Bernoulli **10**(1), 2004, 1–24

Counting process intensity estimation by orthogonal wavelet methods

PRAKASH N. PATIL[1] and ANDREW T.A. WOOD[2]

[1]*School of Mathematics and Statistics, University of Birmingham, Birmingham, B15 2TT, UK. E-mail: pnp@for.mat.bham.ac.uk*
[2]*School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, UK. E-mail: andrew.wood@nottingham.ac.uk*

In recent years wavelet-based methods have been proposed to estimate density and regression functions, most often in settings with independent and identically distributed observations. An attractive feature of these methods is that the rate of convergence is unaffected by the presence of discontinuities in the function being estimated. We provide structure and mean-square analyses of wavelet-based estimators of counting process intensities in the context of the multiplicative intensity model.

Keywords: counting process; mean integrated square error; multiplicative intensity model; smoothness; wavelet thresholding

1. Introduction

Wavelet-based methods provide an attractive and powerful tool for the nonparametric estimation of objects with spatially variable smoothness properties. This is primarily due to the local adaptability to different levels of smoothness in the target function provided by wavelet thresholding. A number of authors have studied wavelet methods in various settings, especially density and regression function estimation. In much of this work, it is assumed that observations (or errors) are independent and identically distributed (i.i.d.), but in some cases dependent observations have been considered. Recent work on statistical applications of wavelets includes Donoho *et al.* (1995; 1996), Kerkyacharian and Picard (1993), Donoho and Johnstone (1994; 1998), Hall and Patil (1995; 1996), Hall *et al.* (1996), Johnstone and Silverman (1997), Marron *et al*. (1998) and Truong and Patil (1996).

Here we consider wavelet-based estimation in Aalen's multiplicative intensity model for counting processes. In this context, observations are neither independent nor identically distributed. Our principal objective is to establish the mean integrated square error (MISE) properties of wavelet-based estimators of the intensity function.

Our main results are formulated in a similar way to Hall and Patil's (1995) results concerning MISE properties of wavelet-based density estimators based on i.i.d. observations. However, a number of new technical issues arise in the proofs of analogous results in the counting process framework. Here, key roles are played by two inequalities for point process martingales. One of these is an exponential probability inequality due to Courbot

1350–7265

(1996) which is analogous to Bennett's inequality in the independent case, but allows random jump sizes. The other is a point process analogue of Rosenthal's inequality; see Wood (1999; 2001) and references therein for further details. These inequalities are stated in Section 3.1.

The kernel-based estimator of an intensity function α was first proposed and studied in the counting process context by Ramlau-Hansen (1983a; 1983b). There he showed that the MISE is of the form

$$\text{MISE} = C_1(nb_n)^{-1} + C_2 b_n^{2r} + o\{(nb_n)^{-1} + b_n^{2r}\}, \tag{1.1}$$

where n denotes sample size, b_n is the bandwidth of the kernel-based estimator, $r \geqslant 2$ is the order of the kernel, and C_1 and C_2 are constants depending on both the kernel and the unknown intensity function. The first term on the right of (1.1) is associated with variance and the second is associated with squared bias. The constant C_2 in the squared bias term is proportional to the integral of the square of the rth derivative of the intensity function, and the MISE expansion for the kernel-based estimator generally fails if α does not have r continuous derivatives. We show that an analogue of (1.1) also holds in the case of nonlinear wavelet estimators, and that it remains valid even if the target intensity function is only piecewise continuous. Thus, as in the case of density estimation based on i.i.d. observations, wavelet methods enjoy good convergence rates in the counting processes framework even when smoothness conditions on the target function are imposed only in a piecewise sense.

Kolaczyk (1999) has also considered wavelet-based estimators of intensity in a counting process framework. To estimate intensity functions of a certain class of burst-like Poisson processes, he proposes an approach based on wavelet shrinkage. An interesting proposal in his paper is to use (asymmetric) upper and lower thresholds of standard order $(n^{-1} \log n)^{1/2}$ to take account of the skewness of the Poisson distribution. Kolaczyk's paper and the developments here are complementary in that we focus on MISE properties of the estimators, whereas he focuses more directly on thresholding issues.

Our main results, Theorems 2.1 and 2.2, are described in Section 2. Auxiliary results required to prove the main results are given in Section 3, and the main results are proved in Section 4.

2. Main results

In Section 2.1 we briefly summarize basic properties of wavelets, and in Section 2.2 we explain how wavelet estimators of the (deterministic component of the) intensity of a counting process can be constructed. In Section 2.3 we introduce our basic assumptions concerning the random component of the counting process intensity (the Y process) and provide discussion of the assumptions made. In Section 2.5 these assumptions are discussed in the context of an example. Our main results, Theorems 2.1 and 2.2, are given in Section 2.4.

2.1. Wavelet transform

Let ψ and ϕ respectively denote mother and father wavelet functions of the rth order, possessing the properties $\int \psi^2 = \int \phi^2 = 1$, $\mu_k \equiv \int x^k \psi(x) \mathrm{d}x = 0$ for $0 \leqslant k \leqslant r-1$, and $\mu_r = r!\kappa$ (say) $\neq 0$. Furthermore, for arbitrary $p > 0$, and defining $p_i = p2^i$ for $i \geqslant 0$, the class of functions $\phi_j(x) = p^{1/2}\phi(px - j)$ and $\psi_{ij}(x) = p_i^{1/2}\psi(p_i x - j)$ form an orthonormal basis for the class of square-integrable functions f. The orthogonality relations may be expressed by $\int \phi_{j_1}\phi_{j_2} = \delta_{j_1 j_2}$, $\int \psi_{i_1 j_1}\psi_{i_2 j_2} = \delta_{i_1 i_2}\delta_{j_1 j_2}$, $\int \phi_i \psi_{jk} = 0$, where δ_{ij} denotes the Kronecker delta.

In addition to the standard properties of wavelets listed above, we suppose that both ϕ and ψ are bounded and compactly supported. For a detailed discussion of wavelets with compact support, see Daubechies (1992).

2.2. Wavelet-based estimators

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\{\mathcal{F}_t, t \in [0, 1]\}$ be an increasing, right-continuous family of sub-sigma-fields of $\mathcal{F}$. We take $\mathcal{F}_t$ to represent the information collected during the period $[0, t]$. In this setting, a counting process N is a stochastic process on [0, 1], adapted to $\{\mathcal{F}_t\}$, where each sample path is a right-continuous step function with $N(0) = 0$ and a finite number of jumps, each of size $+1$. Suppose that $\mathrm{E}N(1) < \infty$. Since N is increasing and hence a submartingale, it follows from the Doob-Meyer decomposition that $N = A + M$, where A is a predictable increasing process and M is a martingale. We also assume that there exists a non-negative left-continuous process λ, adapted to $\{\mathcal{F}_t\}$, with right-hand limits such that $A(t) = \int_0^t \lambda(s)\mathrm{d}s$. Then, by Aalen (1978), $M(t) = N(t) - \int_0^t \lambda(s)\mathrm{d}s$ is a square-integrable martingale with variance process $\langle M\rangle(t) = \int_0^t \lambda(s)\mathrm{d}s$. The process λ is called the intensity process of N, and $\mathrm{E}[\mathrm{d}N(t)|\mathcal{F}_t] = \lambda(t+)$ and $\mathrm{var}(\mathrm{d}N(t)|\mathcal{F}_t) = \lambda(t+)$. We assume that $\lambda(t)$ can be written in the form

$$\lambda(t) = \alpha(t)Y(t), \qquad t \in [0, 1],$$

where α is an unknown non-stochastic function, the underlying intensity function, while Y is an observable stochastic process. The function α is assumed to be left-continuous with right-hand limits, and Y is assumed to be predictable with respect to $\{\mathcal{F}_t\}$. The intensity α is interpreted as the transition intensity on the individual level, and in most applications $Y(t)$ measures the size of the risk population just before time t.

One can describe the above by writing the model

$$\mathrm{d}N_n(t) = \alpha(t)Y_n(t)\mathrm{d}t + \mathrm{d}M_n(t),$$

where $\mathrm{d}M_n(t)$ is noise. An index n is used to indicate the n-dependence of the counting process and is introduced to provide a suitable stochastic framework; see Andersen *et al.* (1993) for examples.

If $\alpha(t)$ is a square integrable then the wavelet transform of $\alpha(t)$ is given by

$$\alpha(t) = \sum_j b_j \phi_j(t) + \sum_{i=0}^{\infty} \sum_j b_{ij} \psi_{ij}(t), \tag{2.1}$$

where $b_j = \int \alpha \phi_j$ and $b_{ij} = \int \alpha \psi_{ij}$ are wavelet coefficients. Then the nonlinear estimator of $\alpha(t)$, obtained by thresholding empirical wavelet coefficients, has the form

$$\hat{\alpha}_n(t) = \sum_j \hat{b}_j \phi_j(t) + \sum_{i=0}^{q-1} \sum_j \hat{b}_{ij} I(|\hat{b}_{ij}| > \delta)\, \psi_{ij}(t), \tag{2.2}$$

where

$$\hat{b}_j = \int_0^1 \phi_j(s) \frac{J_n(s)}{Y_n(s)} \mathrm{d}N_n(s) = \sum_{k=1}^{N(1)} \frac{\phi_j(T_k)}{Y_n(T_k)},$$

$$\hat{b}_{ij} = \int_0^1 \psi_{ij}(s) \frac{J_n(s)}{Y_n(s)} \mathrm{d}N_n(s) = \sum_{k=1}^{N(1)} \frac{\psi_{ij}(T_k)}{Y_n(T_k)},$$

the T_k are the jump times of the counting process $N_n(t)$, and $J_n(s) = I\{Y_n(s) > 0\}$. Also we define $J_n(s)/Y_n(s) = 0$ when $Y_n(s) = 0$. Finally, set

$$\bar{b}_j = \int_0^1 \phi_j(t) J_n(t) \alpha(t) \mathrm{d}t, \qquad \bar{b}_{ij} = \int_0^1 \psi_{ij}(t) J_n(t) \alpha(t) \mathrm{d}t.$$

For notational convenience, the dependence of $\hat{b}_j$, $\hat{b}_{ij}$, $\bar{b}_j$ and $\bar{b}_{ij}$ on n is suppressed.

2.3. The *Y* Process

The following assumptions are made concerning the process $Y_n(t)$:

(A1) For each $n \geqslant 1$, $\{Y_n(t) : t \in [0, 1]\}$ is predictable with respect to the filtration $\{\mathcal{F}_t^{(n)} : t \in [0, 1]\}$.

(A2) There exists an $\epsilon_0 > 0$, independent of n and t, such that $Y_n(t) < \epsilon_0$ implies that $Y_n(t) = 0$.

(A3) The function $u(t) = \mathrm{E}(Y_n(t)/n)$ is continuous and satisfies

$$\inf_{t \in [0,1]} u(t) \geqslant \delta_0 \tag{2.3}$$

for some constant $\delta_0 > 0$.

(A4) For each $\gamma > 0$,

$$\sup_{t \in [0,1]} \mathrm{E}[|n^{-1} Y_n(t) - u(t)|^{\gamma}] = O(n^{-\gamma/2}). \tag{2.4}$$

(A5) For some fixed $\lambda_0 > 0$,

$$P[n^{-1} Y_n(t) < \lambda_0 \text{ for some } t \in [0, 1]] = O(n^{-1}). \tag{2.5}$$

We briefly comment on the above assumptions. In the framework we are considering, (A1) is a natural assumption to make. In many applications of Aalen's model, $Y_n(t)$ is a non-negative integer-valued process, in which case (A2) is satisfied with $\epsilon_0 = 1$. In (A3), the continuity assumption can be weakened, but (2.3) does seem to play an essential role in our proof of Theorem 2.1 below, as do (2.4) and (2.5). Apart from the very mild assumption of left continuity made in (A1), (A5) is the only explicit assumption we make about the sample path properties (as opposed to the moments) of the Y process. Note that, if (A3) is assumed, then (A5) is implied by the following: for some $\eta > 0$,

$$\mathrm{E}\left[\sup_{t\in[0,1]} |n^{-1}Y_n(t) - u(t)|^{\eta}\right] = O(n^{-1}).$$

This assertion follows easily from an application of Chebyshev's inequality.

2.4. Main results

We now introduce some further definitions before stating our main results, Theorems 2.1 and 2.2. Recall the assumption in Section 2.1 that $\int y^h\psi(y) = 0$ for each integer $0 \leqslant h \leqslant r-1$, and

$$\kappa = (r!)^{-1}\int y^r\,\psi(y)\,\mathrm{d}y \neq 0.$$

Suppose that the support of ψ and support of ϕ are both contained in the interval $[-\nu_1, \nu_2]$. Define

$$C_0 = 2\sqrt{\frac{r}{2r+1}\frac{1}{\lambda_0}\sup_{t\in[0,1]}\alpha(t)}, \tag{2.6}$$

where λ_0 is the quantity referred to in assumption (A5).

Theorem 2.1. *Assume the following: that ϕ and ψ satisfy the conditions stated in Section 2.1; that $Y_n(t)$ satisfies conditions* (A1)–(A5) *stated in Section 2.3; and that the intensity $\alpha(t)$ has rth derivative $\alpha^{(r)}$ which is bounded and piecewise continuous and has left and right limits everywhere on* [0, 1]. *Suppose also that*

$$p \to \infty, \qquad q \to \infty, \qquad p_q\delta^2 \to 0, \qquad p^{2r+1}\delta^2 \to \infty, \tag{2.7}$$

where

$$\delta \geqslant C(n^{-1}\log n)^{1/2}, \qquad C > C_0, \tag{2.8}$$

and $C_0 > 0$ is defined in (2.6). *Then*

$$E\left|\int(\hat{\alpha}_n - \alpha)^2 - \left\{n^{-1}p\int\frac{\alpha}{u} + p^{-2r}\kappa^2(1-2^{-2r})^{-1}\int\alpha^{(r)^2}\right\}\right| = o(n^{-1}p + p^{-2r}) \tag{2.9}$$

as $n \to \infty$.

Remark 2.1. Theorem 2.1 may be viewed as a counting process analogue of Theorem 2.1(i) in Hall and Patil (1995), which was obtained for wavelet density estimators. Note the similarity of our C_0, given in (2.6), and $C_0 = 2\{r \sup f/(2r+1)\}^{1/2}$ given by Hall and Patil (1995). However, our C_0 also depends on a quantity, λ_0, which appears in assumption (A5) and has no analogue in Hall and Patil (1995).

Remark 2.2. To implement the estimator (2.2) we need to specify C_0, which depends on r, which is either known a priori or may be guessed, and on unknowns λ_0 and $\sup_{t\in[0,1]}\alpha(t)$. Note that the unknown λ_0 may be estimated by $\hat{\lambda}_0 = \inf_{t\in[0,1]} n^{-1} Y_n(t)$; and $\sup_{t\in[0,1]}\alpha(t)$ may be estimated by $\sup_{t\in[0,1]}\hat{\alpha}(y)$, where $\hat{\alpha}(t)$ is a pilot estimator of $\alpha(t)$. However, more investigations are needed into the practical implementation of such an approach.

Remark 2.3 *Comparison with traditional MISE formulae*. From result (2.9), by taking the expected value on the left-hand side inside the modulus signs, we obtain a wavelet version of the traditional MISE formula:

$$\int \mathrm{E}(\hat{\alpha}_n - \alpha)^2 \sim n^{-1} p \int \frac{\alpha}{u} + p^{-2r}\kappa^2(1 - 2^{-2r})^{-1} \int \alpha^{(r)^2}, \tag{2.10}$$

where '$\sim$' means that the ratio of the left- and right-hand sides converges to 1 as $n \to \infty$. Here, the $n^{-1}p$ term derives from variance, and the p^{-2r} term from squared bias, exactly as in the case of classical formulae for kernel function estimators. To obtain the wavelet formula from its counterpart for kernel methods, albeit with different constants multiplying the bias contribution, simply replace bandwidth b_n by p^{-1}. See, for example, Andersen *et al.* (1993, Chapter IV) for a detailed account of the kernel case.

Of course, the right-hand side of (2.10) is asymptotically minimized by taking $p \sim a n^{1/(2r+1)}$, where $a = \{2r\kappa^2(1 - 2^{-2r})^{-1} \int \alpha^{(r)^2} (\int \alpha/u)^{-1}\}^{1/(2r+1)}$; and the minimum size of (2.10) is const.$n^{-2r/(2r+1)}$.

Theorem 2.2. *Assume all the conditions of Theorem* 2.1, *add the assumption that* $p_q^{2r+1} n^{-2r} \to \infty$, *and impose the condition of r-times differentiability of* α *only in a piecewise sense. That is, we ask that there exist points* $x_0 = 0 < x_1 < \ldots < x_N < 1 = x_{N+1}$ *such that the first r derivatives of* α *exist and are bounded and continuous on* (x_i, x_{i+1}) *for* $0 \leqslant i \leqslant N$, *with left- and right-hand limits. In particular,* α *itself may be only piecewise continuous. Then the result of Theorem* 2.1 *holds.*

Remark 2.4 *Comparison with kernel estimators.* This result is quite different from its analogue for a kernel-based estimator, where the presence of discontinuities can dramatically increase the order of magnitude of MISE. For detailed discussion of this point in the comparable situation of density estimation, see Hall and Patil (1995).

2.5. An example

Consider n similar machines operating independently. For $i = 1, \ldots, n$, define $I_i(t)$ as follows: $I_i(t) = 1$ if machine i is operable and busy at time t; $I_i(t) = 0$ if it is operable and idle at time t; and $I_i(t) = -1$ if it is under repair at time t. Note that, by assumption, the $I_i(t)$ are i.i.d. processes. The problem is to estimate the breakdown rate $\alpha(t)$ of a typical busy machine. Here we define $Y_n(t) = \#\{i : I_i(t-) = 1\}$. Then the intensity of breakdowns at time t is given by $Y_n(t)\alpha(t)$. Assuming that each I_i process is observed, we can formulate the model so that assumption (A1) is satisfied; and, as defined, $Y_n(t)$ is integer-valued and so satisfies (A2). If we make the very mild assumptions that

$$p_0 = P[I_i(t) = 1 \text{ for all } t \in [0, 1]] \in (0, 1)$$

and

$$p_0' = P[I_i(t) \neq 1 \text{ for any } t \in [0, 1]] \in (0, 1),$$

where $p_0 + p_0' < 1$ then, using the fact that $Y_n(t)$ is a binomial random variable with parameters n and $p \in [p_0, 1 - p_0']$ for each $t \in [0, 1]$, it is straightforward to show that (A3)–(A5) are also satisfied, so that Theorems 2.1 and 2.2 apply.

3. Auxiliary results

We now establish some lemmas which are required in the proofs of Theorems 2.1 and 2.2. In proving these lemmas we make use of two results which we present now for convenience. Observe that the notation used in Section 3.1 is sometimes different from that used in the rest of the paper.

3.1. Two inequalities

The first result is a point process analogue of Rosenthal's inequality; see Wood (1999, 2001) for further details. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ denote a filtered probability space satisfying the 'usual conditions'; see, for example, Brémaud (1981) or Rogers and Williams (1987). Suppose that N_t is a counting process (i.e. a right-continuous, non-negative, increasing integer-valued process with jumps of size $+1$) which is adapted to $(\mathcal{F}_t)_{t\geq 0}$. It is assumed that N_t has a predictable intensity λ_t, and that g_t is another predictable process. Assume that:

(a) $\int_0^t \lambda_s \mathrm{d}s < \infty$ almost surely;
(b) $\mathrm{E}[\int_0^t |g_s|\lambda_s \mathrm{d}s] < \infty$.

Then, using Brémaud (1981), (a) implies that $M_t = N_t - \int_0^t \lambda_s \mathrm{d}s$ is a local martingale which has bounded variation on $[0, t]$ with probability one, so that $S_t = \int_0^t g_s \mathrm{d}M_s$ is well defined as a Stieltjes integral. Moreover, (b) implies that S_t is a martingale. The predictable quadratic variation of S_t is given by $\langle S\rangle_t = \int_0^t g_s^2 \lambda_s \mathrm{d}s$. Choose $t > 0$. If (a) and (b) hold, and $\mathrm{E}\langle S\rangle_t < \infty$, then, for any $p \geq 2$,

$$c_p A_{p,t} \leq \mathrm{E}|S_t|^p \leq C_p A_{p,t}, \tag{3.1}$$

where

$$A_{p,t} = \mathrm{E}\left[\langle S\rangle_t^{p/2} + \int_0^t |g_s|^p \lambda_s \mathrm{d}s\right]$$

and c_p and C_p are finite positive quantities which depend only on p.

The second result we shall need is an exponential inequality for local martingales due to Courbot (1996). Let $\mathcal{M}_0^{2,\mathrm{loc}}$ denote the class of locally square-integrable local martingales which are null at zero. This class is defined as follows: $\{X_t : t \geqslant 0\} \in \mathcal{M}_0^{2,\mathrm{loc}}$ if $X_0 = 0$ and there exists a non-decreasing sequence of stopping times $0 = T_0 \leqslant T_1 \leqslant T_2 \ldots,$ with $\lim_{m\to\infty} T_m = \infty$ almost surely, such that, for each $m \geqslant 0$, $\{X_{\min(t,T_m)} : t \geqslant 0\}$ is an L^2 martingale, i.e. a martingale satisfying $\sup_{t\geqslant 0} \mathrm{E}X^2_{\min(t,T_m)} < \infty$. Let $(\Delta X)_t = X_t - X_{t-}$ be the pure jump process associated with X_t, and define $|X|_t^* = \sup_{s\in[0,t]}|X_s|$, $|\Delta X|_t^* = \sup_{s\in[0,t]}|(\Delta X)_s|$. As usual, $\langle X\rangle_t$ denotes the predictable variance process associated with X_t. Also, for $x, y > 0$, define the function $\Psi(x, y)$ by

$$\Psi(x, y) = (x + y)\log\{(y/x) + 1\} - y.$$

The following inequality is due to Courbot (1996, Inequality 3.2(ii)); see also Shorak and Wellner (1986, p. 897) for a closely related result. For x, β, a and t all positive.

$$P[|X|_t^* \geqslant x] \leqslant 2\exp\left\{-\Psi\left(\frac{\beta^2}{a^2}, \frac{x}{a}\right)\right\} + 2P[\langle X\rangle_t > \beta^2] + P[|\Delta X|_t^* > a]. \tag{3.2}$$

3.2. Some lemmas

We now present five lemmas which provide the basis for the proof of Theorem 2.1 and Theorem 2.2.

Lemma 3.1. *Under assumptions* (A2)–(A4)*, the following statements hold.*

(i) *For each* $x > 0$,

$$\sup_{t\in[0,1]} P[|n^{-1}Y_n(t) - u(t)| > x] = O(n^{-\gamma}),$$

for for any positive γ.

(ii) *For each* $\gamma > 0$,

$$\sup_{t\in[0,1]} P[Y_n(t) = 0] = O(n^{-\gamma}).$$

(iii) *For each* $\gamma > 0$,

$$\sup_{t\in[0,1]} \mathrm{E}[\{nJ_n(t)/Y_n(t)\}^\gamma] = O(1).$$

(iv) *We have*

$$\sup_{t\in[0,1]} \mathrm{E}\left|\frac{nJ_n(t)}{Y_n(t)} - \frac{1}{u(t)}\right| = O(n^{-1/2}).$$

Proof. (i) By Chebyshev's inequality,

$$P\big[|n^{-1}Y_n(t) - u(t)| > x\big] \leqslant x^{-2\gamma}\mathrm{E}\big[|n^{-1}Y_n(t) - u(t)|^{2\gamma}\big],$$

so the desired conclusion follows immediately from assumption (A4).

(ii) Since

$$\{\omega : Y_n(t) = 0\} \subseteq \{\omega : |n^{-1}Y_n(t) - u(t)| > \delta_0/2\},$$

where $\delta_0 = \inf_{t\in[0,1]}u(t)$ is positive by assumption (A3), the result follows as a consequence of part (i).

(iii) Define

$$A_{n,t} = \left\{\omega : \frac{nJ_n(t)}{Y_n(t)} \geqslant \rho\delta_0^{-1}\right\},$$

where $\rho > 1$ is a constant and $\delta_0 > 0$ is the quantity in (A3). On $A_{n,t}$,

$$\begin{aligned} n^{-1}Y_n(t) - u(t) &\leqslant \rho^{-1}\delta_0 - u(t) \\ &\leqslant \rho^{-1}\delta_0 - \inf_{t\in[0,1]} u(t) \\ &\leqslant -(1-\rho^{-1})\delta_0. \end{aligned}$$

Thus

$$A_{n,t} \subseteq \{\omega : |n^{-1}Y_n(t) - u(t)| > (1-\rho^{-1})\delta_0\}$$

and, by part (i),

$$\sup_{t\in[0,1]} P[A_{n,t}] \leqslant \sup_{t\in[0,1]} P[|n^{-1}Y_n(t) - u(t)| > (1-\rho^{-1})\delta_0] = O(n^{-\xi}), \tag{3.3}$$

for any $\xi > 0$. Now

$$\left\{\frac{nJ_n(t)}{Y_n(t)}\right\}^{\gamma} \leqslant \begin{cases} (n/\epsilon_0)^{\gamma} & \text{on } A_{n,t} (\text{by (A2)}) \\ (\rho/\delta_0)^{\gamma} & \text{on } A_{n,t}^c. \end{cases}$$

Therefore, choosing ξ in (3.3) to satisfy $\xi > \gamma$, we obtain

$$\sup_{t\in[0,1]} E\left\{\frac{nJ_n(t)}{Y_n(t)}\right\}^{\gamma} \leqslant \left(\frac{n}{\epsilon_0}\right)^{\gamma} \sup_{t\in[0,1]} P[A_{n,t}] + \left(\frac{\rho}{\delta_0}\right)^{\gamma} = O(1).$$

(iv) It is sufficient to show that $\sup_{t\in[0,1]}\mathrm{E}|Z_n(t)| = O(1)$, where

$$Z_n(t) = n^{1/2}\left(\frac{nJ_n(t)}{Y_n(t)} - \frac{1}{u(t)}\right).$$

In view of the identity

$$Z_n(t) = n^{1/2}\{J_n(t) - 1\}/u(t) - \left(\frac{nJ_n(t)}{u(t)Y_n(t)}\right)n^{-1/2}\{Y_n(t) - nu(t)\},$$

the triangle inequality, the Cauchy–Schwarz inequality, and assumption (A3), the conclusion follows since

$$\sup_{t\in[0,1]} \mathrm{E}[n^{1/2}\{1 - J_n(t)\}/u(t)] \leqslant n^{1/2}\delta_0^{-1} \sup_{t\in[0,1]} P[Y_n(t) = 0] = o(1)$$

using part (ii);

$$\sup_{t\in[0,1]} \mathrm{E}\left[\left(\frac{nJ_n(t)}{Y_n(t)}\right)^2\right] = O(1)$$

by part (iii); and

$$\sup_{t\in[0,1]} \mathrm{E}[\{n^{-1/2}|Y_n(t) - nu(t)|\}^2] = O(1)$$

using assumption (A4). □

Lemma 3.2. *Suppose that the interval* $[-\nu_1, \nu_2]$ *contains the support of both* ϕ *and* ψ.

(i) *If* $j \notin [-\nu_2, p + \nu_1]$, *then* $\hat{b}_j$, $\overline{b}_j$ *and* b_j *are all zero.*

(ii) *If* $j \notin [-\nu_2, p_i + \nu_1]$, *then* $\hat{b}_{ij}$, $\overline{b}_{ij}$ *and* b_{ij} *are all zero.*

Proof. Follows immediately from the definitions of the bs. □

In part (iii) and (iv) of Lemma 3.3 below,

$$\langle \hat{b}_j - \overline{b}_j \rangle = \int_0^1 \phi_j(s)^2 \alpha(s) \frac{J_n(s)}{Y_n(s)} \mathrm{d}s$$

is the predictable quadratic variation of $\hat{b}_j - \overline{b}_j$, and a similar definition applies to $\langle \hat{b}_{ij} - \overline{b}_{ij} \rangle$.

Lemma 3.3. *Suppose that assumptions* (A1)–(A4) *are satisfied. Then the following results hold for* b_j, $\hat{b}_j$, $\overline{b}_j$, b_{ij}, $\hat{b}_{ij}$ *and* $\overline{b}_{ij}$ *defined in Section 2.2.*

(i) *For any* $\rho \geqslant 1$ *and* $\gamma > 0$,

$$\sup_j \mathrm{E}|\overline{b}_j - b_j|^\rho = O(n^{-\gamma}).$$

(ii) *For any* $\rho \geqslant 1$ *and* $\gamma > 0$,

$$\sup_{0\leqslant i\leqslant q-1} \sup_j \mathrm{E}|\overline{b}_{ij} - b_{ij}|^\rho = O(n^{-\gamma}).$$

In (iii)–(viii) *below,* $\beta \geqslant 2$ *is fixed but arbitrary.*

(iii) $\sup_j \mathrm{E}\langle \hat{b}_j - \overline{b}_j \rangle^{\beta/2} = O(n^{-\beta/2})$.

(iv) $\sup_{i,j} \mathrm{E}\langle \hat{b}_{ij} - \overline{b}_{ij} \rangle^{\beta/2} = O(n^{-\beta/2})$.

(v) $\sup_j \mathrm{E}|\hat{b}_j - \overline{b}_j|^\beta = O(n^{-\beta/2})$.

(vi) $\sup_{0\leqslant i\leqslant q-1} \sup_j \mathrm{E}|\hat{b}_{ij} - \overline{b}_{ij}|^\beta = O(n^{-\beta/2})$.

(vii) $\sup_j \mathrm{E}|\hat{b}_j - b_j|^\beta = O(n^{-\beta/2})$.

(viii) $\sup_{0\leqslant i\leqslant q-1} \sup_j \mathrm{E}|\hat{b}_{ij} - b_{ij}|^\beta = O(n^{-\beta/2})$.

Proof. The proofs of (i), (iii), (v) and (vii) are very similar to those of (ii), (iv), (vi) and (viii), respectively, so we only give the latter.

(ii) Using Hölder's inequality, Fubini's theorem and the definitions of $\bar{b}_{ij}$ and b_{ij}, we obtain

$$
\begin{aligned}
\mathrm{E}|\bar{b}_{ij} - b_{ij}|^{\rho} &= \mathrm{E}\left|\int_0^1 \{1 - J_n(s)\}\psi_{ij}(s)\alpha(s)\mathrm{d}s\right|^{\rho} \\
&\leqslant \mathrm{E}\int_0^1 \{1 - J_n(s)\}|\psi_{ij}(s)\alpha(s)|^{\rho}\mathrm{d}s \\
&= \int_0^1 P[Y_n(s) = 0]|\psi_{ij}(s)\alpha(s)|^{\rho}\mathrm{d}s \\
&\leqslant \sup_{t\in[0,1]} P[Y_n(t) = 0]\int_0^1 |\psi_{ij}(s)\alpha(s)|^{\rho}\mathrm{d}s \\
&= O(n^{-\gamma}),
\end{aligned}
$$

for any $\gamma > 0$ by Lemma 3.1(ii).

(iv) Using Hölder's inequality (with $p = \beta/(\beta - 2)$, $q = \beta/2$ and $p^{-1} + q^{-1} = 1$), Fubini's theorem and Lemma 3.1(iii), and writing $s_1 = (s + j)/p_i$, we obtain

$$
\begin{aligned}
\sup_{i,j} \mathrm{E}\langle \hat{b}_{ij} - \bar{b}_{ij}\rangle^{\beta/2} &= \sup_{i,j} \mathrm{E}\left(\int_0^1 \psi_{ij}(s)^2\alpha(s)\frac{J_n(s)}{Y_n(s)}\mathrm{d}s\right)^{\beta/2} \\
&\leqslant \sup_{i,j} \mathrm{E}\left(\int_{-\nu_1}^{\nu_2} \psi(s)^2\alpha(s_1)\frac{J_n(s_1)}{Y_n(s_1)}\mathrm{d}s\right)^{\beta/2} \\
&\leqslant \{2(\nu_1 + \nu_2)\}^{(\beta-2)/2} n^{-\beta/2} \\
&\quad \times \int_{-\nu_1}^{\nu_2} |\psi(s)|^{\beta}|\sup_{i,j}\left[\alpha(s_1)|^{\beta/2}\mathrm{E}\left\{\frac{nJ_n(s_1)}{Y_n(s_1)}\right\}^{\beta/2}\right]\mathrm{d}s \\
&= O(n^{-\beta/2}).
\end{aligned}
$$

(vi) We apply Rosenthal's inequality (see (3.1)) with $g_s = \psi_{ij}(s)J_n(s)/Y_n(s)$ and $\lambda_s = \alpha(s)Y_n(s)$. Again writing $s_1 = (s + j)/p_i$, the second term on the right-hand side of (3.1) is given by

$$
\mathrm{E}\int_0^1 |\psi_{ij}(s)|^{\beta}\alpha(s)\{J_n(s)/Y_n(s)\}^{\beta-1}\mathrm{d}s
$$

$$\leqslant n^{-(\beta-1)} p_i^{(\beta-2)/2} \int_{-\nu_1}^{-\nu_2} |\psi(s)|^\beta \alpha(s_1) E\{nJ_n(s_1)/Y_n(s_1)\}^{\beta-1} \mathrm{d}s$$

$$= O(n^{-\beta/2})$$

by Lemma 3.1(iii), for $\beta \geqslant 2$ and $0 \leqslant i \leqslant q-1$. Therefore, using (3.1) and part (iv) of the present lemma,

$$\mathrm{E}|\hat{b}_{ij} - \bar{b}_{ij}|^\beta = O(n^{-\beta/2}).$$

(viii) This follows from parts (ii) and (vi) since, for $\beta \geqslant 0$,

$$\mathrm{E}|\hat{b}_{ij} - b_{ij}|^\beta \leqslant 2^\beta \{\mathrm{E}|\hat{b}_{ij} - \bar{b}_{ij}|^\beta + \mathrm{E}|\bar{b}_{ij} - b_{ij}|^\beta\}. \qquad \square$$

We now present another result which is required in the proof of Theorem 2.1. Let $N_n^{(h)}(t)$ be a counting process with intensity $\lambda_n^{(h)}(t)$, ($h = 0, 1, 2, 3$), defined on the same probability space, where $\lambda_n^{(0)}(t) = \alpha(t)Y_n(t)$, $\lambda_n^{(1)}(t) = n\alpha(t)u(t)$, $\lambda_n^{(2)}(t) = \alpha(t)(Y_n(t) - nu(t))^+$ and $\lambda_n^{(3)}(t) = \alpha(t)(nu(t) - Y_n(t))^+$, with $(x)^+ = \max(x, 0)$. As before, it is assumed that $Y_n(t)$ satisfies (A1). Define the corresponding counting process martingales by

$$M_n^{(h)}(t) = N_n^{(h)}(t) - \int_0^t \lambda_n^{(h)}(s)\mathrm{d}s, \qquad h = 0, 1, 2, 3.$$

It is assumed that the predictable quadratic covariation satisfies

$$\langle M_n^{(h)}, M_n^{(k)} \rangle = 0 \qquad \text{if } h \neq k. \tag{3.4}$$

Note that (3.4) implies that, with probability one, none of the jump times of $M_n^{(h)}(t)$ coincide with those of $M_n^{(k)}$ if $h \neq k$.

Consider the two pairs of superimposed process $M_n^{(0)}(t) + M_n^{(3)}(t)$ and $M_n^{(1)}(t) + M_n^{(2)}$. It is clear from the construction that

$$M_n^{(0)}(t) + M_n^{(3)}(t) = M_n^{(1)}(t) + M_n^{(2)}(t) \tag{3.5}$$

in distribution, the common intensity being

$$\lambda_n^{(0)}(t) + \lambda_n^{(3)}(t) = \lambda_n^{(1)}(t) + \lambda_n^{(2)}(t) = \max\{\alpha(t)Y_n(t), n\alpha(t)u(t)\}.$$

Define

$$\hat{b}_i - \bar{b}_i = \int_0^1 \phi_i(s) \frac{J_n(s)}{Y_n(s)} \mathrm{d}M_n^{(0)}(s)$$

and

$$\hat{d}_i - \bar{d}_i = \frac{1}{n} \int_0^1 \{\phi_i(s)/u(s)\} \mathrm{d}M_n^{(1)}(s).$$

Then we have the following result.

Lemma 3.4. *There exists a $C > 0$, independent of i, j and n, such that*

$$|\mathrm{cov}\{(\hat{b}_i - \bar{b}_i)^2, (\hat{b}_j - \bar{b}_j)^2\} - \mathrm{cov}\{(\hat{d}_i - \bar{d}_i)^2, (\hat{d}_j - \bar{d}_j)^2\}|$$

$$\leqslant \begin{cases} Cn^{-9/4} & \text{if } i, j \in (-\nu_2, p + \nu_1), \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The main idea in the proof is the construction leading to (3.5). The 'zero' case follows from Lemma 3.2(i), so we assume below that $i, j \in [-\nu_2, p + \nu_1]$.

Define

$$X_{hi} = \int_0^1 \phi_i(s) \frac{J_n(s)}{Y_n(s)} \mathrm{d}M_n^{(h)}(s), \qquad h = 0, 1, 2, 3,$$

and

$$Z_i = \int_0^1 \{\phi_i(s)/u(s)\} \mathrm{d}M_n^{(1)}(s).$$

Note that

$$X_{0i} = \hat{b}_i - \bar{b}_i \quad \text{and} \quad Z_i = \hat{d}_i - \bar{d}_i. \tag{3.6}$$

By the construction,

$$(X_{0i} + X_{3i}, X_{0j} + X_{3j}) = (X_{1i} + X_{2i}, X_{1j} + X_{2j}) \qquad \text{in distribution}$$

for each i and j. So, in particular,

$$\mathrm{cov}\{(X_{0i} + X_{3i})^2, (X_{0j} + X_{3j})^2\} = \mathrm{cov}\{(X_{1i} + X_{2i})^2, (X_{1j} + X_{2j})^2\}. \tag{3.7}$$

Bearing (3.6) and (3.7) in mind, the conclusion of the lemma will follow from an application of the triangle inequality if we can establish the inequalities

$$|\mathrm{cov}\{(X_{0i} + X_{3i})^2, (X_{0j} + X_{3j})^2\} - \mathrm{cov}(X_{0i}^2, X_{0j}^2)| \leqslant Cn^{-9/4}, \tag{3.8}$$

$$|\mathrm{cov}\{(X_{1i} + X_{2i})^2, (X_{1j} + X_{2j})^2\} - \mathrm{cov}(X_{1i}^2, X_{1j}^2)| \leqslant Cn^{-9/4}, \tag{3.9}$$

$$|\mathrm{cov}(X_{1i}^2, X_{1j}^2) - \mathrm{cov}(Z_i^2, Z_j^2)| \leqslant Cn^{-5/2}, \tag{3.10}$$

where C does not depend on i, j or n. To establish (3.8)–(3.10), we use the following elementary result: if, for some $K < \infty$, $\{U_{h,n} : h = 1, 2, 3, 4;\ n \geqslant 1\}$ is a family of random variables satisfying $\mathrm{E}U_{h,n}^4 \leqslant K$ for all h, n, then, for any $\epsilon > 0$,

$$|\mathrm{cov}\{(n^{-1/2}U_{1,n} + n^{-\epsilon-1/2}U_{2,n})^2, (n^{-1/2}U_{3,n} + n^{-\epsilon-1/2}U_{4,n})^2\}$$
$$- \mathrm{cov}\{(n^{-1/2}U_{1,n})^2, (n^{-1/2}U_{3,n})^2\}| \leqslant Cn^{-\epsilon-2}, \tag{3.11}$$

where $C = C(K)$ depends only on K and is finite if K is finite. The proof of (3.11), which involves several applications of the Hölder and Lyapunov inequalities, is straightforward and is omitted.

We may use Rosenthal's inequality as in the proof of Lemma 3.3(vi) to show that, for some $K < \infty$ independent of i and n,

$$\mathrm{E}X_{hi}^4 \leqslant n^{-2}K \quad (h = 0, 1), \qquad \mathrm{E}X_{hi}^4 \leqslant n^{-3}K \quad (h = 2, 3)$$

and

$$\mathrm{E}(X_{1i} - Z_i)^4 \leqslant n^{-4}K.$$

To establish (3.8), we apply (3.11) with

$$U_{1,n} = n^{1/2}X_{0i}, \qquad U_{2,n} = n^{3/4}X_{3,i}, \qquad U_{3,n} = n^{1/2}X_{0j}, \qquad U_{4,n} = n^{3/4}X_{3j};$$

to establish (3.9), we apply (3.11) with

$$U_{1,n} = n^{1/2}X_{1i}, \qquad U_{2,n} = n^{3/4}X_{2i}, \qquad U_{3,n} = n^{1/2}X_{1j}, \qquad U_{4,n} = n^{3/4}X_{2j};$$

and, finally, to establish (3.10), we apply (3.11) with

$$U_{1,n} = n^{1/2}X_{1i}, \qquad U_{2,n} = n(X_{1i} - Z_i), \qquad U_{3,n} = n^{1/2}X_{1j}, \qquad U_{4,n} = n(X_{1j} - Z_j).$$

□

Lemma 3.5. *Suppose that the assumptions of Theorem 2.1 are satisfied. Then, for any* δ *satisfying*

$$\delta \geqslant C(n^{-1}\log n)^{1/2},$$

where $C > C_0$ *with* C_0 *defined in* (2.6), *we have*

$$\sup_{0\leqslant i\leqslant q-1} \sup_j P[|\hat{b}_{ij} - \bar{b}_{ij}| > \delta] = \begin{cases} o(n^{-1}p + p^{-2r}) & \text{if } j \in [-\nu_2,\ p_i + \nu_1], \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The 'zero' case follows immediately from Lemma 3.2. To establish the result in the other case, we may apply inequality (3.2). Consider the martingale

$$(\hat{b}_{ij} - \bar{b}_{ij})(t) = \int_0^t \psi_{ij}(s)\frac{J_n(s)}{Y_n(s)}\mathrm{d}M_n(s), \tag{3.12}$$

where $M_n(t) = N_n(t) - \int_0^t \alpha(s)Y_n(s)\mathrm{d}s$ and $N_n(t)$ is an integer-valued counting process; see Section 2.2. Thus $\hat{b}_{ij} - \bar{b}_{ij}$ in previous notation is equal to $(\hat{b}_{ij} - \bar{b}_{ij})(t)$ evaluated at $t = 1$. Let Δ_{ij} denote the largest jump, in absolute value, of the martingale $(\hat{b}_{ij} - \bar{b}_{ij})(t)$ for $t \in [0, 1]$. Then Δ_{ij} is bounded above by the supremum of the integrand in (3.12). Therefore

$$\Delta_{ij} \leqslant p_q^{1/2}\|\psi\|_\infty\left(\sup_{t\in[0,1]}\frac{J_n(t)}{Y_n(t)}\right),$$

where $\|\psi\|_\infty = \sup_{x\in[-\nu_1,\nu_2]}|\psi(x)|$. Now, for any $\epsilon > 0$, we have

$$\{\Delta_{ij} > \epsilon(n\log n)^{-1/2}\} \subseteq \left\{ p_q^{1/2}\|\psi\|_\infty \sup_{t\in[0,1]} \frac{J_n(t)}{Y_n(t)} > \epsilon(n\log n)^{-1/2} \right\}$$

$$\subseteq \{n^{-1}Y_n(t) < \epsilon^{-1}\|\psi\|_\infty (n^{-1}p_q \log n)^{1/2} \text{ for some } t \in [0, 1]\}$$

$$\subseteq \{n^{-1}Y_n(t) < \lambda_0 \text{ for some } t\}$$

when n is sufficiently large since, by hypothesis, $n^{-1}p_q \log n \to 0$. Therefore, using (A5),

$$P[\Delta_{ij} > \epsilon(n\log n)^{-1/2}] = O(n^{-1}) \tag{3.13}$$

for any fixed $\epsilon > 0$.

The predictable quadratic variation of $(\hat{b}_{ij} - \bar{b}_{ij})(t)$ over $t \in [0, 1]$ is given by

$$\langle \hat{b}_{ij} - \bar{b}_{ij}\rangle = \int_0^1 \psi_{ij}(s)^2 \alpha(s) \frac{J_n(s)}{Y_n(s)} \mathrm{d}s \leqslant \|\alpha\|_\infty \sup_{t\in[0,1]} \frac{J_n(t)}{Y_n(t)},$$

where $\|\alpha\|_\infty = \sup_{t\in[0,1]}\alpha(t)$. Therefore,

$$P[\langle \hat{b}_{ij} - \bar{b}_{ij}\rangle > n^{-1}\lambda_0^{-1}\|\alpha\|_\infty] \leqslant P\left[\|\alpha\|_\infty \sup_{t\in[0,1]} \frac{J_n(t)}{Y_n(t)} > n^{-1}\lambda_0^{-1}\|\alpha\|_\infty\right]$$

$$\leqslant P[n^{-1}Y_n(t) < \lambda_0 \text{ for some } t \in [0, 1]]$$

$$= O(n^{-1}) \tag{3.14}$$

using (A5) again.

Put $a = \epsilon(n\log n)^{-1/2}$, $\beta^2 = \|\alpha\|_\infty \lambda_0^{-1} n^{-1}$ and $x = C(n^{-1}\log n)^{1/2}$, where $C > C_0$ with C_0 defined in (2.6). We now apply inequality (3.2). For small $\epsilon > 0$, we have

$$\Psi\left(\frac{\beta^2}{a^2}, \frac{x}{a}\right) = \left[\left(\frac{\|\alpha\|_\infty}{\lambda_0\epsilon^2} + \frac{C}{\epsilon}\right)\log\left(1 + \frac{\lambda_0\epsilon C}{\|\alpha\|_\infty}\right) - \frac{C}{\epsilon}\right]\log n$$

$$= \left[\frac{C^2\lambda_0}{2\|\alpha\|_\infty} + R(\epsilon)\right]\log n,$$

where $R(\epsilon) = O(\epsilon)$. Now we choose ϵ so small that

$$\frac{C^2\lambda_0}{2\|\alpha\|_\infty} + R(\epsilon) > \frac{C_0^2\lambda_0}{2\|\alpha\|_\infty},$$

in which case

$$\exp\left[-\Psi\left(\frac{\beta^2}{a^2}, \frac{x}{a}\right)\right] = o\{n^{-2r/(2r+1)}\} = o(n^{-1}p + p^{-2r}), \tag{3.15}$$

so the desired conclusion follows from (3.2) and (3.13)–(3.15). □

4. Proofs of theorems

Proof of Theorem 2.1. Using the orthogonality properties of $\phi_i(t)$ and $\psi_{ij}(t)$ indicated in Section 2.1, we may write

$$\int(\hat{\alpha}_n - \alpha)^2 = \int\left\{\sum_j(\hat{b}_j - b_j)\phi_j(t)\right.$$

$$\left. + \sum_{i=0}^{q-1}\sum_j(\hat{b}_{ij}I(|\hat{b}_{ij}| > \delta) - b_{ij})\psi_{ij}(t) + \sum_{i=q}^{\infty}\sum_j b_{ij}\psi_{ij}(t)\right\}^2 \mathrm{d}t$$

$$= I + II + III + IV,$$

where

$$I = \sum_j(\hat{b}_j - b_j)^2, \qquad II = \sum_{i=0}^{q-1}\sum_j b_{ij}^2,$$

$$III = \sum_{i=0}^{q-1}\sum_j\{(\hat{b}_{ij} - b_{ij})^2 - b_{ij}^2\}I(|\hat{b}_{ij}| > \delta) \quad \text{and} \quad IV = \sum_{i=q}^{\infty}\sum_j b_{ij}^2.$$

The expectations of $I, \ldots, IV$ are approximated in steps 1 to 4 below.

Step 1. We show that

$$\mathrm{E}\left|\sum_j(\hat{b}_j - b_j)^2 - n^{-1}p\int_0^1 \frac{\alpha(s)}{u(s)}\mathrm{d}s\right| = o(n^{-1}p). \tag{4.1}$$

Using Lemmas 3.2 and 3.3, it will be seen that

$$\mathrm{E}\sum_j(\bar{b}_j - b_j)^2 = o(n^{-1}p)$$

and

$$\left|\mathrm{E}\sum_j(\hat{b}_j - \bar{b}_j)(\bar{b}_j - b_j)\right| \leqslant \sum_j\{\mathrm{E}(\hat{b}_j - \bar{b}_j)^2\}^{1/2}\{\mathrm{E}(\bar{b}_j - b_j)^2\}^{1/2} = o(n^{-1}p).$$

Consequently, (4.1) will be established if we can show that

$$\left|\mathrm{E}\sum_i(\hat{b}_j - \bar{b}_j)^2 - n^{-1}p\int_0^1 \frac{\alpha(s)}{u(s)}\mathrm{d}s\right| = o(n^{-1}p) \tag{4.2}$$

and

$$\mathrm{var}\left\{\sum_j(\hat{b}_j - \bar{b}_j)^2\right\} = o(n^{-2}p^2). \tag{4.3}$$

Now, using Lemma 3.1(iv),

$$\mathrm{E}(\hat{b}_j - \bar{b}_j)^2 = \mathrm{E}\langle \hat{b}_j - \bar{b}_j \rangle = \int_0^1 \phi_i^2(s)\alpha(s)\mathrm{E}\left\{\frac{J_n(s)}{Y_n(s)}\right\}\mathrm{d}s$$

$$= n^{-1}\int_{-j}^{p-j} \phi^2(s)\alpha\{(s+j)/p\}u\{(s+j)/p\}^{-1}\mathrm{d}s + o(n^{-1}), \tag{4.4}$$

where $u(t)$ is given in (A3) and the remainder term is of the stated order uniformly in j. Since the support of ϕ is contained in $[-\nu_1, \nu_2]$, we have the following:

(a) the limits of integration $(-j, p-j)$ may be replaced by $(-\infty, \infty)$ when $j \in (\nu_1, p-\nu_2)$;
(b) the number of integers j which lie in $[-\nu_2, \nu_1] \cup [p-\nu_2, p+\nu_1]$ is bounded above by $2(\nu_1+\nu_2)$; and
(c) when $j \notin [-\nu_2, p+\nu_1]$, $\mathrm{E}[\hat{b}_j - \bar{b}_j]^2 = 0$ by Lemma 3.2(i).

Define $A_1 = \{j \in \mathbb{Z} : j \in (\nu_1, p-\nu_2)\}$; $A_3 = \{j \in \mathbb{Z} : j \in [-\nu_2, \nu_1] \cup [p-\nu_2, p+\nu_1]\}$; and $A_2 = \{j \in \mathbb{Z} : j \notin [-\nu_2, p+\nu_1]\}$. Then, using (a)–(c) and (4.4), we obtain

$$\sum_{j\in\mathbb{Z}} \mathrm{E}[\hat{b}_j - \bar{b}_j]^2 = \left(\sum_{j\in A_1} + \sum_{j\in A_2} + \sum_{j\in A_3}\right)\mathrm{E}[\hat{b}_j - \bar{b}_j]^2$$

$$= \left(\sum_{j\in A_1} \mathrm{E}[\hat{b}_j - \bar{b}_j]^2\right) + O(n^{-1}) + 0$$

$$= \left(n^{-1}p\int_{-\infty}^{\infty}\phi^2(s)p^{-1}\left[\sum_{j\in A_1}\frac{\alpha\{(s+j)/p\}}{u\{(s+j)/p\}}\right]\mathrm{d}s\right) + o(n^{-1}p)$$

$$= \left(n^{-1}p\int_0^1 \frac{\alpha(s)}{u(s)}\mathrm{d}s\right) + o(n^{-1}p).$$

The final step uses the fact that

$$\left|\int_0^1 \frac{\alpha(s)}{u(s)}\mathrm{d}s - p^{-1}\sum_{j\in A_1}\frac{\alpha\{(s+j)/p\}}{u\{(s+j)/p\}}\right| = o(1),$$

which is itself a consequence of the continuity of u and α. Thus (4.2) has been proved.

To prove (4.3) we use Lemma 3.4, arguing as follows:

$$\mathrm{var}\left\{\sum_i(\hat{b}_i-\bar{b}_i)^2\right\}=\sum_i\sum_j \mathrm{cov}\{(\hat{b}_i-\bar{b}_i)^2,\,(\hat{b}_j-\bar{b}_j)^2\}$$

$$=\sum_{a,b=1}^{3}\left(\sum_{i\in A_a}\sum_{j\in A_b}\mathrm{cov}\{(\hat{d}_i-\bar{d}_i)^2,\,(\hat{d}_j-\bar{d}_j)^2\}\right)+O(n^{-9/4}p^2)+O(n^{-2}p)$$

$$=\left(\sum_{i\in A_1}\sum_{j\in A_1}\mathrm{cov}\{(\hat{d}_i-\bar{d}_i)^2,\,(\hat{d}_j-\bar{d}_j)^2\}\right)+o(n^{-2}p^2). \tag{4.5}$$

The moment generating function of $(\hat{d}_i-\bar{d}_i,\,\hat{d}_j-\bar{d}_j)$ is given by

$$M_{ij}(\theta_1,\,\theta_2)=\mathrm{E}\exp\left[n^{1/2}\int_0^1\frac{\theta_1\phi_i(s)+\theta_2\phi_j(s)}{u(s)}\mathrm{d}M_n^{(1)}(s)\right]$$

$$=\exp\left[n\int_0^1\frac{\alpha(s)}{u(s)}\Gamma(\frac{\theta_1\phi_i(s)+\theta_2\phi_j(s)}{nu(s)})\mathrm{d}s\right],$$

where $\Gamma(x)=\exp(x)-1-x$. Using the moment generating function, it can be shown by direct calculation that

$$\sup_{i,j}\left|\mathrm{cov}\{(\hat{d}_i-\bar{d}_i)^2,\,(\hat{d}_j-\bar{d}_j)^2\}-2\{\mathrm{cov}(\hat{d}_i-\bar{d}_i,\,\hat{d}_j-\bar{d}_j)\}^2\right|\leqslant Cn^{-3}, \tag{4.6}$$

where C is a constant. Also, for $i,\,j\in A_1$,

$$\mathrm{cov}(\hat{d}_i-\bar{d}_i,\,\hat{d}_j-\bar{d}_j)=n^{-1}\int_0^1\phi_i(s)\phi_j(s)\frac{\alpha(s)}{\alpha(s)}\mathrm{d}s$$

$$=n^{-1}\int_{-i}^{p-i}\phi(s)\phi(s+i-j)\frac{\alpha\{(s+i)/p\}}{u\{(s+i)/p\}}\mathrm{d}s$$

$$=o(n^{-1})$$

uniformly in i, j and n, due to the orthogonality properties of ϕ. Therefore

$$\sum_{i\in A_1}\sum_{j\in A_1}\{\mathrm{cov}(\hat{d}_i-\bar{d}_i,\,\hat{d}_j-\bar{d}_j)\}^2=o(n^{-2}p^2). \tag{4.7}$$

Putting (4.5), (4.6) and (4.7) together, we obtain the desired conclusion.

Step 2. We show that

$$\mathrm{E}\left|\sum_{i=0}^{q-1}\sum_j b_{ij}^2-p^{-2r}\kappa^2(1-2^{-2r})^{-1}\int_0^1\{\alpha^{(r)}(s)\}^2\mathrm{d}s\right|=o(p^{-2r}).$$

Note that no random terms appear above, so we may ignore the expectation operator. Using Taylor expansion of α and the fact that the first $r-1$ moments of ψ are zero, we obtain

$$\begin{aligned}
b_{ij} &= \int_0^1 \psi_{ij}(s)\alpha(s)\mathrm{d}s \\
&= p_i^{-1/2}\int_{-j}^{p_i-j} \psi(s)\alpha\{(s+j)/p_i\}\mathrm{d}s \\
&= p_i^{-1/2}\int_{-j}^{p_i-j} \psi(s)\{r\}^{-1}(s/p_i)^r\{\alpha^{(r)}(j/p_i)+o(1)\}\mathrm{d}s \\
&= \begin{cases} \kappa p_i^{-(r+1/2)}\{\alpha^{(r)}(j/p_i)+o(1)\} & \text{if } j\in(\nu_1,\ p_i-\nu_2), \\ O(p_i^{-(r+1/2)}) & \text{if } j\in[-\nu_2,\,\nu_1]\cup[p_i-\nu_2,\ p_i+\nu_1], \\ 0 & \text{otherwise,} \end{cases}
\end{aligned} \tag{4.8}$$

where the order terms are uniform in i and j. Consequently,

$$\begin{aligned}
\sum_{i=0}^{q-1}\sum_j b_{ij}^2 &= \kappa^2\sum_{i=0}^{q-1}\sum_j p_i^{-2r-1}\left[\{\alpha^{(r)}(j/p_i)\}^2+o(1)\right] \\
&= \kappa^2 p^{-2r}\sum_{i=0}^{q-1}2^{-2ir}\left[\int_0^1\{\alpha^{(r)}(s)\}^2\mathrm{d}s+o(1)\right] \\
&= \kappa^2 p^{-2r}(1-2^{-2r})^{-1}\int_0^1\{\alpha^{(r)}(s)\}^2\mathrm{d}s+o(p^{-2r}),
\end{aligned}$$

as required.

Step 3. We show that

$$\mathrm{E}\left|\sum_{i=0}^{q-1}\sum_j\left\{(\hat{b}_{ij}-b_{ij})^2-b_{ij}^2\right\}I(|\hat{b}_{ij}|>\delta)\right| = o(n^{-1}p+p^{-2r}). \tag{4.9}$$

It is sufficient to show that

$$\left(\sup_{i,j}P[|\hat{b}_{ij}|>\delta]\right)\sum_{i=0}^{q-1}\sum_j b_{ij}^2 = o(n^{-1}p+p^{-2r}) \tag{4.10}$$

and

$$\sum_{i=0}^{q-1}\sum_j \mathrm{E}\{(\hat{b}_{ij}-b_{ij})^2 I(|\hat{b}_{ij}|>\delta)\} = o(n^{-1}p+p^{-2r}). \tag{4.11}$$

Now for any positive $\beta_1, \beta_2, \beta_3$ satisfying $\sum\beta_k=1$,

$$I(|\hat{b}_{ij}|>\delta)\leqslant I(|\hat{b}_{ij}-\bar{b}_{ij}|>\beta_1\delta)+I(|\bar{b}_{ij}-b_{ij}|>\beta_2\delta)+I(|b_{ij}|>\beta_3\delta),$$

and therefore

$$\sup_{i,j} P[|\hat{b}_{ij}| > \delta] \leqslant \sup_{i,j} P[|\hat{b}_{ij} - \bar{b}_{ij}| > \beta_1\delta] + \sup_{i,j} P[|\bar{b}_{ij} - b_{ij}| > \beta_2\delta] + \sup_{i,j} I(|b_{ij}| > \beta_3\delta). \tag{4.12}$$

Chose β_2 and β_3 so small that β_1 is sufficiently close to 1 for the inequality $\beta_1 C > C_0$ to hold. Then by Lemma 3.5,

$$\sup_{i,j} P[|\hat{b}_{ij} - \bar{b}_{ij}| > \beta_1\delta] = o(n^{-1}p + p^{-2r}). \tag{4.13}$$

Also, by Lemma 3.3(ii) combined with Chebyshev's inequality,

$$\sup_{i,j} P[|\bar{b}_{ij} - b_{ij}| > \beta_2\delta] = o(n^{-1}p + p^{-2r}); \tag{4.14}$$

and, since $p^{2r+1}\delta^2 \to \infty$ and $|b_{ij}| \leqslant Bp_i^{-r-(1/2)}$ for some $B > 0$, it follows that

$$\sup_{i,j} I(|b_{ij}| > \beta_3\delta) = 0 \tag{4.15}$$

for n sufficiently large. Putting (4.12)–(4.15) together, it will be seen that

$$\sup_{i,j} P[|\hat{b}_{ij}| > \delta] = o(n^{-1}p + p^{-2r});$$

and, by step 2, $\sum_{i=0}^{q-1}\sum_j b_{ij}^2$ is bounded. So (4.10) has been proved.

To establish (4.11), we use Hölder's inequality to obtain

$$\begin{aligned}
&\sum_{i=0}^{q-1}\sum_j \mathrm{E}\{(\hat{b}_{ij} - b_{ij})^2 I(|\hat{b}_{ij}| > \delta)\} \\
&\quad \leqslant \sum_{i=0}^{q-1}\sum_j \Big\{\mathrm{E}|\hat{b}_{ij} - b_{ij}|^{2/\beta_1}\Big\}^{\beta_1} \Big\{P[|\hat{b}_{ij}| > \delta]\Big\}^{\beta_2} \\
&\quad \leqslant \left\{\sup_{i,j} P[|\hat{b}_{ij}| > \delta]\right\}^{\beta_2} \sum_{i=0}^{q-1}\sum_j \Big\{\mathrm{E}|\hat{b}_{ij} - b_{ij}|^{2/\beta_1}\Big\}^{\beta_1},
\end{aligned} \tag{4.16}$$

where β_1 and β_2 are positive numbers which sum to 1. By Lemma 3.5 and (4.12)–(4.15), we may choose β_2 so close to 1 that

$$\left\{\sup_{i,j} P[|\hat{b}_{ij}| > \delta]\right\}^{\beta_2} = o(n^{-1}p + p^{-2r}); \tag{4.17}$$

and by Lemma 3.3(viii) we have, for any $\beta_1 > 0$,

$$\sup_{i,j}\Big\{\mathrm{E}|\hat{b}_{ij} - b_{ij}|^{2/\beta_1}\Big\}^{\beta_1} = O(n^{-1}), \tag{4.18}$$

and therefore

$$\sum_{i=0}^{q-1}\sum_j \{\mathrm{E}|\hat{b}_{ij} - b_{ij}|^{2/\beta_1}\}^{\beta_1} = \sum_{i=0}^{q-1} O(n^{-1}p_i) = O(n^{-1}p_q) = o(1) \tag{4.19}$$

by (2.7) and (2.8). Putting (4.17) and (4.19) together, we obtain (4.11).

Step 4. We show that

$$\sum_{i=q}^{\infty}\sum_j b_{ij}^2 = o(p^{-2r}).$$

By (4.8), $|b_{ij}| \leqslant C p_i^{-(r+(1/2))}$ for some constant C independent of i, j and n. Consequently,

$$\begin{aligned}\sum_{i=q}^{\infty}\sum_j b_{ij}^2 &\leqslant C^2 \sum_{i=q}^{\infty}\sum_j p_i^{-(2r+1)} \\ &= O\left(\sum_{i=q}^{\infty} p_i^{-2r}\right) \\ &= O(p_q^{-2r}) \\ &= o(p^{-2r})\end{aligned}$$

since $p_q = p2^q$ and $q \to \infty$.

Putting steps 1–4 together, the proof of Theorem 2.1 is complete. □

Proof of Theorem 2.2. We only outline the proof of Theorem 2.2 since its derivation mostly follows from careful inspection and modification of the proof of Theorem 2.1.

Let $\mathcal{D}$ denote the finite set of set of points where $\alpha^{(m)}$ has a point of discontinuity for some $0 \leqslant m \leqslant r$. Since the support of ϕ is contained in $[-\nu_1, \nu_2]$, both b_j and $\hat{b}_j$ are constructed entirely from (Stieltjes) integrals on a finite number of intervals in which $\alpha^{(m)}$ is everywhere continuous unless

$$j \in \mathcal{K} = \{k : k \in (ps - \nu_1,\ ps + \nu_2) \text{ for some } s \in \mathcal{D}\}.$$

Also, since the support of ψ is contained in $[-\nu_1, \nu_2]$ both b_{ij} and $\hat{b}_{ij}$ are constructed from an interval as just described unless

$$j \in \mathcal{K}_i = \{k : k \in (p_i s - \nu_1,\ p_i s + \nu_2) \text{ for some } s \in \mathcal{D}\}.$$

Therefore,

$$\begin{aligned}\int(\hat{\alpha}_n - \alpha)^2 = {}& I(\mathcal{K}) + II(\mathcal{K}_i) + III(\mathcal{K}_i) + IV(\mathcal{K}_i) \\ &+ I(\tilde{\mathcal{K}}) + \mathrm{II}(\tilde{\mathcal{K}}_i) + \mathrm{III}(\tilde{\mathcal{K}}_i) + \mathrm{IV}(\tilde{\mathcal{K}}_i),\end{aligned}$$

where $\tilde{\mathcal{S}}$ denotes the complement of $\mathcal{S}$ in $\mathbb{Z}$ and

$$I(\mathcal{S}) = \sum_{j\in\mathcal{S}} (\hat{b}_j - b_j)^2, \qquad II(\mathcal{S}_i) = \sum_{i=0}^{q-1} \sum_{j\in\mathcal{S}_i} b_{ij}^2,$$

$$III(\mathcal{S}_i) = \sum_{i=0}^{q-1} \sum_{j\in\mathcal{S}_i} \Big\{ (\hat{b}_{ij} - b_{ij})^2 - b_{ij}^2 \Big\} I(|\hat{b}_{ij}| > \delta) \qquad \text{and} \qquad IV(\mathcal{S}_i) = \sum_{i=q}^{\infty} \sum_{j\in\mathcal{S}_i} b_{ij}^2.$$

Using the proof of Theorem 2.1 it is clear that $\mathrm{E}[I(\tilde{\mathcal{K}} + \mathrm{II}(\tilde{\mathcal{K}}_i) + \mathrm{III}(\tilde{\mathcal{K}}_i) + \mathrm{IV}(\tilde{\mathcal{K}}_i)]$ has the asymptotic properties claimed for $\int(\hat{\alpha}_n - \alpha)^2$ in Theorem 2.1. Since $\mathcal{K}$ can have no more than $(\nu_1 + \nu_2)(\#\mathcal{D})$ elements it obviously follows that $\mathrm{E}[I(\mathcal{K})] = o(n^{-1}p)$, and since $\mathcal{K}_i$ contains no more than $(\nu_1 + \nu_2)(\#\mathcal{D})$ elements it follows that $IV(\mathcal{K}) = o(n^{-2r/(2r+1)})$ as $p_q = o(n^{-2r/(2r+1)})$ and at discontinuity points b_{ij}^2 is of size p_i^{-1}.

To show that $\mathrm{E}[II(\mathcal{K}_i) + III(\mathcal{K}_i)]$ is also of smaller size, write

$$II(\mathcal{K}_i) + III(\mathcal{K}_i) = \sum_{i=0}^{q-1} \sum_{j\in\mathcal{K}_i} b_{ij}^2 I(|\hat{b}_{ij}| < \delta) + \sum_{i=0}^{q-1} \sum_{j\in\mathcal{K}_i} (\hat{b}_{ij} - b_{ij})^2 I(|\hat{b}_{ij}| > \delta). \tag{4.20}$$

From (4.16)–(4.18) and the fact that $q = O(\log n)$ it follows that the expected value of the second term in (4.20) is of size $o(n^{-1}p)$.

For the first term in (4.20) note that, since

$$I(|\hat{b}_{ij}| \leqslant \delta) \leqslant I\{|b_{ij}| \leqslant (1+\epsilon)\,\delta\} + I(|\hat{b}_{ij} - b_{ij}| > \epsilon\delta)$$

and

$$I(|b_{ij}| \leqslant (1-\epsilon)\,\delta) \leqslant I(|\hat{b}_{ij}| \leqslant \delta) + I(|\hat{b}_{ij} - b_{ij}| > \epsilon\delta),$$

then

$$s_1 - \Delta \leqslant \sum_{i=0}^{q-1} \sum_{j\in\mathcal{K}_i} b_{ij}^2 I(|\hat{b}_{ij}| < \delta) \leqslant s_2 + \Delta,$$

where

$$s_1 = \sum_{i=0}^{q-1} \sum_{j\in\mathcal{K}_i} b_{ij}^2\, I\{|b_{ij}| \leqslant (1+\epsilon)\,\delta\},$$

$$s_2 = \sum_{i=0}^{q-1} \sum_{j\in\mathcal{K}_i} b_{ij}^2\, I\{|b_{ij}| \leqslant (1-\epsilon)\,\delta\},$$

$$\Delta = \sum_{i=0}^{q-1} \sum_{j\in\mathcal{K}_i} b_{ij}^2\, I(|\hat{b}_{ij} - b_{ij}| > \epsilon\delta).$$

Now, again owing to the fact that $\mathcal{K}_i$ contains no more than $(\nu_1 + \nu_2)(\#\mathcal{D})$ elements, both s_1 and s_2 are equal to $o(n^{-2r/(2r+1)})$. Further, in view of finite number of points in $\mathcal{K}_i$, if follows from (4.13) and (4.14) that $\mathrm{E}[\Delta] = o(n^{-2r/(2r+1)})$. □

Acknowledgements

We are grateful to the two referees and Professor Peter Hall for helpful comments.

References

Aalen, O.O. (1978) Nonparametric inference for a family of counting processes. *Ann. Statist.*, **6**, 701–726.

Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993) *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Brémaud, P. (1981) *Point Processes and Queues: Martingale Dynamics.* New York: Springer-Verlag.

Courbot, B. (1996) Exponential inequalities: a semimartingale theory approach. Preprint. IRMAR, University of Rennes, I.

Daubechies, I. (1992) *Ten Lectures on Wavelets*. Philadelphia: SIAM.

Donoho, D. and Johnstone, I.M. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

Donoho, D. and Johnstone, I.M. (1998) Minimax estimation via wavelet shrinkage. *Ann. Statist.*, **26**, 879–921.

Donoho, D., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: asymptopia? (with discussion). *J. Roy. Statist. Soc. Ser. B*, **57**, 301–369.

Donoho, D., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996) Density estimation by wavelet thresholding. *Ann. Statist.*, **24**, 508–539.

Hall, P. and Patil, P.N. (1995) Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.*, **23**, 905–928.

Hall, P. and Patil, P.N. (1996) On the choice of smoothing parameter, threshold and truncation in nonparametric regression by nonlinear wavelet methods. *J. Roy. Statist. Soc. Ser. B*, **58**, 361–378.

Hall, P., McKay, I. and Turlach, B.A. (1996) Performance of wavelet methods for functions with many discontinuities. *Ann. Statist.*, **24**, 2462–2476.

Johnstone, I.M. and Silverman, B.W. (1997) Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B*, **59**, 319–351.

Kerkyacharian, G. and Picard, D. (1993) Introduction aux ondelettes et estimation de densité, 1: Introduction aux ondelettes et á l'analyse multiresolution. Lecture Notes, Univ. Paris XII.

Kolaczyk, E.D. (1999) Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. *Statist. Sinica*, **9**, 119–135.

Marron, J.S., Adak, S., Johnstone, I.M., Neumann, M. and Patil, P.N. (1998) Exact risk analysis of wavelet regression. *J. Comput. Graph. Statist.*, **7**, 278–309.

Ramlau-Hansen, H. (1983a) Smoothing counting process intensity by means of kernel functions. *Ann. Statist.*, **11**, 453–466.

Ramlau-Hansen, H. (1983b) The choice of a kernel function in the graduation of counting process intensities. *Scand. Actuarial J.*, 165–182.

Rogers, L.C.G. and Williams, D. (1987) *Diffusions, Markov Processes and Martingales, Volume 2: Itô Calculus.* Chichester: Wiley.

Shorack, G.R. and Wellner, J. (1986) *Empirical Processes.* New York: Wiley.

Truong, Y.K. and Patil, P.N. (1996) On a local property of wavelet estimators involving time series. *J. Nonparametr. Statist.*, **6**, 143–156.

Wood, A.T.A. (1999) Rosenthal's inequality for point process martingales. *Stochastic Process. Appl.*, **81**, 213–246.

Wood, A.T.A. (2001) Acknowledgement of priority. Comment on 'Rosenthal's inequality for point process martingales'. *Stochastic Process. Appl.*, **93**, 349.

Received July 2001 and revised October 2002