Supplementary materials to "Adaptive efficient estimation for generalized semi-Markov Big Data models"

Vlad Stefan Barbu [1] **- Slim Beltaief** [2] **- Serguei Pergamenchtchikov** [3]

Monte-Carlo simulations

In this section we give the results of Monte-Carlo experiments for the model selection procedure (39). In (3) we chose a 1-periodic function which, for $0 \le t \le 1$, is defined as

$$S(t) = \begin{cases} |t-0,5| & \text{if} \quad 0,25 \le t \le 0,75\,, \\ 0,25\,, & \text{elsewhere.} \end{cases} \tag{1}$$

We simulate the process (3) in which $\xi_t = 0.5 w_t + 0.5 z_t$ and z_t is the semi-Markov process defined in (8) with i.i.d. Gaussian $\mathcal{N}(0,1)$ sequence $(\zeta_j)_{j\ge1}$ and $(\tau_k)_{k\ge1}$ used in (10) are taken as $\tau_k \sim \chi_3^2$. For this case we use the model selection procedure (39) constructed through the weight vectors (48) with the following parameters: $k^* = 100 + \sqrt{\ln T}$, $r_i = i/\ln T$, $m = [\ln^2 T]$ and $\delta = (3+\ln(n))^{-2}$. We define the empirical risk as

$$\overline{\mathcal{R}} = \frac{1}{p}\sum_{j=1}^{p} \widehat{\mathbf{E}}\left(\widehat{S}_T(t_j) - S(t_j)\right)^2 \quad \text{and} \quad t_j = \frac{j}{p}\,. \tag{2}$$

Here the observation frequency was taken $p = 100001$ and the expectation was taken as an average over $M = 10000$ replications, i.e.

$$\widehat{\mathbf{E}}\left(\widehat{S}_T(.) - S(.)\right)^2 = \frac{1}{M}\sum_{l=1}^{M}\left(\widehat{S}_T^l(\cdot) - S(\cdot)\right)^2\,.$$

We use also the relative risk

$$\overline{\mathcal{R}}_* = \frac{\overline{\mathcal{R}}}{\|S\|_p^2} \quad \text{and} \quad \|S\|_p^2 = \frac{1}{p}\sum_{j=1}^{p} S^2(t_j)\,. \tag{3}$$

In our case $\|S\|_p^2 = 0.1883601$. Table 1 below gives the values for the sample risks (2) and (3) for different numbers of observations T.

[1] Laboratoire de Mathématiques Raphaël Salem, UMR 6085 CNRS-Université de Rouen Normandie, France, E-mail: vladstefan.barbu@univ-rouen.fr

[2] ALTEN de Toulouse, 9 Rue Alain Fournier, 31300 Toulouse France, E-mail: beltaiefslim@hotmail.fr

[3] Laboratoire de Mathématiques Raphaël Salem, UMR 6085 CNRS-Université de Rouen Normandie, Avenue de l'Université, BP.12, 76801 Saint-Etienne-du-Rouvray, France and International Laboratory of Statistics of Stochastic Processes and Quantitative Finance, Tomsk State University, Tomsk, Russia, E-mail: Serge.Pergamenshchikov@univ-rouen.fr

T	$\overline{\mathcal{R}}$	$\overline{\mathcal{R}}_*$
20	0.0398	0.211
100	0.0091	0.0483
200	0.0067	0.0355
1000	0.0022	0.0116

Table 1 Empirical risks

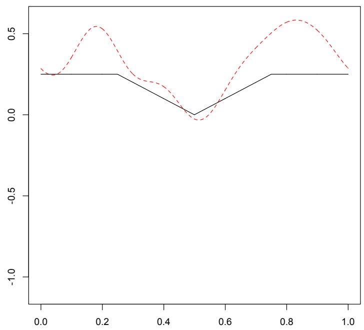

Fig. 1 Estimator of S for $T = 20$

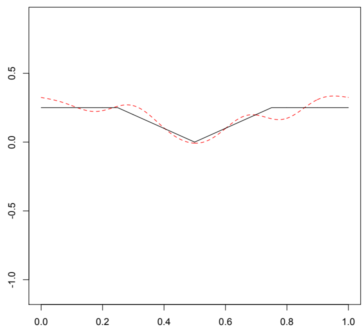

Fig. 2 Estimator of S for $T = 100$

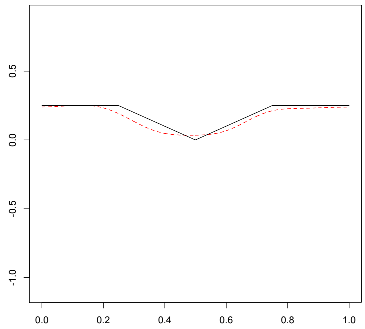

Fig. 3 Estimator of S for $T = 200$

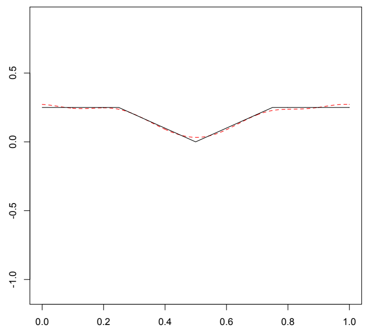

Fig. 4 Estimator of S for $T = 1000$

Figures 1–4 show the behaviour of the regression function and of its estimates obtained by the model selection procedure (39), for several values of observation periods T. The black full line is the regression function (1) and the red dotted line is the associated estimator.

Remark 1 From numerical simulations of the procedure (39) with various observation durations T we may conclude that the quality of the proposed procedure is good for practical needs, i.e. for reasonable (non large) number of observations. We can also add that the quality of the estimation improves as the number of observations increases.