Some limit results on the Haar-Fisz transform for inhomogeneous Poisson signals

Thorsten Schmidt[1] and Ling Xu[2]

August 24, 2007

forthcoming in *Journal for Analysis and its Applications*

One method to estimate the intensity of inhomogeneous Poisson processes, suggested in Fryzlewicz and Nason (2004), is first to preprocess the data using the so-called Haar-Fisz transform and then to apply wavelet methods to the outcome of the first step. For this procedure it is necessary, that the outcomes of the preprocessing step can be approximated by a normal distribution. In this paper we establish the necessary weak convergence results for the case of inhomogeneous Poisson processes which show that the outcome of the preprocessed data can be approximated by Gaussian random variables and wavelet shrinkage with a global threshold may be applied. A small simulation studies the application to shot-noise models. It suggests that this method is able to detect small peaks while at the same time it does not over-smooth large peaks in comparison with kernel estimators or standard wavelet estimators.

Keywords: inhomogeneous Poisson process, intensity estimation, wavelets, shot-noise process

1 Introduction

Poisson processes have a long history in insurance and finance as risk-arrival processes and are well analysed. The homogeneous case is obviously the most convenient one as far as estimation is concerned. However, in practice quite often inhomogeneous processes are more suitable due to seasonalities or changes in the considered environment and so on. In particular, this research was inspired by the work of Dassios and Jang (2003), where the pricing of reinsurance claims which are subject to catastrophes has been analysed. If a catastrophe occurs, the number of claims rises sharply, but after a time this effect fades away. The aim is at estimating this intensity from the claim data. Other applications of shot-noise processes in finance include Gaspar and Schmidt (2005) and Schmidt and Stute (2007).

More precisely, this paper considers inhomogeneous Poisson processes and proposes to estimate the intensity using wavelets. The approach considers a method proposed in Fryzlewicz and Nason (2004) and gives the necessary weak convergence results for the case of inhomogeneous Poisson processes. This generalizes the results in Fryzlewicz and Nason (2004), which obtained weak convergence only for the homogeneous case.

The considered method follows ideas which go back to Fisz (1955). The key tool is a transformation, the so-called *Haar-Fisz transform*, which transforms the observations to approximately normal distributed random variables. In his paper, Fisz used this property to test the hypothesis that two Poisson variables have equal means, and the hypothesis that their means are both equal

[1]Department of Mathematics, University of Leipzig, D-04081 Leipzig, Germany. Email: thorsten.schmidt@math.uni-leipzig.de

[2]Department of Mathematics, University of Leipzig, 04081 Leipzig, Germany. Email: ling.xu@math.uni-leipzig.de

to given number. Later, Fryzlewicz and Nason proposed in Fryzlewicz and Nason (2004) the analyzed algorithm. One first preprocesses a vector of Poisson random variables (rvs) using a nonlinear wavelet-based transformation and then treat the preprocessed vector as if it was Gaussian. In Fryzlewicz and Nason (2004) it was proved that, in the case of a homogeneous Poisson process, the transformed vector is approximately normal and the elements are asymptotically uncorrelated. We extend their results to the inhomogeneous case.

The estimation procedure, as considered here, consists of two steps: the preprocessing step and the wavelet analysis of the preprocessed part. In this paper we provide theoretical considerations on the first step. Fryzlewicz (2007) provides a mean-square consistency result for the complete estimation procedure in a more general setting of the data-driven wavelet-Fisz estimation. However, asymptotic normality of the preprocessed vector is not considered, which is the topic of this work. The extension of the Haar-Fisz transform to other wavelets is considered in Jansen (2006) which also provide some partial limit results. Fryzlewicz and Nason (2006) apply the Haar-Fisz transform and related ideas for estimating the local variance of a locally stationary Gaussian time series.

The paper is organized as follows. In section two we describe the estimation procedure. In section three we establish weak convergence results for inhomogeneous Poisson processes. In section four we give some simulation results which illustrate the applicability of the chosen approach to intensities of the shot-noise type and compare it to kernel estimators.

2 The procedure

The main goal of this paper is a suitable transformation of the observed Poisson process which will allow the application of well-established wavelet techniques.

Inhomogeneous Poisson processes An inhomogeneous Poisson process P is pure-jump process with independent and Poisson distributed increments. It is determined by its intensity $\lambda : \mathbb{R}^+ \mapsto \mathbb{R}^+$ and we have for $0 \leq s < t$

$$\mathbb{P}(P_t - P_s = k) = \exp\Big(- \int_s^t \lambda(s)ds \Big) \frac{(\int_s^t \lambda(s)ds)^k}{k!}, \quad k = 0, 1, 2, \dots$$

We assume that we observe P on the interval $[0, T]$ and aim at estimating λ on this interval.

Haar wavelets play a key role in the used transformation. For the reader's convenience we give a short introduction. For a full treatment of wavelets we refer to Mallat (1999).

Haar wavelets Consider $J \in \mathbb{N}$. We divide the observation interval $[0, T]$ in $2^J =: N$ intervals of equal length. The Haar wavelet filters are given by a family of vectors $\boldsymbol{\psi}^{j,l} \in \mathbb{R}^N$ with $j \in \{1, \dots, J\}$ and $l \in \{1, \dots, 2^{J-j}\}$. Here j is a scale parameter and l is a location parameter. It is more convenient to consider $k = k(l, j) = (l-1)2^j + 1$ instead of l. Throughout the paper we simply write k instead of $k(l, j)$. The Haar wavelet filters $\boldsymbol{\psi}^{j,l} = (\psi_1^{j,l}, \dots, \psi_N^{j,l})$ are defined by

$$\psi_n^{j,l} = \mathbf{1}_{\{n \in [k, k+2^{j-1})\}} - \mathbf{1}_{\{n \in [k+2^{j-1}, k+2^j)\}}.$$

For a given degree of fineness N, we also need to take care of the overall scale on the considered intervals. In analogy to $\boldsymbol{\psi}$ we therefore introduce the Haar scaling filters $\boldsymbol{\phi}^{j,l} \in \mathbb{R}^N$ defined by

$$\phi_n^{j,l} = \mathbf{1}_{\{n \in [k, k+2^j)\}}.$$

To have an easy access to the used indices we define for $J_0 \in \{1, \dots, J\}$ the set

$$A^J_{J_0} := \{(j,l) | (j,l) \in \mathbb{Z}^2 \text{ and } 1 \le j \le J_0,\ 1 \le l \le 2^{J-j}\}.$$

Clearly, the Haar wavelet filters and Haar scaling filter satisfy $\boldsymbol{\psi}^{j,l} = \boldsymbol{\phi}^{j-1,2l-1} - \boldsymbol{\phi}^{j-1,2l}$ and $\boldsymbol{\phi}^{j,l} = \boldsymbol{\phi}^{j-1,2l-1} + \boldsymbol{\phi}^{j-1,2l}$ for any $(j,l) \in A^J_J \setminus A^J_1$.

Remark 2.1. Haar filters constitute an orthogonal basis of $\mathbb{R}^N$. More precisely, we have that[3]

$$\langle \boldsymbol{\psi}^{j,l}, \boldsymbol{\phi}^{J,1} \rangle = \langle \boldsymbol{\psi}^{j,l}, \mathbf{1} \rangle = 0, \qquad \langle \boldsymbol{\psi}^{j,l}, \boldsymbol{\psi}^{s,t} \rangle = 2^j \mathbf{1}_{\{(j,l)=(s,t)\}}.$$

Letting $\tilde{\boldsymbol{\psi}}^{j,l} = 2^{-j/2}\ \boldsymbol{\psi}^{j,l}$ and $\tilde{\boldsymbol{\phi}}^{j,l} = 2^{-j/2}\ \boldsymbol{\phi}^{j,l}$ we find that $\{\tilde{\boldsymbol{\phi}}^{J,1}, \tilde{\boldsymbol{\psi}}^{j,l} : (j,l) \in A^J_J\}$ is an orthonormal basis of $\mathbb{R}^N$.

The Haar-Fisz transformation The transformation we use was proposed by Fryzlewicz and Nason (2004). We give a formulation which is suitable for obtaining convergence results. As the wavelet filters constitute a basis of $\mathbb{R}^N$ we have for $\mathbf{v} \in \mathbb{R}^N$ the following decomposition:

$$\mathbf{v} = \mathcal{A}_{J,1}(\mathbf{v})\ \tilde{\boldsymbol{\phi}}^{J,1} + \sum_{(j,l) \in A^J_J} \mathcal{D}_{j,l}(\mathbf{v})\ \tilde{\boldsymbol{\psi}}^{j,l}$$

with

$$\mathcal{A}_{J,1}(\mathbf{v}) = \langle \mathbf{v}, \tilde{\boldsymbol{\phi}}^{J,1} \rangle, \quad \mathcal{D}_{j,l}(\mathbf{v}) = \langle \mathbf{v}, \tilde{\boldsymbol{\psi}}^{j,l} \rangle.$$

Here, $\mathcal{A}_{J,1}(\mathbf{v})$ gives the overall scale of the wavelet decomposition while $\mathcal{D}_{j,l}(\mathbf{v})$ refer to the fine structure on the considered detail level.

Now we are in the position to precisely state the transform.

Definition 2.2. *The Haar-Fisz transform is the function $\mathcal{F} : \mathbb{R}^N \to \mathbb{R}^N$ defined by*

$$\mathcal{F}(\mathbf{v}) = \mathcal{A}_{J,1}(\mathbf{v})\ \tilde{\boldsymbol{\phi}}^{J,1} + \sum_{(j,l) \in A^J_J} \mathcal{G}_{j,l}(\mathbf{v})\ \tilde{\boldsymbol{\psi}}^{j,l}, \tag{1}$$

with

$$\mathcal{G}_{j,l}(\mathbf{v}) = \frac{\langle \mathbf{v}, \boldsymbol{\psi}^{j,l} \rangle}{\langle \mathbf{v}, \boldsymbol{\phi}^{j,l} \rangle^{\frac{1}{2}}} \mathbf{1}_{\{\langle \mathbf{v}, \boldsymbol{\phi}^{j,l} \rangle > 0\}}.$$

We will often refer to components of the vector $\mathcal{F}(\mathbf{v})$ and therefore simply set $\mathcal{F}(\mathbf{v}) = (\mathcal{F}_1(\mathbf{v}), \dots, \mathcal{F}_N(\mathbf{v}))$.

The reason for using $\mathcal{G}$ instead of $\mathcal{D}$ is simply normalization. For a discretization level J set $\Delta := T/N$. Then the discretized observation $\boldsymbol{\xi}_N := (P_\Delta - P_0, \dots, P_{N\Delta} - P_{(N-1)\Delta})$ is a vector of independent Poisson rvs with mean $\boldsymbol{\lambda}_N$ where $\lambda_n = \int_{(n-1)\Delta}^{n\Delta} \lambda(s)\, ds$. Then, under some assumptions, $\mathcal{G}_{j,l}(\boldsymbol{\xi}_N)$ will converge to a normally distributed rv with unit variance as we will see later. Hence we are able to apply well-established wavelet denoising techniques with global thresholding.

2.1 The estimation procedure

For fixed N we consider the discretized observation $\boldsymbol{\xi}_N$ as above. The estimation procedure consists of the following three steps proposed by Fryzlewicz and Nason (2004):

[3]Denote by $\langle \cdot, \cdot \rangle$ the inner product in $\mathbb{R}^N$.

1. Given the vector $\boldsymbol{\xi}_N$ of independent Poisson rvs, we first preprocess it using $\mathcal{F}(\boldsymbol{\xi}_N)$. As will be shown in Theorem 3.4, $\mathcal{F}(\boldsymbol{\xi}_N)$ is $\mathcal{F}(\boldsymbol{\lambda}_N)$ plus approximately white noise.
2. Denoise $\mathcal{F}(\boldsymbol{\xi}_N)$ with standard wavelet techniques and denote the outcome by $\tilde{\mathcal{F}}(\boldsymbol{\xi}_\mathbf{N})$. These steps are outlined detailed in Section 4.
3. The inverse Haar-Fisz transform gives the estimator: $\mathcal{F}^{-1}(\tilde{\mathcal{F}}(\boldsymbol{\xi}_N))$.

Remark 2.3. The estimating procedure basically consists of two steps. The first step, is the transformation of the data, given by the function $\mathcal{F}$. The second step consists in the application of standard wavelet techniques. Finally , the inverse Haar-Fisz transform is applied.

The second step requires that the transformation $\mathcal{F}$ has an outcome which can be approximated by normally distributed random variables. This paper mainly concentrates on this and gives the necessary weak convergence results in our main result, Theorem 3.4.

3 Convergence results for the Haar-Fisz transform

This section gives precise results for convergence of the preprocessing step. By $\xrightarrow{\mathscr{L}}$ we denote convergence in distribution. Central to the following argumentation will be the following result provided in Fisz (1955).

Theorem 3.1. *Let ξ_1 and ξ_2 be independent and Poisson distributed with intensity λ_1 and λ_2, respectively. If $\lambda_1 \to \infty$, $\lambda_2 \to \infty$ and $\frac{\lambda_1}{\lambda_2} \to 1$, then we have that*

$$\eta := \frac{\xi_1 - \xi_2}{\sqrt{\xi_1 + \xi_2}}\, \mathbf{1}_{\{\xi_1+\xi_2>0\}} - \frac{\lambda_1 - \lambda_2}{\sqrt{\lambda_1 + \lambda_2}} \xrightarrow{\mathscr{L}} \mathcal{N}(0,1).$$

Furthermore, from the proof of this result we learn that

$$\frac{\xi_1 - \xi_2}{\sqrt{\xi_1 + \xi_2}}\mathbf{1}_{\{\xi_1+\xi_2>0\}} - \frac{\xi_1 - \xi_2}{\sqrt{\lambda_1 + \lambda_2}} \xrightarrow{\mathbb{P}} 0, \tag{2}$$

where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability. In Lemma 6.2 we proof uniform L_3-boundedness of $\{\eta : \lambda_1, \lambda_2 > 1\}$ and so from convergence to $\mathcal{N}(0,1)$ it follows that

$$\mathbb{E}(\eta) \to 0,\ \mathbb{E}(\eta^2) \to 1, \quad \text{and} \quad \mathrm{Var}(\eta) \to 1. \tag{3}$$

It will prove useful to have a convenient representation of vectors like $\boldsymbol{\psi}$, $\boldsymbol{\phi}$, $\mathcal{G}$ and others. In the following we therefore write simply $(\mathcal{G}_1, \dots, \mathcal{G}_{N-1})$ for $(\mathcal{G}_{1,1}, \mathcal{G}_{1,2}, \dots, \mathcal{G}_{1,2^{J-1}}, \mathcal{G}_{2,1}, \dots, \mathcal{G}_{2,2^{J-2}}, \dots, \mathcal{G}_{J,1})$ and similar for $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$. The different indexation should suffice to avoid confusion. Set $\mathcal{G} := (\mathcal{G}_1, \dots, \mathcal{G}_{N-1})$ and let $c_i^n := \tilde{\psi}_n^i$ and $\mathbf{c}^n = (c_1^n, \dots, c_{N-1}^n)$, $n = 1, \dots, N$. Note that then

$$\left(\begin{pmatrix} \mathbf{c}^1 \\ 2^{-\frac{J}{2}} \end{pmatrix}, \dots, \begin{pmatrix} \mathbf{c}^N \\ 2^{-\frac{J}{2}} \end{pmatrix} \right) = (\tilde{\boldsymbol{\psi}}^1, \dots, \tilde{\boldsymbol{\psi}}^{N-1}, \tilde{\boldsymbol{\phi}}^{J,1})^\top. \tag{4}$$

First, we derive a useful representation of $\mathcal{F}(\mathbf{v})$. In this simpler notation, (1) reads

$$\mathcal{F}_n(\mathbf{v}) = 2^{-J/2}\mathcal{A}_{J,1}(\mathbf{v}) + \langle \mathbf{c}^n, \mathcal{G}(\mathbf{v}) \rangle. \tag{5}$$

A first step is the following generalization of Theorem 3.1. By $\mathbf{I}_N$ we denote the identity matrix.

Theorem 3.2. *Consider a vector* $\boldsymbol{\xi}$ *of independent Poisson random variables with mean* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$*. If* $\lambda_i \to \infty$ *and* $|\frac{\lambda_i}{\lambda_j} - 1| \to 0$ *for all* i *and* j*, then*

$$\mathcal{G}(\boldsymbol{\xi}) - \mathcal{G}(\boldsymbol{\lambda}) \xrightarrow{\mathscr{L}} \mathcal{N}(0, \mathbf{I}_N).$$

Proof. The proof mainly relies on (2). To apply this result, we denote by $\tilde{\mathcal{G}}_i(\boldsymbol{\xi}) = \langle \boldsymbol{\xi}, \boldsymbol{\psi}^i \rangle \cdot \langle \boldsymbol{\lambda}, \boldsymbol{\phi}^i \rangle^{-1/2}$ and set $\tilde{\mathcal{G}} = (\tilde{\mathcal{G}}_1, \ldots, \tilde{\mathcal{G}}_{N-1})$. Note that

$$\tilde{\mathcal{G}}_i(\boldsymbol{\xi}) - \mathcal{G}_i(\boldsymbol{\lambda}) = \frac{\langle \boldsymbol{\xi} - \boldsymbol{\lambda}, \boldsymbol{\psi}^i \rangle}{\langle \boldsymbol{\lambda}, \boldsymbol{\phi}^i \rangle^{1/2}} = \langle \frac{\boldsymbol{\xi} - \boldsymbol{\lambda}}{\sqrt{\lambda_1}} , \frac{\sqrt{\lambda_1} \boldsymbol{\psi}^i}{\langle \boldsymbol{\lambda}, \boldsymbol{\phi}^i \rangle^{\frac{1}{2}}} \rangle$$

hence $\tilde{\mathcal{G}}(\boldsymbol{\xi}) - \mathcal{G}(\boldsymbol{\lambda})$ is of the form $\frac{(\boldsymbol{\xi} - \boldsymbol{\lambda})^\top}{\sqrt{\lambda_1}} \boldsymbol{\Psi}$ with appropriate $\boldsymbol{\Psi} \in \mathbb{R}^{N \times N}$. Furthermore, $\frac{\sqrt{\lambda_1} \boldsymbol{\psi}^i}{\langle \boldsymbol{\lambda}, \boldsymbol{\phi}^i \rangle^{\frac{1}{2}}} \to \frac{\boldsymbol{\psi}^i}{\langle \mathbf{1}, \boldsymbol{\phi}^i \rangle^{\frac{1}{2}}} = \tilde{\boldsymbol{\psi}}^i$ hence $\boldsymbol{\Psi}^\top \boldsymbol{\Psi} \to \mathbf{I}_N$.

On the other hand, using that the components of $\boldsymbol{\xi}$ are independent, we obtain by Lemma 6.1 $\frac{\boldsymbol{\xi} - \boldsymbol{\lambda}}{\sqrt{\lambda_1}} \xrightarrow{\mathscr{L}} \mathcal{N}(0, \mathbf{I_N})$ and we therefore have shown that $\tilde{\mathcal{G}}(\boldsymbol{\xi}) - \mathcal{G}(\boldsymbol{\lambda}) \xrightarrow{\mathscr{L}} \mathcal{N}(0, \mathbf{Id}_N)$. Using (2), we have $\mathcal{G}_i(\boldsymbol{\xi}) - \tilde{\mathcal{G}}_i(\boldsymbol{\xi}) \xrightarrow{\mathbb{P}} 0$, and so we get $\mathcal{G}(\boldsymbol{\xi}) - \tilde{\mathcal{G}}(\boldsymbol{\xi}) \xrightarrow{\mathbb{P}} \mathbf{0}$ and the desired result follows. ■

Up to now, N was always fixed. Our main result, Theorem 3.4 considers the case where N goes to infinity and the λ_i relate appropriately. Recall that we want to approximate a rescaled Poisson random variable by a Gaussian distribution and it is therefore necessary that the intensity is sufficiently high. In the inhomogeneous case, additionally the intensities must scale properly which leads to the following Assumption 3.3. We therefore introduce ρ and let λ_i be increasing with ρ. Besides letting the grid getting finer and finer by increasing N we also increase the intensity of the Poisson signals by increasing ρ.

The precise assumptions are as follows. Recall that the observation $\boldsymbol{\xi}_N$ was Poisson distributed with mean $\boldsymbol{\lambda}_N$. We need some uniform convergence of the components of $\boldsymbol{\lambda}_N$ and to be able to state this conveniently we assume that $\boldsymbol{\lambda}_N = \boldsymbol{\lambda}_N(\rho) = (\lambda_1(\rho), \ldots, \lambda_N(\rho))$ with $\rho > 0$. Then we examine convergence for N and ρ going to infinity.

Assumption 3.3. *Assume that for any* N *and* $\epsilon, \delta > 0$ *there exits a* ρ_0*, s.t. for* $\rho > \rho_0$

$$\inf_{1 \le i \le N} \lambda_i(\rho) > \delta, \qquad \sup_{1 \le i,j \le N} |\frac{\lambda_i(\rho)}{\lambda_j(\rho)} - 1| < \epsilon, \tag{6}$$

where ϵ, δ, ρ_0 *do not depend on* N*.*

We are ready to state the main result.

Theorem 3.4. *Assume Assumption 6 holds and* $\langle \boldsymbol{\lambda}_N, \boldsymbol{\phi}^{J,1} \rangle^{1/2}/N \to 0$ *for* $\rho, N \to \infty$*. Then*

$$\mathcal{F}_n(\boldsymbol{\xi}_N) - \mathcal{F}_n(\boldsymbol{\lambda}_N) \xrightarrow[\rho, N \to \infty]{\mathscr{L}} \mathcal{N}(0, 1) \tag{7}$$

as well as $\operatorname{Cov}(\mathcal{F}_m(\boldsymbol{\xi}_N), \mathcal{F}_n(\boldsymbol{\xi}_N)) \to 0$ *for any* $m \neq n$*.*

Example 3.5. Consider the case where $\lambda_i(\rho) = \rho$ for all i. Then, Assumption 3.3 is trivially satisfied. Next,

$$\frac{\langle \boldsymbol{\lambda}_N, \boldsymbol{\phi}^{J,1} \rangle^{1/2}}{N} = \frac{\sqrt{N\rho}}{N} = \sqrt{\frac{\rho}{N}}.$$

If we choose, for example, $\rho = \sqrt{N}$ the assumptions in Theorem 3.4 are satisfied.

The proof extends the ideas in Fryzlewicz and Nason (2001) to the inhomogeneous setting.

Proof. Let $\mathbf{z} := \mathcal{G}(\boldsymbol{\xi}_N) - \mathcal{G}(\boldsymbol{\lambda}_N)$. By (5), we obtain

$$\mathcal{F}_n(\boldsymbol{\xi}_N) - \mathcal{F}_n(\boldsymbol{\lambda}_N) = 2^{-J/2}(\mathcal{A}_{J,1}(\boldsymbol{\xi}_N) - \mathcal{A}_{J,1}(\boldsymbol{\lambda}_N)) + \langle \mathbf{c}^n, \mathbf{z}\rangle := M + X_n. \tag{8}$$

First, we show that M converges to zero. To this, note that $\langle \boldsymbol{\xi}_N, \boldsymbol{\phi}^{J,1}\rangle$ is Poisson with mean $\langle \boldsymbol{\lambda}_N, \boldsymbol{\phi}^{J,1}\rangle)$. Hence, by Lemma 6.1 in the appendix, $Y := (\langle \boldsymbol{\xi}_N, \boldsymbol{\phi}^{J,1}\rangle - \langle \boldsymbol{\lambda}_N, \boldsymbol{\phi}^{J,1}\rangle)\ \langle \boldsymbol{\lambda}_N, \boldsymbol{\phi}^{J,1}\rangle^{-1/2}$ converges in distribution to $\mathcal{N}(0,1)$. Set $\vartheta = \vartheta(N, \boldsymbol{\lambda}_N) := \langle \boldsymbol{\lambda}_N, \boldsymbol{\phi}^{J,1}\rangle^{1/2}/N$. Then we obtain that

$$M = Y\vartheta \xrightarrow{\mathbb{P}} 0.$$

Furthermore, as $\mathrm{Var}(Y) = 1$, we have that

$$\mathrm{Var}(M) = \mathrm{Var}(Y)\ \vartheta^2 \longrightarrow 0. \tag{9}$$

Second, we consider the convergence of X_n. For this part, the condition that ϑ converges to zero is not necessary and we directly show that X_n converges to $\mathcal{N}(0,1)$ as both ρ and J go to infinity. The main idea is to split X_n up in a part which converges to a r.v. which is arbitrary close to standard normal and a part which is arbitrary close to zero. Fix a J_0 and consider $J > J_0$. Recall the definition of $\mathbf{c}^n$ prior to (4) and set $\mathbf{c}_1^n = (c_1^n, \dots, c_{2^J-2^{J-J_0}}^n)^\top$ as well as $\mathbf{c}_2^n = (c_{2^J-2^{J-J_0}+1}^n, \dots, c_{N-1}^n)^\top$. We similarly divide $\mathbf{z}$ in $\mathbf{z}_1$ and $\mathbf{z}_2$. Then

$$X_n = X_1^n + X_2^n, \tag{10}$$

where $X_i^n = \langle \mathbf{c}_i^n, \mathbf{z}_i\rangle$. Note that X_1^n does not depend on J. Denote by $\mathbf{b}(n) = (b_1(n), b_2(n), \dots)$ the binary representation of the integer n. Then[4]

$$\begin{aligned} X_1^n &= \sum_{j=1}^{J_0} (-1)^{b_j(n)}\ 2^{-j/2}\left(\mathcal{G}_{j,\lceil n/2^j\rceil}(\boldsymbol{\xi}_N) - \mathcal{G}_{j,\lceil n/2^j\rceil}(\boldsymbol{\lambda}_N)\right) \\ &\xrightarrow{\mathscr{L}} \mathcal{N}(0, \sum_{j=1}^{J_0}\left[(-1)^{b_j(n)} 2^{\frac{-j}{2}}\right]^2) = \mathcal{N}(0, 1-2^{-J_0}) \end{aligned}$$

follows.

On the other hand, we have that by the definition of $\mathbf{c}^n$ that $\|\mathbf{c}_2^n\|_1^2 \le (\sqrt{2}+1)2^{-J_0/2}$. A consequence of (18) provided in the appendix is that there exists a ρ_0, such that for all $\rho > \rho_0$

$$\sup_{J,(j_i,l_i)\in A_J^J} \mathbb{E}[(z_i)^2] \le 2.$$

Hence for $\rho > \rho_0$,

$$\mathbb{E}[(X_2^n)^2] \le 2\ \|\mathbf{c}_2^n\|_1^2 \le \tilde{c}\ 2^{-J_0/2}$$

with $\tilde{c} = 2(\sqrt{2}+1)$. By the Markov-inequality we get for any $\epsilon > 0$ and $\rho > \rho_0$ that

$$\mathbb{P}(|X_2^n| > \epsilon) \le \frac{1}{\epsilon^2}\mathbb{E}[(X_2^n)^2] \le \epsilon^{-2}\ \tilde{c}\ 2^{-J_0/2}. \tag{11}$$

[4]Denote by $\lceil x\rceil$ the smallest integer $i \ge x$. Note that by the definition of Haar wavelets we have $c_i^n = \tilde{\psi}_n^i$ and for some j, l

$$\tilde{\psi}_n^{j,l} = (-1)^{b_j(n)} 2^{\frac{-j}{2}} \mathbf{1}_{\{l=\lceil n/2^j\rceil\}}.$$

These two results suffice for the claim, as we will show now. For any x we have that for $\rho > \rho_0$

$$\mathbb{P}(X_n \le x) = \mathbb{P}(X_n \le x, |X_2^n| \le \epsilon) + \mathbb{P}(|X_2^n| > \epsilon).$$

For the first term we use that $\{X_1^n + X_2^n \le x, |X_2^n| \le \epsilon\} \subset \{X_1^n \le x + \epsilon\}$ and obtain

$$\begin{aligned}\lim_{J,\rho\to\infty} \mathbb{P}(X_n \le x) &\le \lim_{\rho\to\infty} \mathbb{P}(X_1^n \le x+\epsilon) + \epsilon^{-2}\, \tilde{c}\, 2^{-J_0/2} \\ &= \Phi\left(\frac{x+\epsilon}{\sqrt{1-2^{-J_0}}}\right) + \epsilon^{-2}\, \tilde{c}\, 2^{-J_0/2}.\end{aligned}$$

As J_0 was arbitrary, we obtain $\lim_{J,\rho\to\infty} \mathbb{P}(X_n \le x) \le \Phi(x)$. For the other inclusion, observe that for $\rho > \rho_0$

$$\begin{aligned}\mathbb{P}(X_n \le x) &\ge \mathbb{P}(X_1^n + X_2^n \le x, |X_2^n| \le \epsilon) \\ &\ge \mathbb{P}(X_1^n + \epsilon \le x) + \mathbb{P}(|X_2^n| \le \epsilon) - 1 \\ &\ge \Phi(x-\epsilon) + (1 - \epsilon^{-2}\tilde{c}2^{-J_0}) - 1 = \Phi(x-\epsilon) - \epsilon^{-2}\tilde{c}2^{-J_0}.\end{aligned}$$

Again, as J_0 and ϵ are arbitrary, we obtain $\lim_{J,\rho\to\infty} \mathbb{P}(X_n \le x) \ge \Phi(x)$ and therefore the limit equals $\Phi(x)$.

The second part of theorem is a statement about covariances. The essential ingredient is again Lemma 6.3. First, observe that

$$\begin{aligned}|\operatorname{Cov}(f_m, f_n)| &= |\operatorname{Var}(M) + \operatorname{Cov}(M, X_m) + \operatorname{Cov}(M, X_n) + \operatorname{Cov}(X_m, X_n)| \\ &\le \operatorname{Var}(M) + \operatorname{Var}^{\frac{1}{2}}(M)[\operatorname{Var}^{\frac{1}{2}}(X_m) + \operatorname{Var}^{\frac{1}{2}}(X_n)] + |\operatorname{Cov}(X_m, X_n)|\end{aligned}$$

Recall that $\operatorname{Var}(M) \to 0$. In the following we show that $\operatorname{Cov}(X_m, X_n) \to 0$, as $\rho, J \to \infty$ and that $\operatorname{Var}(X_m)$ is bounded. As in (7), we choose a J_0 and use the decomposition $X_n = X_1^n + X_2^n$. With

$$|\operatorname{Cov}(X_m, X_n)| \le \sum_{i,j=1}^{2} |\operatorname{Cov}(X_i^m, X_i^n)| \tag{12}$$

we consider, letting $N_0 = 2^J - 2^{J-J_0}$,

$$\begin{aligned}|\operatorname{Cov}(X_1^m, X_1^n)| &= |\mathbf{c}_1^{m\top} \operatorname{Cov}(\mathbf{z}_1, \mathbf{z}_1)\mathbf{c}_1^n| \\ &\le |\mathbf{c}_1^{m\top}(\operatorname{Cov}(\mathbf{z}_1, \mathbf{z}_1) - \mathbf{I}_{N_0})\mathbf{c}_1^n| + |\langle \mathbf{c}_1^m, \mathbf{c}_1^n\rangle|.\end{aligned}$$

Recalling that $c_i^n = \tilde{\psi}_n^i$, we have that

$$\|\mathbf{c}_2^n\|_1 = \sum_{(j,l)\in A_J^J \setminus A_{J_0}^J} |\tilde{\psi}_n^{j,l}| = \sum_{j=J_0+1}^{J} \sum_{l=1}^{2^{J-j}} |\tilde{\psi}_n^{j,l}| = \sum_{j=J_0+1}^{J} 2^{-\frac{j}{2}} \le (\sqrt{2}+1)2^{-J_0/2}$$

and similarly, $\|\mathbf{c}_1^n\|_1 \le \sqrt{2}+1$ and $\|\mathbf{c}_2^n\|_2^2 = 2^{-J_0} - 2^{-J}$. With $|\langle \mathbf{c}_2^m, \mathbf{c}_2^n\rangle| \le \|\mathbf{c}_2^m\|_2^{1/2}\|\mathbf{c}_2^n\|_2^{1/2}$ we obtain

$$|\langle \mathbf{c}_1^m, \mathbf{c}_1^n\rangle| = |\langle \mathbf{c}^m, \mathbf{c}^n\rangle - \langle \mathbf{c}_2^m, \mathbf{c}_2^n\rangle| \le 2^{-J_0}.$$

On the other hand, from Lemma 6.3 we obtain that for given ϵ there exists ρ_0, such that for $\rho > \rho_0$

$$\sup_{J\ge J_0, 1\le i,k\le 2^J(1-2^{J_0})} |\operatorname{Cov}(z_i, z_k) - \mathbf{1}_{\{i=k\}}| < \epsilon$$

and hence

$$|\operatorname{Cov}(X_1^m, X_1^n)| \leq \ \epsilon \, \|\mathbf{c}_1^m\|_1 \|\mathbf{c}_1^n\|_1 + 2^{-J_0} = \epsilon(1+\sqrt{2})^2 + 2^{-J_0}.$$

For the remaining terms, observe that Lemma 6.3 also gives

$$\sup_{J \geq J_0, 1 \leq i < 2^J} |\operatorname{Var}(z_i) - 1| < \epsilon$$

and so

$$\begin{aligned}
&\operatorname{Var}(X_2^n) = \mathbf{c}_2^{n\top} \operatorname{Cov}(\mathbf{z}_2, \mathbf{z}_2) \mathbf{c}_2^n \leq (1+\epsilon)\|\mathbf{c}_2^n\|_1^2 \leq (1+\epsilon)(2^{-J_0} - 2^{-J}), \\
&\operatorname{Var}(X_1^n) \leq (1+\epsilon)\|\mathbf{c}_1^n\|_1^2 \leq (\sqrt{2}+1)^2(1+\epsilon).
\end{aligned}$$

This directly implies

$$\begin{aligned}
&|\operatorname{Cov}(X_1^m, X_2^n)| \leq \operatorname{Var}^{\frac{1}{2}}(X_1^m) \operatorname{Var}^{\frac{1}{2}}(X_2^n) \leq (\sqrt{2}+1)(1+\epsilon)\sqrt{2^{-J_0} - 2^{-J}}, \\
&|\operatorname{Cov}(X_2^m, X_2^n)| \leq (1+\epsilon)\,(2^{-J_0} - 2^{-J})
\end{aligned}$$

and as J_0 and ϵ were arbitrary, we obtain that all terms in (12) converge to zero as $J, \rho \to \infty$. Furthermore, as $\operatorname{Var}(X_1^m)$ is bounded, we also have that $\operatorname{Var}(M)\operatorname{Var}(X_1^m) \to 0$ and so we conclude $|\operatorname{Cov}(f_n, f_m)| \to 0$. □ ■

4 Simulation results

In this section, we compare the performance of the estimation procedure to kernel-based estimators in the case where the intensity is of the shot-noise type. Further, comprehensive simulation results may be found in Fryzlewicz and Nason (2004).

The analysed scenario is inspired by Dassios and Jang (2003). The authors discuss an insurance problem where the insurance claims arrive through a Cox process with shot noise intensity. Here, we concentrate on a somewhat simpler case, namely the case of inhomogeneous Poisson processes where the intensity has a shot noise form. The shot noise form is motivated by an occurring catastrophe which induces thereafter a high number of claims, but as time passes by this effect fades away. The question analysed is: if we observe the number of claims, how can we estimate the claim intensity. Our method is nonparametric. If, as analysed in Dassios and Jang (2005), one would like to specify a stochastic model for the intensity, one typically would use filtering methods to estimate the unknown intensity.

The denoising procedure. We shortly illustrate the denoising procedure. In principle, an arbitrary wavelet basis can be used. For convenience, we illustrate the denoising procedure using Haar wavelets. Recall the definitions of the normalized Haar wavelets $\tilde{\boldsymbol{\psi}}$ and $\tilde{\boldsymbol{\phi}}$ in Section 2. From Remark 2.1 we obtain the the set of normalized Haar wavelets constitutes an orthonormal basis of $\mathbb{R}^N$.

Then for a threshold $\delta > 0$, the estimator of $\mathbf{v} \in \mathbb{R}^N$ using a so-called hard thresholding takes the following form:

$$\hat{\mathbf{v}} = \sum_{(J_1,l) \in A_{J_1}^J \setminus A_{J_1-1}^J} \langle \mathbf{v}, \tilde{\boldsymbol{\phi}}^{J_1,l} \rangle \, \tilde{\boldsymbol{\phi}}^{J_1,l} + \sum_{(j,l) \in A_{J_1}^J} \langle \mathbf{v}, \tilde{\boldsymbol{\psi}}^{j,l} \rangle \mathbf{1}_{\{|\langle \mathbf{v}, \tilde{\boldsymbol{\psi}}^{j,l} \rangle| > \delta\}} \, \tilde{\boldsymbol{\psi}}^{j,l}. \tag{13}$$

Including the denoising steps in the whole procedure we estimate as follows:

1. *Apply transform.* First, we preprocess the data using the transform $\mathcal{F}$ and obtain $\mathbf{v} := \mathcal{F}(\boldsymbol{\xi}_N)$ which is approximately Gaussian.
2. *Discrete wavelet transform.* Choose a wavelet basis and a level J_1. We select threshold as follows: using the approximate standard normality of $\mathbf{v}$ we set the threshold to $\delta = \sqrt{2 \log N}$ and therewith compute denoised $\hat{\mathbf{v}}$ as in (13).
3. *Inverse transform.* Finally, the transformation is inverted and the estimator of the Poisson intensity is $\mathcal{F}^{-1}(\hat{\mathbf{v}})$.

One of the most important things is the choice of the proper wavelet scaling function[5]. Generally speaking, for appropriate denoising the wavelet scaling function should have properties which are similar to the original signal. Therefore, in our estimation we choose to denoise $\mathcal{F}(\boldsymbol{\xi}_N)$ with Daubechies-3 wavelets. The illustration above just uses Haar wavelets as we already introduced them. However, if we used Haar wavelets for the estimation, the result would be very poor.

The chosen estimation method is very well suited to estimate intensities which have extremely sharp spikes. However, when the procedure is used for the estimation of smoother intensities, its performance is as good as that of kernel methods. We illustrate this with some simulations.

Simulation results. In Figure 1 we compare the estimation method based on the Haar-Fisz transform with a kernel estimator of the intensity and a standard wavelet estimator. We rely on a standard, i.e. symmetric and smooth kernel (Gaussian) as well as we use standard, symmetric and smooth wavelets (Daubechies-3). The intensity is, as already mentioned, assumed to be subject to certain shocks and therefore has a peaky, shot-noise like shape. The number of jumps in the considered intervals clearly reflects this. Figure 1 gives four plots. The first and the second pair of plots differ in the true intensity. On the left side, we compare the estimation method based on the Haar-Fisz transform (H-F estimator) with the outcomes from a kernel estimator. On the right side we compare the H-F estimator with a standard wavelet approach not using the Haar-Fisz transform. All plots suggests the advantage of the H-F estimator over the other methods. In particular, this estimator is able to capture the large peaks without over-smoothing it. It is notably, that the standard wavelet approach is quite close to the H-F estimator. As was to be expected, the kernel estimator shows an over-smoothing of these peaks. Furthermore, in the left plot, the kernel estimator is not able to detect the two smaller peaks following each other.

5 Conclusion

This paper considers a wavelet based method for the estimation of the intensity of an inhomogeneous Poisson process. The procedure first transforms the observation to a vector which is approximately Gaussian and then applies well established wavelets methods. In this paper we establish the necessary weak convergence results which provide asymptotic normality of the preprocessed data. A small simulation study considers the application to inhomogeneous Poisson processes with intensities of the shot-noise type.

6 Appendix

First, we recall the following well-known result:

Lemma 6.1. *Let $\xi \sim Poiss(\lambda)$, then $\frac{\xi - \lambda}{\sqrt{\lambda}} \xrightarrow{\mathscr{L}} \mathcal{N}(0,1)$, as $\lambda \to \infty$.*

[5]We refer to Nason (2002) for more information on the choice of the wavelet scaling function in practice.

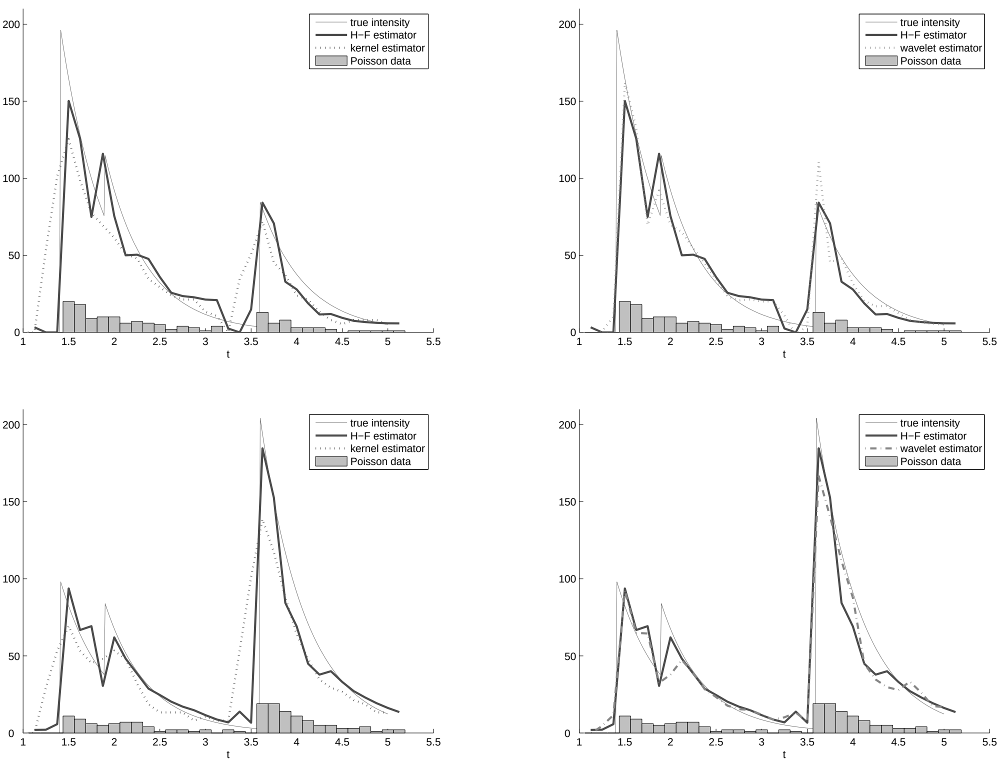

Figure 1: Simulation and estimation of inhomogeneous Poisson processes for two different, given intensities. The solid line gives the true intensity and the bars show the number of jumps in the considered intervals. The plots give the different estimators: a kernel estimator, a standard wavelet estimator and the estimator based on the Haar-Fisz transform (H-F estimator) as given in Section 2.1.

Next, we give a result on boundedness of certain transforms of Poisson r.v.

Lemma 6.2. *Consider, as in Theorem 3.1, independent* $\xi_i \sim \mathrm{Poisson}(\lambda_i)$*,* $i = 1, 2$ *and set*

$$\eta = \eta(\lambda_1, \lambda_2) := \frac{\xi_1 - \xi_2}{\sqrt{\xi_1 + \xi_2}} \, \mathbf{1}_{\{\xi_1+\xi_2>0\}} - \frac{\lambda_1 - \lambda_2}{\sqrt{\lambda_1 + \lambda_2}}.$$

Then $\{\eta(\lambda_1, \lambda_2) : \lambda_1, \lambda_2 \geq 1\}$ *is uniformly bounded in* L_3*.*

Proof. Set $Y(\lambda_1) := (\xi_1 - \lambda_1) \cdot \lambda_1^{-1/2}$. Then $\mathbb{E}(Y^6) = 15 + 25\frac{1}{\lambda_1} + \frac{1}{\lambda_1^2}$ and hence $\{Y(\lambda_1) : \lambda_1 \geq 1\}$ is uniformly bounded in L_6. Denote by $\| \cdot \|_p$ the norm in L_p, for any $p > 0$. Observe that

$$\| \, \eta \, \|_3 \leq \left\| \frac{\xi_1 - \xi_2}{\sqrt{\lambda_1 + \lambda_2}} \, \frac{\sqrt{\lambda_1 + \lambda_2}}{\sqrt{\xi_1 + \xi_2}} \mathbf{1}_{\{\xi_1+\xi_2>0\}} - \frac{\lambda_1 - \lambda_2}{\sqrt{\lambda_1 + \lambda_2}} \right\|_3$$

The second term is smaller than

$$\left\| \frac{\xi_1 - \lambda_1}{\sqrt{\lambda_2}} \right\|_3 + \left\| \frac{\xi_2 - \lambda_2}{\sqrt{\lambda_2}} \right\|_3$$

and hence it is uniformly L_3 bounded by the remark above. To the first term, we have that

$$\begin{aligned}
&\left\| \frac{\xi_1 - \xi_2}{\sqrt{\xi_1 + \xi_2}} \mathbf{1}_{\{\xi_1+\xi_2>0\}} - \frac{\xi_1 - \xi_2}{\sqrt{\lambda_1 + \lambda_2}} \right\|_3 \\
&\quad \leq \left\| (\xi_1 - \xi_2) \cdot \frac{\sqrt{\xi_1 + \xi_2} - \sqrt{\lambda_1 + \lambda_2}}{\sqrt{(\xi_1 + \xi_2)(\lambda_1 + \lambda_2)}} \mathbf{1}_{\{\xi_1+\xi_2>0\}} \right\|_3 \\
&\quad \leq \left\| (\xi_1 - \xi_2) \frac{\xi_1 + \xi_2 - (\lambda_1 + \lambda_2)}{\sqrt{(\xi_1 + \xi_2)(\lambda_1 + \lambda_2)}\,(\sqrt{\xi_1 + \xi_2} + \sqrt{\lambda_1 + \lambda_2})} \mathbf{1}_{\{\xi_1+\xi_2>0\}} \right\|_3 \\
&\quad \leq \left\| \frac{\xi_1 - \xi_2}{\xi_1 + \xi_2} \mathbf{1}_{\{\xi_1+\xi_2>0\}} \right\|_6^{1/2} \cdot \left\| \frac{\xi_1 + \xi_2 - (\lambda_1 + \lambda_2)}{\sqrt{\lambda_1 + \lambda_2}} \right\|_6^{1/2}.
\end{aligned}$$

The first term is smaller than 1 and the second term is uniformly bounded as noted above. Thus we have shown the result. □ ■

The following lemma is used in the proof of Theorem 3.4.

Lemma 6.3. *Assume that Assumption 3.3 holds. Then $\forall\, \epsilon > 0$, $\exists\, \rho_0$, such that $\forall\, \rho > \rho_0$,*

$$\sup_{2^J, 1 \leq i < 2^J} |\operatorname{Var}(\mathcal{G}_i(\boldsymbol{\xi}_{2^J})) - 1| < \epsilon. \tag{14}$$

Furthermore, for any $J_0 \in \mathbb{N}$, $\forall\, \epsilon > 0$, $\exists\, \rho_0$, such that $\forall\, \rho > \rho_0$,

$$\sup_{J \geq J_0,\, 1 \leq i \neq k \leq 2^J(1-2^{J_0})} |\operatorname{Cov}(\mathcal{G}_i(\boldsymbol{\xi}_{2^J}), \mathcal{G}_k(\boldsymbol{\xi}_{2^J}))| < \epsilon. \tag{15}$$

Proof. Consider some J and set $N = 2^J$. Recall the notational convention $(\mathcal{G}_1, \ldots, \mathcal{G}_{N-1}) = (\mathcal{G}_{1,1}, \mathcal{G}_{1,2}, \ldots, \mathcal{G}_{1,2^{J-1}}, \mathcal{G}_{2,1}, \ldots, \mathcal{G}_{2,2^{J-2}}, \ldots, \mathcal{G}_{J,1})$. So the supremum is over suitable i's, and hence this is equivalent to consider all suitable (j,l)'s which are precisely all $(j,l) \in A_J^J$. It is essential to observe that for increasing N but fixed (j,l) the vectors $\boldsymbol{\psi}^{j,l}$ get filled up with zeros, so the value of $\langle \boldsymbol{\psi}^{j,l}, \mathbf{v} \rangle$, for any $\mathbf{v}$ does not change if N is increased.

With $k = (l-1)2^j + 1$ set

$$\eta_1 = \langle \mathbf{1}_{[k, k+2^{j-1})}, \boldsymbol{\xi}_N \rangle, \;\; \eta_2 = \langle \mathbf{1}_{[k+2^{j-1}, k+2^j)}, \boldsymbol{\xi}_N \rangle,$$

and $\nu_i = \mathbb{E}(\eta_i)$. Observe that $\eta_i \sim \operatorname{Poiss}(\nu_i)$, and η_1 and η_2 are independent. Then

$$\eta = \mathcal{G}_i(\boldsymbol{\xi}_N) - \mathcal{G}_i(\boldsymbol{\lambda}_N) = \frac{\langle \boldsymbol{\psi}^i, \boldsymbol{\xi}_N \rangle}{\langle \boldsymbol{\phi}^i, \boldsymbol{\xi}_N \rangle^{\frac{1}{2}}} - \frac{\langle \boldsymbol{\psi}^i, \boldsymbol{\lambda}_N \rangle}{\langle \boldsymbol{\phi}^i, \boldsymbol{\lambda}_N \rangle^{\frac{1}{2}}} = \frac{\eta_1 - \eta_2}{\sqrt{\eta_1 + \eta_2}} - \frac{\nu_1 - \nu_2}{\sqrt{\nu_1 + \nu_2}}.$$

By (3) we have that $\operatorname{Var}(\eta) \to 1$ if ν_1, ν_2 both converge to infinity and ν_1/ν_2 converges to 1. Hence, $\forall\, \epsilon > 0$, $\exists\, \epsilon_0$, δ_0, such that $\forall\, \nu_1, \nu_2 > \delta_0$ and $|\frac{\nu_1}{\nu_2} - 1| < \epsilon_0$ it holds that $|\operatorname{Var}(\eta) - 1| < \epsilon$. By Assumption 6 for this ϵ_0 and δ_0 there exits a ρ_0, such that for all $\rho > \rho_0$,

$$\inf_{i \in \mathbb{N}} \lambda_i(\rho) > \delta_0, \quad \sup_{i,j \in \mathbb{N}} |\frac{\lambda_i(\rho)}{\lambda_j(\rho)} - 1| < \epsilon_0. \tag{16}$$

Hence for all $\rho > \rho_0$ we have that $|\operatorname{Var}(\eta) - 1| < \epsilon$ and so (14) is proved.

For the second result, we start with a simple observation on the considered covariances. Using Theorem 3.2 together with (3) we have that for $i \neq k$, $\operatorname{Cov}(\mathcal{G}_i(\boldsymbol{\xi}_N), \mathcal{G}_k(\boldsymbol{\xi}_N)) \to 0$.

Next, the observation $\boldsymbol{\xi}_N$ is split up into parts of fixed length 2^{J_0}, which we denote by $\mathbf{u}_n$, s.t. we have $\boldsymbol{\xi}_N^\top = (\mathbf{u}_1^\top, \ldots, \mathbf{u}_{\tilde{J}}^\top)$ with $\tilde{J} = 2^J/2^{J_0}$. Of course, all $\mathbf{u}_n$ are independent as are the components of $\boldsymbol{\xi}_N$. For fixed $1 \leq i \neq k \leq (2^{J_0} - 1)$ it follows from the remark above that $\forall \epsilon > 0$, there exists $\rho_{i,k}$, such that $\forall \rho > \rho_{i,k}$,

$$\sup_n \operatorname{Cov}(\mathcal{G}_i(\mathbf{u}_n), \mathcal{G}_k(\mathbf{u}_n)) < \epsilon. \tag{17}$$

Set $\rho_0 = \max_{1 \leq i \neq k < 2^{J_0}} \rho_{i,k}$. Clearly, for any $\rho > \rho_0$, (17) holds. Now consider arbitrary $J \geq J_0$. As is clear from the definition, $\mathcal{G}_{j,l}(\boldsymbol{\xi}_N)$ is equal to $\mathcal{G}_{\tilde{j},\tilde{l}}(\mathbf{u}_{\tilde{n}})$ with appropriate $\tilde{j}, \tilde{l}, \tilde{n}$. Combining $\operatorname{Cov}(\mathcal{G}_i(\mathbf{u}_n), \mathcal{G}_k(\mathbf{u}_m)) = 0$ for $n \neq m$ because of independence with (17) we obtain

$$\sup_{J \geq J_0,\, 1 \leq i \neq k \leq 2^J(1-2^{J_0})} |\operatorname{Cov}(\mathcal{G}_i(\boldsymbol{\xi}_{2^J}), \mathcal{G}_k(\boldsymbol{\xi}_{2^J}))| < \epsilon.$$

and therefore the proof is finished. □ ■

Remark 6.4. It will also be useful to have a result on the second moments instead of the variances. Note that with $\operatorname{Var}(\eta) \to 1$ we have also $\mathbb{E}(\eta^2) \to 1$, compare (3). A analogous argument to the one used in Theorem 6.3 then yields that for any $\epsilon > 0$ there exists ρ_0, such that thereafter

$$\sup_{2^J, 1 \leq i < 2^J} |\mathbb{E}((\mathcal{G}_i(\boldsymbol{\xi}_{2^J})^2) - 1| < \epsilon. \tag{18}$$

Acknowledgement

We thank an anonymous referee for his suggestions which helped to improve the exposition of the paper substantially.

References

Dassios, A. and J. Jang (2003). Pricing of catastrophe reinsurance & derivatives using the Cox process with shot noise intensity. *Finance and Stochastics 7*(1), 73 – 95.

Dassios, A. and J.-W. Jang (2005). Kalman-Bucy filtering for linear systems driven by the Cox process with shot noise intensity and its application to the pricing of reinsurance contracts. *Journal of Applied Probability 42*(1), 93–107.

Fisz, M. (1955). The limiting distribution of a function of two independent random variables and its statistical application. *Cloolquium Mathematicum 3*, 138–146.

Fryzlewicz, P. (2007). Data-driven wavelet-Fisz methodology for nonparametric function estimation. *submitted*.

Fryzlewicz, P. and G. P. Nason (2001). Poisson intensity estimation using wavelets and the fisz transformation. citeseer.ist.psu.edu/608155.html.

Fryzlewicz, P. and G. P. Nason (2004). A Haar-Fisz algorithm for Poisson intensity estimation. *Journal of Computational and Graphical Statistics 13*(3), 621 – 638.

Fryzlewicz, P. and G. P. Nason (2006). Haar-Fisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society Series B 68*, 611 – 634.

Gaspar, R. M. and T. Schmidt (2005). Quadratic portfolio credit risk models with shot-noise effects. *Stockholm School of Economics Working Paper Series - N0. 616*.

Jansen, M. (2006). Multiscale poisson data smoothing. *Journal of the Royal Statistical Society, Series B 68*(1), 17 – 48.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing* (2nd ed.). Academic Press.

Nason, G. P. (2002). Choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage. *Statistics and Computing 12*, 219–227.

Schmidt, T. and W. Stute (2007). General shot-noise processes and the minimal martingale measure. *Statistics & Probability Letters 77*, 1332–1338.