

# Answer to the Referee Report I

- A consistency result of Proposition 3.3.

I believe that the following type of result can be obtained:  $\Pr(E_T^l) \rightarrow 1$ . Theorem 3.1 is for testing purpose. In certain application one might be interested in such consistency result. Basically one needs to study the behavior of  $q_T(\alpha)$  when  $\alpha \rightarrow 0$ .

- Estimation of long run variance using autoregressive processes.

The authors considered estimating  $\sigma^2$  using AR processes. A limitation is that the order  $p$  is fixed and finite. It appears that the latter limitation can be relaxed. For a stationary process  $\varepsilon_t$  (not necessarily linear), one can fit an AR process with large  $p$

$$\varepsilon_t = \sum_{j=1}^p a_j \varepsilon_{t-j} + \eta_t,$$

properties of fitted  $\hat{a}_1, \dots, \hat{a}_p$  can be obtained from the results in the following papers

- W. B. Wu and Mohsen Pourahmadi (2009): Banding Sample Covariance Matrices of Stationary Processes, *Statistica Sinica* 19 1755-1768.
- H. Xiao and W. B. Wu (2012). Covariance Matrix Estimation for Stationary Time Series. *Annals of Statistics*, Volume 40, Number 1 (2012), 466-493.

A similar version of the authors estimate (4.14) can be used. Rate of convergence (cf. Proposition 4.1) can be derived with rate  $T^{-1/2}$  therein possibly replaced by a larger term of the form  $T^{-c}$  with  $c < 1/2$ .

- Real data application.

The authors analyzed the yearly mean Central England temperature data. It will be interesting to apply their approach to the global temperature data. In the paper “Isotonic regression: another look at the change point problem. *Biometrika*, 88, 793-804, 2001”, an increasing trend function is fitted. It will be important to know which period the sequence is increasing/decreasing.

- Simulation Study

In the simulation study the authors considered AR(1) processes with relatively weaker dependence:  $a_1 \in \{-0.5, -0.25, 0.25, 0.5\}$ . One should consider the stronger positive/negative dependence case with  $a = \pm 0.9$  (say). How does the strength of dependence affect the performance of the procedure?

# Answer to the Referee Report II

1. Section 3.2: The authors recommend computing the quantiles for the independent Gaussian case by simulation. This suggestion is already in the original SiZer paper (Chaudhuri and Marron, 1999). However in the late 1990s computing power was not sufficient to make this suggestion feasible. This led to the use of approximation such as in Hannig and Marron (2006). I would like to ask how does the simulation based quantile compare to the approximation in Hannig and Marron (2006).
  - Compare multiscale test with and without additive correction terms  $\lambda(h)$  by “parallel coordinate plots” as in Figure 5 of Hannig & Marron (2006).
2. Page 12, line 15: What is random here? After a spending some time I believe that it is the  $\Pi_T$  but on first reading I thought  $E_T$ s are non-random. Please explain these various objects better.
3. Page 18, line 52: Please remove the speculative statements about what can be shown unless you actually show it in this paper.
4. Section 4.1: This section does not contain any truly new material and should be removed.
5. Section 5.2: I understand that you are doing comparisons to SiZer out of the box. However, some of the comparison might not be quite fair. SiZer is adjusting multiplicity “row-wise” while the proposed method is attempting a global multiple control. What would happen if your  $G_T$  only focused on one scale?

What’s done (here  $\alpha$  is always taken to be 0.05):

- For each setting, each sample size, and each bandwidth row we report the percentage of realizations of the data in which there were some red or blue pixels in that row. The reports are given in “parallel coordinate plots” as in Figure 5 of Hannig and Marron (2006). Each plot has the results of SiZer, the results of our global method and the results of our rowwise method respectively. The settings are as follows:
  - (a) Under the null for  $a_1 = -0.25, 0, 0.25$  and sample size  $T = 250$ .
  - (b) Under the alternative for  $a_1 = -0.25, 0$  and sample size  $T = 250$ . The trend function is linear increasing on the whole interval with the slope  $\beta = 1.25$ .
  - (c) Under the alternative for  $a_1 = 0.25$  and sample size  $T = 250$ . The trend function is linear increasing on the whole interval with the slope  $\beta = 2.25$ .

- Representative SiZer maps (for SiZer, our global method and our rowwise method) for simulated data from the null distribution for  $a_1 = -0.25, 0, 0.25$ . Sample size is  $T = 250$ . 1,000 such SiZer maps were simulated for each setting and the population of them was ordered in terms of number of pixels that flag significant structure by being red or blue. For each setting and for each method we show 500th of these maps (essentially the median of the population), the 750th (the third quartile), the 850th and the 950th member of the ordered population.
6. Page 25, line 1-26: I do not quite understand this figure. Would it be possible to rather reproduce the colorful SiZer figures that show the results of the test at various scales and locations? Also you should use several different signals. I believe that a single relatively large bump is not sufficient test bed. A good collection of signals can be found in Donoho and Johnstone (1995). Also, would Hannig et al. (2013) be helpful in comparing the results?

What's done (here  $\alpha$  is always taken to be 0.05):

- Plots comparing SiZer maps under the null:
  - (a) Plots comparing SiZer and global method and rowwise methods respectively for sample size is  $T = 250$ . AR(1) parameters are  $a_1 = -0.25, 0, 0.25$  and  $\sigma_\eta^2 = 1$ .
- Plots comparing SiZer maps under the alternative:
  - (a) Plots comparing SiZer and global method and rowwise methods respectively for sine curve plus AR(1) noise (as in Park, Hannig and Kang (2009)) for sample size  $T = 250$ . AR(1) parameters are  $a_1 = -0.25, 0, 0.25$  and  $\sigma_\eta^2 = 1$ .
  - (b) Plots comparing SiZer and global method and rowwise methods respectively for blocks plus AR(1) noise (signal from Donoho, Johnstone, Kerkycharian and Picard (1995), recentered and normalized as in Hannig and Marron (2006)) for sample size  $T = 250$ . AR(1) parameters are  $a_1 = -0.25, 0, 0.25$  and  $\sigma_\eta^2 = 0.01$ .
  - (c) Plots comparing SiZer and global method and rowwise methods respectively for blocks plus AR(1) noise (signal from Donoho, Johnstone, Kerkycharian and Picard (1995), recentered and normalized as in Hannig and Marron (2006)) for sample size  $T = 1000$ . AR(1) parameters are  $a_1 = -0.25, 0, 0.25$  and  $\sigma_\eta^2 = 0.01$ .

Different signals:

- Power results could be reported in terms of ability to find the jump points in the blocks example.

7. Page 31, line 1-39: Can you plot the SiZer results on this data?

## References

- DONOHO, D., JOHNSTONE, I., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society: Series B*, **57** 301–369.
- HANNIG, J. and MARRON, J. S. (2006). Advanced distribution theory for SiZer. *Journal of the American Statistical Association*, **101** 484–499.
- PARK, C., HANNIG, J. and KANG, K.-H. (2009). Improved SiZer for time series. *Statistica Sinica*, **19** 1511–1530.