \int_0^1 \psi_{j,k} \psi_{j',k'} \\ &= 1_{(j,k)=(j',k')}. \end{aligned} \tag{11}$$

We deduce that the $\widetilde{\xi}_{j,k}$ are independent standard Gaussian.

We next bound the $\widetilde{\eta}_{j,k}$. Setting

$$M := \max_l M_l,$$

we have that

$$\mathbb{E}[M^{2+\varepsilon/2}] \le \sum_l \mathbb{E}[M_l^{2+\varepsilon/2}] = O\big(n^{-\varepsilon/8}\big),$$

so $M = O_p(n^{-\varepsilon'})$ for some $\varepsilon' > 0$. Using (11), we also have

$$\begin{aligned} \mathbb{E}\Big[\exp\Big(u \widetilde{\eta}_{j,k} - \frac{1}{2} u^2 M \Big) \Big] &\le \mathbb{E}\Big[\prod_l \exp\Big(u b_{j,k,l} \eta_l - \frac{1}{2} u^2 b_{j,k,l}^2 M_l \Big) \Big] \\ &\le 1. \end{aligned}$$

The desired result follows by applying Lemma 2.

Finally, we control the $\widetilde{\upsilon}_{j,k}$. Since we are only interested in $\theta = \theta_0, \widehat{\theta}$, we may assume Θ is bounded. For $\theta, \theta' \in \Theta$, $|\theta - \theta'| = O(n^{-1/2})$, we then have

$$\sup_{j,k,\theta,\theta'} 2^{j/2} \bigl|\widetilde{\beta}_{j,k}(\theta) - \widetilde{\beta}_{j,k}(\theta')\bigr|$$

$$= \max_{j,k} O\bigl(n^{-3/2} 2^{j/2}\bigr) \sum_{i=0}^{n-1} \bigl|\psi_{j,k}(t_i)\bigr| \bigl(1 + |Y_i|\bigr),$$

using Lemma 3(i),

$$= O\bigl(n^{-1/2}\bigr)\left(1 + \max_k n^{-1/2} \sum_{i=ns_k}^{ns_{k+1}-1} |Y_i|\right)$$

$$= O\bigl(n^{-1/2}\bigr)\left(1 + \left(\max_k n^{-1/2} \sum_{i=ns_k}^{ns_{k+1}-1} Y_i^2\right)^{1/2}\right),$$

by Cauchy–Schwarz,

$$= O_p\bigl(n^{-1/2}\bigr), \tag{12}$$

using Lemma 3(iv). We deduce that

$$\sup_{j,k} 2^{j/2} |\widetilde{\upsilon}_{j,k}| = O_p(1).$$

To prove part (ii), we first claim we may assume the $\widehat{X}_i = X_{t_i}$. To prove the claim, we define terms

$$Z_i'(\theta) := \bigl(Y_i - \mu(\theta, t_i, X_{t_i})\bigr)/\sigma(\theta, t_i, X_{t_i})$$

and

$$\widetilde{\beta}'_{j,k}(\theta) := n^{-1} \sum_{i=0}^{n-1} \psi_{j,k}(t_i) \bigl(Z_i'(\theta) - \mathbb{E}[Z_i'(\theta) | \mathcal{F}_{t_i}]\bigr).$$

We then have

$$\sup_{j,k,\theta\in\Theta} \bigl|\widetilde{\beta}_{j,k}(\theta) - \widetilde{\beta}'_{j,k}(\theta)\bigr|$$

$$= O\bigl(n^{-1}\bigr) \max_{j,k} \sum_{i=0}^{n-1} \bigl|\psi_{j,k}(t_i)\bigr| \bigl(1 + |Y_i|\bigr) \|\widehat{X}_i - X_{t_i}\|,$$

using Lemma 3(i),

$$= O\bigl(n^{-1/2}\bigr)\left(\max_{j,k} n^{-1} \sum_{i=0}^{n-1} \psi_{j,k}^2(t_i)\bigl(1 + Y_i^2\bigr)\right)^{1/2}$$

$$\times \left(\sum_{i=0}^{n-1} \|\widehat{X}_i - X_{t_i}\|^2\right)^{1/2},$$

by Cauchy–Schwarz,

$$= O_p(n^{-1/2})\left(1 + \max_k n^{-1/2} \sum_{i=ns_k}^{ns_{k+1}-1} Y_i^2\right)^{1/2},$$

since $\mathbb{E}[\sum_{i=0}^{n-1} \|\widehat{X}_i - X_{t_i}\|^2] = O(1)$,

$$= O_p(n^{-1/2}),$$

using Lemma 3(iv).

We may thus assume the $\widehat{X}_i = X_{t_i}$, and so apply Lemma 5(ii). On an extended probability space, we obtain a filtration $\mathcal{G}_l$, constants $A_l = O(1)$, and variables M_l as in the statement of the lemma. Setting

$$M := \max_l (A_l + M_l),$$

we obtain that $M = O_p(1)$, and

$$\sup_{\theta \in \Theta} \mathbb{E}\left[\exp\left(un^{1/2}\widetilde{\beta}_{j,k}(\theta) - \frac{1}{2}u^2 M\right)\right] \leq 1,$$

arguing as in part (i). Letting $\underline{\Theta}_n$ denote a $n^{-1/2}$-net for $\Theta \subset \mathbb{R}^p$, of size $O(n^{p/2})$, we thus have

$$\begin{aligned}\max_{j,k,\theta \in \underline{\Theta}_n} |\widetilde{\beta}_{j,k}(\theta)| &= O_p(n^{-1/2} \log(n^{p/2})^{1/2}) \\ &= O_p(n^{-1/2} \log(n)^{1/2}),\end{aligned}$$

using Lemma 2.

Next, for any $\theta \in \Theta$, we have a point $\underline{\theta} \in \underline{\Theta}_n$ with $\theta - \underline{\theta} = O(n^{-1/2})$. Using (12), we deduce that

$$\sup_{j,k,\theta \in \Theta} |\widetilde{\beta}_{j,k}(\theta) - \widetilde{\beta}_{j,k}(\underline{\theta})| = O_p(n^{-1/2}).$$

We conclude that

$$\sup_{j,k,\theta \in \Theta} |\widetilde{\beta}_{j,k}(\theta)| = O_p(n^{-1/2} \log(n)^{1/2}).$$

□

Next, we prove a lemma bounding the bias of our estimated scaling and wavelet coefficients $\widehat{\alpha}_{0,0}(\theta)$, $\widehat{\beta}_{j,k}(\theta)$.

LEMMA 7. *Suppose the $\widehat{X}_i = O(1)$, and for $j = 0, \ldots, J-1$, $k = 0, \ldots, 2^J - 1$ and $\theta \in \Theta$, define the wavelet-coefficient bias terms*

$$\overline{\beta}_{j,k}(\theta) := n^{-1} \sum_{i=0}^{n-1} \psi_{j,k}(t_i) \mathbb{E}[Z_i(\theta) | \mathcal{F}_{t_i}] - \beta_{j,k}(\theta).$$

Similarly, define scaling-coefficient bias terms $\overline{\alpha}_{0,0}(\theta)$ using $\varphi_{0,0}$:

(i) *Under* H_0, *suppose* $\widehat{\theta} - \theta_0 = O(n^{-1/2})$. *Then*

$$\max_{j,k} |\overline{\alpha}_{0,0}(\widehat{\theta})|, 2^{j/2}|\overline{\beta}_{j,k}(\widehat{\theta})| = O_p(n^{-1/2}).$$

(ii) *Under* H_1, *suppose* Θ *is bounded. Then*

$$\sup_{j,k,\theta\in\Theta} |\overline{\alpha}_{0,0}(\theta)|, |\overline{\beta}_{j,k}(\theta)| = O_p(n^{-1/2}).$$

PROOF. We will bound the wavelet-coefficient bias terms $\overline{\beta}_{j,k}(\theta)$; we note we may include the scaling-coefficient bias terms $\overline{\alpha}_{0,0}(\theta)$ similarly. For $t \in [0, 1]$, define $\underline{t} := \lfloor nt \rfloor / n$, and set

$$\underline{\beta}_{j,k}(\theta) := \int_0^1 \psi_{j,k}(t)\big(S_{\underline{t}}(\theta) - S_t(\theta)\big)\,dt.$$

In each part (i) and (ii), we will show that $\overline{\beta}_{j,k}(\theta)$ is close to $\underline{\beta}_{j,k}(\theta)$, which is small.

We note that in either part we may assume Θ is bounded, since in part (i), we are only interested in $\theta = \theta_0, \widehat{\theta}$. We then have

$$\begin{aligned} |\overline{\beta}_{j,k}(\theta) - \underline{\beta}_{j,k}(\theta)| &\le n^{-1}\sum_{i=0}^{n-1} |\psi_{j,k}(t_i)||\mathbb{E}[Z_i(\theta)|\mathcal{F}_{t_i}] - S_{t_i}(\theta)| \\ &\quad + \int_0^1 |\psi_{j,k}(t) - \psi_{j,k}(\underline{t})||S_{\underline{t}}(\theta)|\,dt \\ &= O(n^{-1/2}2^{-j/2}), \end{aligned} \tag{13}$$

using Lemma 3(ii) and (iii). It thus remains to bound the $\underline{\beta}_{j,k}(\theta)$.

To prove part (i), we note that

$$\underline{\beta}_{j,k}(\theta_0) = \sum_{i=0}^{n-1} \zeta_{i,j,k},$$

where the $\mathcal{F}_{t_{i+1}}$-measurable summands

$$\zeta_{i,j,k} := -\int_{t_i}^{t_{i+1}} \psi_{j,k}(t)\big(S_t(\theta_0) - S_{t_i}(\theta_0)\big)\,dt.$$

Using Lemma 3(ii) and Taylor's theorem, we also have that

$$\begin{aligned} S_t(\theta_0) - S_{t_i}(\theta_0) &= c_i(\mu_t - \mu_{t_i}) + d_i^T(X_t - X_{t_i}) \\ &\quad + O\big(|\mu_t - \mu_{t_i}|^2 + \|X_t - X_{t_i}\|^2 + n^{-1}\big), \end{aligned}$$

for bounded $\mathcal{F}_{t_i}$-measurable random variables $c_i \in \mathbb{R}$, $d_i \in \mathbb{R}^q$.

We deduce that

$$\mathbb{E}[\zeta_{i,j,k}|\mathcal{F}_{t_i}] = O\big(n^{-2}2^{j/2}\big),$$

and similarly

$$\mathbb{V}\mathrm{ar}[\zeta_{i,j,k}|\mathcal{F}_{t_i}] \le \mathbb{E}[\zeta_{i,j,k}^2|\mathcal{F}_{t_i}] = O(n^{-3}2^j).$$

Furthermore, for fixed j and k, we have that all but $O(n2^{-j})$ of the $\zeta_{i,j,k}$ are almost-surely zero. We thus have

$$\mathbb{E}[\underline{\beta}_{j,k}(\theta_0)^2] = O(n^{-2}).$$

We deduce that

$$\begin{aligned}\mathbb{E}\Big[\max_{j,k}\underline{\beta}_{j,k}(\theta_0)^2\Big] &\le \sum_{j,k}\mathbb{E}[\underline{\beta}_{j,k}(\theta_0)^2] \\ &\le O(n^{-2})\sum_j 2^j \\ &= O(n^{-3/2}),\end{aligned}$$

so $\max_{j,k}|\underline{\beta}_{j,k}(\theta_0)| = O_p(n^{-3/4})$. We also have

$$\underline{\beta}_{j,k}(\theta_0) - \underline{\beta}_{j,k}(\widehat{\theta}) = O(n^{-1/2}2^{-j/2}),$$

using Lemma 3(ii). We conclude that

$$\begin{aligned}\max_{j,k} 2^{j/2}|\overline{\beta}_{j,k}(\widehat{\theta})| &\le \max_{j,k} 2^{j/2}|\underline{\beta}_{j,k}(\theta_0)| \\ &\quad + \max_{j,k} 2^{j/2}|\underline{\beta}_{j,k}(\theta_0) - \underline{\beta}_{j,k}(\widehat{\theta})| \\ &\quad + \max_{j,k} 2^{j/2}|\underline{\beta}_{j,k}(\widehat{\theta}) - \overline{\beta}_{j,k}(\widehat{\theta})| \\ &= O_p(n^{-1/2}),\end{aligned}$$

using (13).

To prove part (ii), using Lemma 3(ii), we have

$$S_{\underline{t}}(\theta) - S_t(\theta) = O(|\mu_{\underline{t}} - \mu_t| + \|X_{\underline{t}} - X_t\| + n^{-1}).$$

We deduce that

$$\begin{aligned}&\sup_{j,k,\theta\in\Theta}|\underline{\beta}_{j,k}(\theta)| \\ &\quad = O(1)\sup_{j,k}\int_0^1 |\psi_{j,k}(t)|(|\mu_{\underline{t}} - \mu_t| + \|X_{\underline{t}} - X_t\| + n^{-1})\,dt \\ &\quad = O(1)\Big(\sup_{j,k}\int_0^1 \psi_{j,k}^2(t)\,dt\Big)^{1/2} \\ &\qquad \times \Big(\int_0^1 (|\mu_{\underline{t}} - \mu_t|^2 + \|X_{\underline{t}} - X_t\|^2 + n^{-2})\,dt\Big)^{1/2},\end{aligned}$$

by Cauchy–Schwarz,

$$= O_p(n^{-1/2}),$$

since $\int_0^1 \psi_{j,k}^2(t)\,dt = 1$, and

$$\mathbb{E}\left[\int_0^1 \left(|\mu_{\underline{t}} - \mu_t|^2 + \|X_{\underline{t}} - X_t\|^2 + n^{-2}\right) dt\right] = O(n^{-1}).$$

Using (13), we conclude that

$$\sup_{j,k,\theta\in\Theta} |\overline{\beta}_{j,k}(\theta)| = O_p(n^{-1/2}).$$

□

We can now prove our limit theorem for the statistic $\widehat{T}(\widehat{\theta})$.

PROOF OF THEOREM 1. We first note that our estimated scaling and wavelet coefficients are equivalently given by

$$\widehat{\alpha}_{0,0}(\theta) = n^{-1}\sum_{i=0}^{n-1} \varphi_{0,0}(t)Z_i(\theta), \qquad \widehat{\beta}_{j,k}(\theta) = n^{-1}\sum_{i=0}^{n-1} \psi_{j,k}(t)Z_i(\theta).$$

We may thus make the variance-bias decomposition

$$\widehat{\alpha}_{0,0}(\theta) - \alpha_{0,0}(\theta) = \widetilde{\alpha}_{0,0}(\theta) + \overline{\alpha}_{0,0}(\theta),$$

$$\widehat{\beta}_{j,k}(\theta) - \beta_{j,k}(\theta) = \widetilde{\beta}_{j,k}(\theta) + \overline{\beta}_{j,k}(\theta),$$

where the terms $\widetilde{\alpha}_{0,0}$, $\overline{\alpha}_{0,0}$, $\widetilde{\beta}_{j,k}$ and $\overline{\beta}_{j,k}$ are defined by Lemmas 6 and 7. We will proceed to bound the distribution of $\widehat{T}(\widehat{\theta})$ using these lemmas.

We begin by showing we may assume the estimated covariates $\widehat{X}_i = O(1)$. We note that

$$\mathbb{E}\left[\max_i \|\widehat{X}_i\|^2\right] \le \mathbb{E}\left[\sup_t \|X_t\|^2\right] + \sum_i \mathbb{E}\left[\|\widehat{X}_i - X_{t_i}\|^2\right] = O(1),$$

so $\max_i \|\widehat{X}_i\| = O_p(1)$. For a constant $R > 0$, define the variables

$$\widetilde{X}_i := \begin{cases} \widehat{X}_i, & \|\widehat{X}_i\| \le R, \\ X_{t_i}, & \text{otherwise.} \end{cases}$$

Then as $R \to \infty$, the probability that the $\widetilde{X}_i$ and $\widehat{X}_i$ agree tends to one, uniformly in n. It thus suffices to prove our results replacing the $\widehat{X}_i$ with the $\widetilde{X}_i$; equivalently, we may assume the $\widehat{X}_i = O(1)$.

We now prove part (i). Since $\widehat{\theta} - \theta_0 = O_p(n^{-1/2})$, we may similarly assume $\widehat{\theta} - \theta_0 = O(n^{-1/2})$. Let $J_2 = \lfloor J/2 \rfloor$, and write

$$\widehat{T}(\theta) = \max\left(\overline{T}(\theta), \widetilde{T}(\theta)\right),$$

where the terms

$$\overline{T}(\theta) := \max_{0 \le j < J_2, k} \bigl|\widehat{\alpha}_{0,0}(\theta)\bigr|, \bigl|\widehat{\beta}_{j,k}(\theta)\bigr|,$$

$$\widetilde{T}(\theta) := \max_{J_2 \le j < J, k} \bigl|\widehat{\beta}_{j,k}(\theta)\bigr|.$$

Under H_0, using Lemmas 6(i) and 7(i), we can then write

$$n^{1/2}\overline{T}(\widehat{\theta}) = \max_{0 \le j < J_2, k} |\widetilde{\xi}_{j,k}| + O_p(1),$$

$$n^{1/2}\widetilde{T}(\widehat{\theta}) = \max_{J_2 \le j < J, k} |\widetilde{\xi}_{j,k}| + O_p(n^{-\varepsilon'}),$$

for some $\varepsilon' > 0$, and independent standard Gaussians $\widetilde{\xi}_{j,k}$.

By standard Gumbel limits, we also have

$$a_{2^{J_2}}^{-1}\Bigl(\max_{0 \le j < J_2, k} |\widetilde{\xi}_{j,k}| - b_{2^{J_2}}\Bigr) \xrightarrow{d} G,$$

$$a_{2^J}^{-1}\Bigl(\max_{J_2 \le j < J, k} |\widetilde{\xi}_{j,k}| - b_{2^J}\Bigr) \xrightarrow{d} G,$$

we note that in the second limit, we may use the constants a_{2^J} and b_{2^J}, rather than $a_{2^J - 2^{J_2}}$ and $b_{2^J - 2^{J_2}}$, as the difference is negligible. We deduce that

$$\mathbb{P}\bigl[\widehat{T}(\widehat{\theta}) = \widetilde{T}(\widehat{\theta})\bigr] \to 1,$$

and so

$$\begin{aligned} a_{2^J}^{-1}\bigl(n^{1/2}\widehat{T}(\widehat{\theta}) - b_{2^J}\bigr) &= a_{2^J}^{-1}\bigl(n^{1/2}\widetilde{T}(\widehat{\theta}) - b_{2^J}\bigr) + o_p(1) \\ &= a_{2^J}^{-1}\Bigl(\max_{J_2 \le j < J, k} |\widetilde{\xi}_{j,k}| - b_{2^J}\Bigr) + o_p(1) \\ &\xrightarrow{d} G. \end{aligned}$$

Next, we prove part (ii). As before, since $\widehat{\theta} = O_p(1)$, we may assume $\widehat{\theta} = O(1)$, and hence that Θ is bounded. Using Lemmas 6(ii) and 7(ii), we then have

$$\begin{aligned} \widehat{T}(\widehat{\theta}) - T(\widehat{\theta}) &= O(1) \max_{0 \le j < J, k} \bigl|\widehat{\alpha}_{0,0}(\widehat{\theta}) - \alpha_{0,0}(\widehat{\theta})\bigr|, \bigl|\widehat{\beta}_{j,k}(\widehat{\theta}) - \beta_{j,k}(\widehat{\theta})\bigr| \\ &= O_p\bigl(n^{-1/2}\log(n)^{1/2}\bigr). \end{aligned}$$

Finally, we note that the rates of convergence proved depend only upon the bounds assumed on the inputs. They therefore hold uniformly over models satisfying our assumptions. □

Next, we will prove our results on test coverage and detection rates.

PROOF OF THEOREM 2. We first note that part (i) is immediate from Theorem 1(i). To prove part (ii), we consider separately the cases (a) and (b). In each case, we will prove that with probability tending to one, the event E_n implies

$$T(\widehat{\theta}) \geq M_n' n^{-1/2} \log(n)^{1/2},$$

for a fixed sequence $M_n' \to \infty$. The result will then follow from Theorem 1(ii).

In case (a), we note that arguing as in Theorem 1, we may assume Θ is bounded. Let $\underline{\Theta}_n$ be an $n^{-1/4}$-net for Θ, of size $O(n^{p/4})$, and $\underline{\widehat{\theta}}$ be an element of $\underline{\Theta}_n$ satisfying

$$\underline{\widehat{\theta}} - \widehat{\theta} = O(n^{-1/4}).$$

Using Lemma 3(ii), we have

$$S_t(\widehat{\theta}) = S_t(\underline{\widehat{\theta}}) + O(n^{-1/4}),$$

so on E_n,

$$\|S(\underline{\widehat{\theta}})\|_\infty \geq \|S(\widehat{\theta})\|_\infty - O(n^{-1/4}) \geq M_n n^{-1/4} \log(n)^{1/2}/2,$$

for large n. We may thus assume further that $\widehat{\theta} \in \underline{\Theta}_n$.

We then apply Lemma 4, obtaining processes $\widetilde{S}_t(\theta)$, $\overline{S}_t(\theta)$ and times τ_i. On the event E_n, for some point $u \in [0, 1]$, we have

$$|S_u(\widehat{\theta})| \geq M_n n^{-1/4} \log(n)^{1/2}.$$

We thus have $u \in [\tau_i, \tau_{i+1})$ for some $i < N - 1$, or $u \in [\tau_i, \tau_{i+1}]$ for $i = N - 1$. From Lemma 4(iii), with probability tending to one we also have

$$\tau_{i+1} - \tau_i \geq 2^{1-J},$$

and so there exists a point $v \in [\tau_i + 2^{-J}, \tau_{i+1} - 2^{-J}]$, $|u - v| \leq 2^{-J}$.

We deduce that with probability tending to one,

$$\begin{aligned}
&|\alpha_{0,0}(\widehat{\theta})| + \sum_{j=0}^{J-1} 2^{j/2} |\beta_{j,2^{-j}\lfloor 2^j v \rfloor}(\widehat{\theta})| \\
&\quad = |\alpha_{0,0}(\widehat{\theta})\varphi_{0,0}(v)| + \sum_{0 \leq j < J, k} |\beta_{j,k}(\widehat{\theta})\psi_{j,k}(v)| \\
&\quad \geq \left|\alpha_{0,0}(\widehat{\theta})\varphi_{0,0}(v) + \sum_{0 \leq j < J, k} \beta_{j,k}(\widehat{\theta})\psi_{j,k}(v)\right| \\
&\quad = |P_J S_v(\widehat{\theta})|,
\end{aligned}$$

writing the projection P_J in terms of the wavelet functions $\psi_{j,k}$,

$$\begin{aligned}
&\geq |S_u(\widehat{\theta})| - |S_v(\widehat{\theta}) - S_u(\widehat{\theta})| - |R_J S_v(\widehat{\theta})| \\
&\geq |S_u(\widehat{\theta})| - |\widetilde{S}_v(\widehat{\theta}) - \widetilde{S}_u(\widehat{\theta})| - |R_J \widetilde{S}_v(\widehat{\theta})|,
\end{aligned}$$

since $\overline{S}_t(\widehat{\theta})$ is constant within a distance 2^{-J} of v,

$$\geq M_n n^{-1/4} \log(n)^{1/2}/2,$$

using Lemma 4(i) and (ii). We deduce that

$$\begin{aligned}
T(\widehat{\theta}) &\geq \max\bigl(\bigl|\alpha_{0,0}(\widehat{\theta})\bigr|, \bigl|\beta_{j,2^{-j}\lfloor 2^j v\rfloor}(\widehat{\theta})\bigr| : j = 0, \dots, J-1\bigr), \\
&\geq 2^{-(J+3)/2}\left(\bigl|\alpha_{0,0}(\widehat{\theta})\bigr| + \sum_{j=0}^{J-1} 2^{j/2}\bigl|\beta_{j,2^{-j}\lfloor 2^j v\rfloor}(\widehat{\theta})\bigr|\right) \\
&\geq M_n' n^{-1/2} \log(n)^{1/2},
\end{aligned}$$

for a sequence $M_n' \to \infty$.

In case (b), on the event E_n, we likewise have

$$\begin{aligned}
&\bigl|\alpha_{0,0}(\widehat{\theta})\bigr| + \sum_{j=0}^{j_n-1} 2^{j/2}\bigl|\beta_{j,2^{-j}\lfloor 2^{j-j_n}k_n\rfloor}(\widehat{\theta})\bigr| \\
&\quad\geq \bigl|P_{j_n} S_{2^{-j_n}k_n}(\widehat{\theta})\bigr| \\
&\quad= 2^{j_n}\left|\int_{2^{-j_n}k_n}^{2^{-j_n}(k_n+1)} S_t(\widehat{\theta})\,dt\right| \\
&\quad\geq M_n 2^{j_n/2} n^{-1/2} \log(n)^{1/2},
\end{aligned}$$

for some $j_n = 0, \dots, J$ and k_n. The result then follows as in part (i). $\square$

Finally, we can prove our lower bound on detection rates.

PROOF OF THEOREM 3. In each case (i) and (ii), we will reduce the statement to a known testing inequality. We will consider the model

$$Y_i := \delta_n^{1/2} n^{-1/2} 2^{j_n}(B_{t_i \vee \tau} - B_\tau) + \varepsilon_i,$$

where B_t is an adapted Brownian motion, the independent $\mathcal{F}_{t_{i+1}}$-measurable variables ε_i are standard Gaussian given $\mathcal{F}_{t_i}$, $\tau \in [0, 1]$ is to be defined, and in case (i) we set $j_n := J$. It can be checked that this model satisfies our assumptions.

Under H_0, we set $\tau := 1$, so we have mean and variance functions

$$\mu := 0, \qquad \sigma^2 := 1.$$

Under H_1, we instead set $\tau := t_m$, where $m := \lfloor n(1 - 2^{-j_n})\rfloor$. We then have

$$S_t = \delta_n^{1/2} n^{-1/2} 2^{j_n}(B_{t\vee\tau} - B_\tau),$$

so in case (i),

$$\mathbb{P}[E_n] = \mathbb{P}\bigl[\|S\|_\infty \geq \delta_n n^{-1/4}\bigr] \to 1.$$

Similarly, in case (ii),

$$\mathbb{P}[E_n] \geq \mathbb{P}\left[2^{j_n/2}\left|\int_{1-2^{-j_n}}^{1} S_t\,dt\right| \geq \delta_n n^{-1/2}\right] \to 1.$$

It remains to show that no sequence of critical regions C_n can satisfy $\limsup_n \mathbb{P}[C_n] < 1$ under H_0, and $\mathbb{P}[C_n] \to 1$ under H_1. We note that under H_0, we have $Y \sim N(0, I)$, while under H_1, $Y \sim N(0, I + \delta_n \Sigma)$, for a covariance matrix

$$\Sigma_{k,l} = 0 \vee 2^{2j_n}(k \wedge l - m)/n^2.$$

As Σ is nonnegative definite, and has Frobenius norm $O(1)$, the result follows from Lemma 2.1 of Munk and Schmidt-Hieber (2010). □

Acknowledgements. All research data was randomly generated using software given in Bull (2016a).

SUPPLEMENTARY MATERIAL

Supplement to "Semimartingale detection and goodness-of-fit tests" (DOI: 10.1214/16-AOS1484SUPP; .pdf). Proofs of technical results.

REFERENCES

AÏT-SAHALIA, Y. (1996). Testing continuous-time models of the spot interest rate. *Rev. Financ. Stud.* **9** 385–426.

AÏT-SAHALIA, Y. and PARK, J. Y. (2012). Stationarity-based specification tests for diffusions when the process is nonstationary. *J. Econometrics* **169** 279–292. MR2947885

ANDERSEN, T. G., BOLLERSLEV, T., DIEBOLD, F. X. and EBENS, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics* **61** 43–76.

BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 253–280. MR1904704

BARNDORFF-NIELSEN, O. E. and VERAART, A. (2009). Stochastic volatility of volatility in continuous time. *CREATES Research Paper* **25**.

BIBINGER, M., JIRAK, M. and VETTER, M. (2015). Nonparametric change-point analysis of volatility. Preprint. Available at arXiv:1502.00043.

BULL, A. D. (2016a). Software for "Semimartingale detection and goodness-of-fit tests". Available at http://dx.doi.org/10.17863/CAM.125.

BULL, A. D. (2016b). Supplement to "Semimartingale detection and goodness-of-fit tests." DOI:10.1214/16-AOS1484SUPP.

CHEN, Q., ZHENG, X. and PAN, Z. (2015). Asymptotically distribution-free tests for the volatility function of a diffusion. *J. Econometrics* **184** 124–144. MR3277557

CONT, R. and TANKOV, P. (2004). *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, Boca Raton, FL. MR2042661

CORRADI, V. and WHITE, H. (1999). Specification tests for the variance of a diffusion. *J. Time Series Anal.* **20** 253–270. MR1693173

DETTE, H. and PODOLSKIJ, M. (2008). Testing the parametric form of the volatility in continuous time diffusion models—A stochastic process approach. *J. Econometrics* **143** 56–73. MR2384433

DETTE, H., PODOLSKIJ, M. and VETTER, M. (2006). Estimation of integrated volatility in continuous-time financial models with applications to goodness-of-fit testing. *Scand. J. Stat.* **33** 259–278. MR2279642

DETTE, H. and VON LIERES UND WILKAU, C. (2003). On a test for a parametric form of volatility in continuous time financial models. *Finance Stoch.* **7** 363–384. MR1994914

DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: Asymptopia? *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 301–369. MR1323344

GONZÁLEZ-MANTEIGA, W. and CRUJEIRAS, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST* **22** 361–411. MR3093195

HOFFMANN, M., MUNK, A. and SCHMIDT-HIEBER, J. (2012). Adaptive wavelet estimation of the diffusion coefficient under additive error measurements. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 1186–1216. MR3052408

INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models. Lecture Notes in Statistics* **169**. Springer, New York. MR1991446

JACOD, J. and REISS, M. (2014). A remark on the rates of convergence for integrated volatility estimation in the presence of jumps. *Ann. Statist.* **42** 1131–1144. MR3224283

JACOD, J., LI, Y., MYKLAND, P. A., PODOLSKIJ, M. and VETTER, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.* **119** 2249–2276. MR2531091

KLEINOW, T. (2002). Testing the diffusion coefficient. Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes.

LEE, S. (2006). The Bickel–Rosenblatt test for diffusion processes. *Statist. Probab. Lett.* **76** 1494–1502. MR2245570

LEE, S. and WEE, I.-S. (2008). Residual empirical process for diffusion processes. *J. Korean Math. Soc.* **45** 683–693. MR2410237

MANCINI, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scand. J. Stat.* **36** 270–296. MR2528985

MUNK, A. and SCHMIDT-HIEBER, J. (2010). Lower bounds for volatility estimation in microstructure noise models. In *Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown. Inst. Math. Stat. Collect.* **6** 43–55. IMS, Beachwood, OH. MR2798510

MYKLAND, P. A. and ZHANG, L. (2006). ANOVA for diffusions and Itô processes. *Ann. Statist.* **34** 1931–1963. MR2283722

NGUYEN, C. M. (2010). Time-changed residual based tests for asset pricing models in continuous time. Unpublished manuscript, Indiana Univ., Bloomington, IN.

PAPANICOLAOU, A. and GIESECKE, K. (2014). Variation-based tests for volatility misspecification. Available at SSRN 2450776.

PAPAPANTOLEON, A. (2008). An introduction to Lévy processes with applications in finance. Preprint. Available at arXiv:0804.0482.

PODOLSKIJ, M. and ZIGGEL, D. (2008). A range-based test for the parametric form of the volatility in diffusion models. CREATES Research Paper 22.

REISS, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *Ann. Statist.* **39** 772–802. MR2816338

REISS, M., TODOROV, V. and TAUCHEN, G. (2014). Nonparametric test for a constant beta over a fixed time interval. Preprint. Available at arXiv:1403.0349.

VETTER, M. (2012). Estimation of integrated volatility of volatility with applications to goodness-of-fit testing. Preprint. Available at arXiv:1206.5761.

VETTER, M. and DETTE, H. (2012). Model checks for the volatility under microstructure noise. *Bernoulli* **18** 1421–1447. MR2995803

STATISTICAL LABORATORY
UNIVERSITY OF CAMBRIDGE
WILBERFORCE ROAD
CAMBRIDGE CB3 0WB
UNITED KINGDOM
E-MAIL: a.bull@statslab.cam.ac.uk