SUPPLEMENT TO "TAIL-GREEDY BOTTOM-UP DATA DECOMPOSITIONS AND FAST MULTIPLE CHANGE-POINT DETECTION"

By Piotr Fryzlewicz*

London School of Economics

1. Extending the scope of TGUH methodology to dependent non-Gaussian data. The purpose of this section is to construct a TGUH estimator of change-points $\eta_1, \ldots, \eta_N$ in a setting in which the noise ε_t is possibly dependent and/or non-Gaussian (including the case in which its tails are heavier than those of the Gaussian distribution), in a sense specified more precisely in the remainder of this section. In order to perform this extension without increasing the rate of the threshold used in the iid Gaussian model (that is, $O(\log^{1/2}(T))$, as in Theorems 3.1–3.3 of the main article), the estimators $\tilde{f}$, $\tilde{\tilde{f}}$ and $\hat{f}$ need to be altered. The first step is to transform the initial estimator $\tilde{f}$ into a new initial estimator $\tilde{f}^r$ by rearranging the Unbalanced Haar basis on which $\tilde{f}$ is built into a new UH basis (the "r" in f^r stands for "rearranged"). Informally speaking, this is done in order for us to be able to only rely on the behaviour of those sums $\sum_{i=t_1}^{t_2} \varepsilon_i$ for which $t_2 - t_1$ is "large", in the analysis of the new estimator $\tilde{f}^r$. This is of importance as it enables the application of certain strong asymptotic normality arguments to the terms $\sum_{i=t_1}^{t_2} \varepsilon_i$, which in turn enables the use of Gaussian-magnitude thresholds $O(\log^{1/2}(T))$ in the construction of $\tilde{f}^r$ and its derivative estimators $\tilde{\tilde{f}}^r$ and $\hat{f}^r$, the latter two being the rearranged-basis counterparts of $\tilde{\tilde{f}}$ and $\hat{f}$, respectively. The ability to use thresholds of magnitude $O(\log^{1/2}(T))$ even for dependent non-Gaussian data is important as this is the lowest permitted threshold rate, and the use of higher-rate thresholds has a detrimental impact of the covergence rates of the implied estimators, as well as (frequently) degrading the practical performance of the method. We start by explaining the mechanism of basis rearrangement in Section 1.1.

1.1. *Basis rearrangement.* For the TGUH basis chosen from the data in the construction of $\tilde{f}$, we separate the coefficients $d_{p,q,r}^{(j,k)}$ into "short-winged"

*Work supported by the Engineering and Physical Sciences Research Council grant no. EP/L014246/1.

and "long-winged" ones. Informally, long-winged coefficients will be defined as those for which both $r - q$ and $q - p$ are large. We will want to re-arrange the TGUH basis chosen so that any short-winged coefficients appear only at fine scales. This is because we will be setting them to zero in the new estimator $\tilde{f}^r$, which will create the least damage to the quality of the estimator if they appear at fine scales. This will be done in order to enable us to rely exclusively on those sums $\sum_{i=t_1}^{t_2} \varepsilon_i$ for which $t_2 - t_1$ is "large" in Theorem 1.1 below, which is the analogue of Theorem 3.1 of the main article for the rearranged-basis estimator $\tilde{f}^r$. This is important as certain strong asymptotic normality arguments will then apply, and we will be able to use thresholds of the same magnitude as for iid Gaussian data, which would be difficult without the basis rearrangement.

We define the sets of location indices for short-winged and long-winged coefficients (respectively) at each scale j as follows. The parameter a will be specified later and will be a feature of the estimator $\tilde{f}^r$; we will therefore sometimes write $\tilde{f}^r(a)$ to emphasise the dependence of $\tilde{f}^r$ on a.

$$\begin{aligned}
\mathcal{W}_j^0(a) &= \{1 \le k \le K(j) \ : \ d_{p,q,r}^{(j,k)} \text{ is such that } r - q \le a \text{ or } q - p + 1 \le a\}, \\
\mathcal{W}_j^1(a) &= \{1, \ldots, K(j)\} \setminus \mathcal{W}_j^0(a).
\end{aligned}$$

The first step in the basis re-arrangement, which we will now describe, is to perform standard hard (unconnected) thresholding of each coefficient $d_{p,q,r}^{(j,k)}$ individually, using threshold $\lambda_.$. The threshold $\lambda_.$ is to be specified later and will be such that if $d_{p,q,r}^{(j,k)}$ is long-winged (with parameter a to be specified later) and survives the thresholding, then the interval $[p, r]$ is certain to contain a change-point, on a set with a large probability.

We now explain our TGUH basis re-arrangement. Let $d_{p,q,r}^{(j,k)}$ be any coefficient such that $k \in \mathcal{W}_j^0(a)$, such that $\exists\ (j', k') \in \mathcal{C}_{j,k}$ satisfying $k' \in \mathcal{W}_{j'}^1(a)$ and $|d_{p',q',r'}^{(j',k')}| > \lambda_.$, and such that no parent coefficient of $d_{p,q,r}^{(j,k)}$ satisfies these conditions. If there are no such coefficients $d_{p,q,r}^{(j,k)}$, no basis re-arrangement is needed. The basis re-arrangement proceeds as follows.

1. Remove $d_{p,q,r}^{(j,k)}$ from the set of TGUH coefficients, and replace it with $d_{p,q',r}^{(j,k)}$. Because $(j', k') \in \mathcal{C}_{j,k}$, we must have $q' - p \ge q' - p'$ and $r - q' \ge r' - q'$, and therefore after the replacement, $k \in \mathcal{W}_j^1(a)$.
2. Remove $d_{p',q',r'}^{(j',k')}$ from the set of TGUH coefficients, and replace it with $d_{p,q,q'}$ if $q' > q$, or with $d_{q',q,r}$ otherwise. Note that the new coefficient becomes a direct child of $d_{p,q',r}^{(j,k)}$, hence the scale of the new coefficient will be $j - 1$.

3. For $i = j-1, \ldots, j'+1$ (if this range is non-empty), take each coefficient $d^{(i,k_i)}_{p_i,q_i,r_i}$ that is both a child of the old coefficient $d^{(j,k)}_{p,q,r}$ and a parent of the old coefficient $d^{(j',k')}_{p',q',r'}$, and modify p_i and/or r_i so that both new coefficients: $d^{(j,k)}_{p,q',r}$ and $d_{p,q,q'}$ if $q' > q$ or $d_{q',q,r}$ otherwise, as well as any coefficients already modified in the current loop over i, become its parents. There is a unique way in which this can be done. The parameter q_i should remain unchanged. Note that the scale of each thus-modified coefficient changes from i to $i-1$.

Some comments are in order.

(i) After the procedure described above, the short-winged coefficient $d^{(j,k)}_{p,q,r}$ gets transformed into a long-winged coefficient $d^{(j,k)}_{p,q',r}$. We then proceed iteratively: perform standard hard (unconnected) thresholding of each coefficient $d^{(j,k)}_{p,q,r}$ individually, using threshold λ., and then again steps (a)–(c) above, until there are no more short-winged coefficients with threshold-exceeding long-winged children.

(ii) Importantly, basis re-arrangement does not increase the number of coefficients $d^{(j,k)}_{p,q,r}$ for which $[p, r]$ contains a change-point. To see this, note that all the coefficients being modified in steps (a)–(c) overlap with a change-point prior to their modification. This is because $|d^{(j',k')}_{p',q',r'}|$ overlaps with a change-point, since $|d^{(j',k')}_{p',q',r'}| > \lambda$..

(iii) After steps (a)–(c), by construction, the UH vectors corresponding to the modified coefficients are mutually orthonormal.

(iv) After steps (a)–(c), the partition of the interval $[p, r]$ induced by the parameters q of the modified coefficients is the same as before steps (a)–(c), since the set of the parameters q does not change. Therefore, the UH vectors corresponding to the modified coefficients are orthonormal to those corresponding to all other coefficients that have not been modified.

(v) As a result of the basis re-arrangement, there are no short-winged coefficients with threshold-exceeding long-winged children. Therefore, if a coefficient has a threshold-exceeding long-winged child, it must itself be long-winged.

1.2. *Theoretical behaviour of rearranged-basis estimators.* We start by describing the mean-square behaviour of the rearranged-basis estimator $\tilde{f}^r$.

THEOREM 1.1. *Let the distribution of ε_t in model (1) of the main article be such that* *(a)* ε_t *satisfies Cramer's conditions, that is*

$$E|\varepsilon_t|^k \leq c^{k-2} k! E(\varepsilon_t^2) < \infty, \quad t = 1, \ldots, T, \quad k = 3, 4, \ldots,$$

where c *is a certain positive constant;*

(b) the stationary sequence $\{\varepsilon_t\}_t$ *is* m*-dependent, that is the dependence between the variables* $(\ldots, \varepsilon_{t-1}, \varepsilon_t)$ *and* $(\varepsilon_{t+m+1}, \varepsilon_{t+m+2}, \ldots)$ *vanishes for each* t*.*

Further, let $\overline{f'} = \max_t f_t - \min_t f_t$ *be bounded in* T*. Let the estimator* $\tilde{f}^r(a)$ *use the UH basis rearranged as described in Section 1.1 with* $a = C_1 \log(T)$ *for a certain large enough constant* C_1*, and let it estimate each* $\mu^{(j,k)}$ *for* $j \geq 1$ *via*

$$\hat{\mu}^{(j,k)} = d^{(j,k)} \mathbb{I}\{\exists\ (j', k') \in \mathcal{C}_{j,k}\ \ |d^{(j',k')}| > \lambda. \ \wedge \ k' \in \mathcal{W}^1_{j'}(a)\}$$

(that is, using the same connected thresholding as that used by $\tilde{f}$*, plus in addition setting all short-winged coefficients* $d^{(j,k)}$ *to zero). Let the threshold* $\lambda.$ *satisfy* $\lambda. = C \log^{1/2}(T)$*, for a large enough* C*. On the set* $\mathcal{A}^r_T$*, defined by*

$$\mathcal{A}^r_T = \left\{ \forall\ 1 \leq t_1 \leq t_2 \leq T \quad s.t. \quad t_2 - t_1 \geq C_1 \log(T) \quad (t_2 - t_1 + 1)^{-1/2} \left| \sum_{t=t_1}^{t_2} \varepsilon_t \right| \leq C \log^{1/2}(T) \right\},$$

which satisfies $P(\mathcal{A}^r_T) \geq 1 - C_3/T$ *for a certain constant* C_3*, we have*

$$\|\tilde{f}^r - f\|^2_T \leq \tilde{C} T^{-1} N \lceil \log(T) / \log\{(1-\rho)^{-1}\} \rceil \log(T),$$

for a certain constant $\tilde{C}$*.*

Proof. We first examine the behaviour of the set $\mathcal{A}^r_T$. Since ε_t is m-dependent, its α-mixing coefficients $\alpha(l)$ must be such that $\alpha(l) = 0$ for $l > m$. We now consider the single sum $a^{-1/2} \sum_{t=1}^{a} \varepsilon_t$. All below applies also to a sum in which we start or end at indices other than 1 and a, provided the difference between them is at least $a - 1$. From Theorem 1.4 in Bosq (1998), we obtain

$$P\left(\left| \sum_{t=1}^{a} \varepsilon_t \right| > a^{1/2} \lambda \right) \leq a_1 \exp\left(- \frac{q \frac{\lambda^2}{a}}{25 m_2^2 + 5c \frac{\lambda}{\sqrt{a}}} \right),$$

where

$$\begin{aligned} a_1 &= 2aq^{-1} + 2 \left(1 + \frac{\frac{\lambda^2}{a}}{25 m_2^2 + 5c \frac{\lambda}{\sqrt{a}}} \right), \\ m_2^2 &= \max_{1 \leq t \leq a} E(\varepsilon_t^2), \end{aligned}$$

provided that $[a/(q+1)] \geq m+1$, which is possible to achieve by setting $q = [c_1 a]$ with c_1 suitably small. With this, setting $\lambda = C\log^{1/2}(T)$, and $a = C_1 \log(T)$ for certain C, C_1, we have that a_1 is bounded by a constant and

$$\exp\left(-\frac{q\frac{\lambda^2}{a}}{25m_2^2 + 5c\frac{\lambda}{\sqrt{a}}}\right) \leq \exp(-C_2\log(T)) = T^{-C_2},$$

where C_2 is a positive constant that can be set arbitrarily large. Since the number of sums of the form $\sum_{t=t_1}^{t_2} \varepsilon_t$ does not exceed T^2, applying the Bonferroni correction, we have the bound

$$P\left(\forall\ 1 \leq t_1 \leq t_2 \leq T \quad \text{s.t.} \quad t_2 - t_1 \geq C_1\log(T) \quad (t_2-t_1+1)^{-1/2}\left|\sum_{t=t_1}^{t_2}\varepsilon_t\right| \leq C\log^{1/2}(T)\right) \geq 1 - \frac{C_3}{T}$$

as $T \to \infty$ for certain large enough C, C_1, and a certain C_3. We now turn to the behaviour of $\tilde{f}^r$. Note that J below may well be different from that in Theorem 3.1 of the main article, but this will not interfere with our analysis as the number of coefficients overlapping with change-points has not increased, as argued in item (ii) of Section 1.1. We have

$$\begin{aligned}
\|\tilde{f} - f\|_T^2 &= T^{-1}\sum_{j=1}^{J}\sum_{k=1}^{K(j)}(d^{(j,k)}\mathbb{I}\{\exists\ (j',k') \in \mathcal{C}_{j,k}\ \ |d^{(j',k')}| > \lambda.\ \wedge\ k' \in \mathcal{W}_{j'}^1(a)\} - \mu^{(j,k)})^2 \\
&+ T^{-1}(s_{1,T} - \mu^{(0,0)})^2 \\
&= T^{-1}\sum_{j=1}^{J}\left(\sum_{k\in\mathcal{S}_j^0} + \sum_{k\in\mathcal{S}_j^1\cap\mathcal{W}_j^0(a)} + \sum_{k\in\mathcal{S}_j^1\cap\mathcal{W}_j^1(a)}\right) \\
&\quad (d^{(j,k)}\mathbb{I}\{\exists\ (j',k') \in \mathcal{C}_{j,k}\ \ |d^{(j',k')}| > \lambda.\ \wedge\ k' \in \mathcal{W}_{j'}^1(a)\} - \mu^{(j,k)})^2 \\
&+ T^{-1}(s_{1,T} - \mu^{(0,0)})^2 =: I + II + III + IV.
\end{aligned}$$

Term I. Since $k \in \mathcal{S}_j^0$, then on set $\mathcal{A}_T^r$, $\forall\ (j',k') \in \mathcal{C}_{j,k}$ if $k' \in \mathcal{W}_{j'}^1(a)$, then $|d^{(j',k')}| \leq \lambda.$. Also, $\mu^{(j,k)} = 0$ since $k \in \mathcal{S}_j^0$. Therefore $I = 0$.

Term II. If it were true that $\exists\ (j',k') \in \mathcal{C}_{j,k}\ \ |d^{(j',k')}| > \lambda.\ \wedge\ k' \in \mathcal{W}_{j'}^1(a)$, then by item (v) of Section 1.1, we would have to have $k \in \mathcal{W}_j^1(a)$, which contradicted the $k \in \mathcal{W}_j^0(a)$ in the definition of term II. Therefore, term II

simplifies to

$$T^{-1}\sum_{j=1}^{J}\sum_{k\in\mathcal{S}_j^1\cap\mathcal{W}_j^0(a)}(\mu^{(j,k)})^2.$$

We now bound the individual term $(\mu^{(j,k)})^2$. From formula (3) in Fryzlewicz (2014), we have

$$(\mu^{(j,k)})^2 \le \frac{(q-p+1)(r-q)}{r-p+1}(\overline{f'})^2. \tag{1}$$

W.l.o.g., let us assume that $q-p+1 \le a$. Denote $\omega_1 = q-p+1$, $\omega_2 = r-q$. Noting that $\omega_1\omega_2/(\omega_1+\omega_2)$ is an increasing function of ω_1, the right-hand side of (1) is futher bounded from above by

$$\frac{a\omega_2}{a+\omega_2}(\overline{f'})^2 \le a(\overline{f'})^2.$$

How many such terms $(\mu^{(j,k)})^2$ are included in term II? No more than the overall number of coefficients overlapping with change-points, which has not increased compared to the setting of Theorem 3.1, by remark (ii) of Section 1.1. From the proof of Theorem 3.1, this number is bounded from above by $N\lceil\log(T)/\log\{(1-\rho)^{-1}\}\rceil$. Therefore, we obtain

$$II \le T^{-1}N\lceil\log(T)/\log\{(1-\rho)^{-1}\}\rceil a(\overline{f'})^2.$$

Recalling that $(\overline{f'})^2 = \text{const}$ and $a = O(\log(T))$, we have

$$II \le CT^{-1}N\lceil\log(T)/\log\{(1-\rho)^{-1}\}\rceil\log(T).$$

Term III. Denote $\mathcal{B} = \{\exists\ (j',k') \in \mathcal{C}_{j,k}\ \ |d^{(j',k')}| > \lambda_.\ \wedge\ k' \in \mathcal{W}_{j'}^1(a)\}$ and compute

$$\begin{aligned}(d^{(j,k)}\mathbb{I}\{\mathcal{B}\}-\mu^{(j,k)})^2 &= (d^{(j,k)}\mathbb{I}\{\mathcal{B}\}-d^{(j,k)}+d^{(j,k)}-\mu^{(j,k)})^2\\ &\le 2(d^{(j,k)})^2\mathbb{I}(|d^{(j,k)}|\le\lambda_.\vee k\in\mathcal{W}_j^0(a))+2(d^{(j,k)}-\mu^{(j,k)})^2\\ &= 2(d^{(j,k)})^2\mathbb{I}(|d^{(j,k)}|\le\lambda_.)+2(d^{(j,k)}-\mu^{(j,k)})^2\\ &\le 2\lambda_.^2+\bar{C}\log(T),\end{aligned}$$

where the $\bar{C}\log(T)$ bound comes from the definition of $\mathcal{A}_T^r$. Recalling that $\lambda_.^2 = O(\log(T))$, we have

$$III \le CT^{-1}N\lceil\log(T)/\log\{(1-\rho)^{-1}\}\rceil\log(T).$$

Finally, noting that $IV = o(II, III)$, we obtain

$$\|\tilde{f} - f\|_T^2 \leq \tilde{C} T^{-1} N \lceil \log(T) / \log\{(1-\rho)^{-1}\} \rceil \log(T),$$

which completes the proof.

We now construct the intermediate estimator $\tilde{\tilde{f}}^r$ from $\tilde{f}^r$ in the same way as the estimator $\tilde{\tilde{f}}$ was constructed from $\tilde{f}$. To be more precise, we follow the process described in Stage 1 of the post-processing of Section 2.4 of the main article with $\tilde{f}^r$ on input and with threshold λ. of Theorem 1.1. The properties of the intermediate estimator $\tilde{\tilde{f}}^r$ are described in the following theorem.

THEOREM 1.2. *Let the distribution of ε_t in model (1) of the main article be as in Theorem 1.1. Further, let $\overline{f'} = \max_t f_t - \min_t f_t$ be bounded in T. Let the estimator $\tilde{\tilde{f}}^r$ be constructed from $\tilde{f}^r$ of Theorem 1.1 via Stage 1 of the post-processing of Section 2.4 of the main article, with threshold λ. of Theorem 1.1. On the set $\mathcal{A}_T^r$ (defined in Theorem 1.1), which satisfies $P(\mathcal{A}_T^r) \geq 1 - C_3/T$ for a certain constant C_3, we have*

$$\|\tilde{\tilde{f}}^r - f\|_T^2 = O(NT^{-1} \log^2(T)).$$

Further, with the convention $\eta_0 = 1$, $\eta_{N+1} = T + 1$, the piecewise-constant estimator $\tilde{\tilde{f}}^r$ contains at most two change-points between each pair (η_i, η_{i+1}) of change-points in f, for $i = 0, \ldots, N$. Therefore the number $\tilde{\tilde{N}}$ of change-points in $\tilde{\tilde{f}}^r$ satisfies $\tilde{\tilde{N}} \leq 2(N+1)$.

Proof. The only difference in comparison to the proof of Theorem 3.2 of the main article is that we can no longer claim that the coefficients inherited from $\tilde{B}$ (in the notation of the proof of Theorem 3.2 of the main article) live on no more than J scales. However, as argued in the proof of Theorem 1.1, any increases to the original number of scales do *not* increase the number of coefficients that overlap with change-points. Therefore, the mean-square bounds from Theorem 1.1 still apply and therefore we still have

$$\|\tilde{\tilde{f}} - f\|_T^2 = O(NT^{-1} \log^2(T)).$$

The rest of the proof proceeds exactly the same as the proof of Theorem 3.2 of the main article. This completes the proof.

We now construct our final estimator $\hat{f}^r$ from $\tilde{\tilde{f}}^r$ in the same way as the estimator $\hat{f}$ was constructed from $\tilde{\tilde{f}}$. To be more precise, we follow the process described in Stage 2 of the post-processing of Section 2.4 of the main article with $\tilde{\tilde{f}}^r$ on input and with threshold $\lambda.$ of Theorem 1.1. The properties of the final estimator $\hat{f}^r$ are described in the following theorem.

THEOREM 1.3. *Let the distribution of ε_t in model (1) of the main article be as in Theorem 1.1. Further, let $\overline{f'} = \max_t f_t - \min_t f_t$ be bounded in T. Let the estimator $\hat{f}^r$ be constructed from $\tilde{\tilde{f}}^r$ of Theorem 1.2 via Stage 2 of the post-processing of Section 2.4 of the main article, with threshold $\lambda.$ of Theorem 1.1. Denote the number of change-points in $\hat{f}$ by $\hat{N}$ and their locations, in increasing order, by $\hat{\eta}_1, \ldots, \hat{\eta}_{\hat{N}}$. Let the number N of change-points in f be finite. Denoting their locations, in increasing order, by $\eta_1, \ldots, \eta_N$, assume that $\min_{i=1,\ldots,N} |f_{\eta_i} - f_{\eta_i - 1}| \geq \underline{f} > 0$, and $\min_{i=1,\ldots,N+1} |\eta_i - \eta_{i-1}| \geq b_T$ where b_T is such that $\log^2(T) = o(b_T)$. Then, on the set $\mathcal{A}_T^r$ (defined in Theorem 1.1), which satisfies $P(\mathcal{A}_T^r) \geq 1 - C_3/T$ for a certain constant C_3, and for T large enough, we have*

$$\begin{aligned} \hat{N} &= N, \\ |\hat{\eta}_i - \eta_i| &\leq \tilde{C} \log^2(T) \quad \textit{for all } i = 1, \ldots, N, \end{aligned}$$

where $\tilde{C}$ is a constant.

Proof. The proof of Theorem 1.3 proceeds in exactly the same way as that of Theorem 3.3 of the main article.

We end this section with two further remarks.

1. In the case of ε_t being a stationary Gaussian process with absolutely summable autocovariance, it is possible to prove analogues of Theorems 3.1–3.3 of the main article for estimators without the basis rearrangement, involving thresholds of the same order $O(\log^{1/2}(T))$ as in the iid Gaussian case, but with different constants. We omit the details.
2. In the case of ε_t satisfying Cramer's conditions with geometric α-mixing, Theorem 1.4 of Bosq (1998) stipulates that thresholds of a higher order of magnitude than $O(\log^{1/2}(T))$ would have to be used, which would impact consistency rates in (the equivalents of) Theorems 1.1 – 1.3. We do not investigate this case in the current work.

2. Refinements to post-processing. In this section, we mention two other possible refinements to the post-processing methodology described in Section 2.4 of the main article.

(A) By Theorem 3.3 of the main article, in the notation of that theorem, we must have

$$1 - \frac{c\,b_T}{T} \geq \frac{\hat{\eta}_{i+1} - \hat{\eta}_i}{\hat{\eta}_i - \hat{\eta}_{i-1}} \geq \frac{c\,b_T}{T}$$

for all i and a certain constant c. Therefore, it may be beneficial for the practical performance of the final estimator to remove, via an iterative procedure analogous to those described in Stages 1 and 2 of Section 2.4 of the main article, any estimated change-points $\hat{\eta}_i$ for which

$$\frac{\hat{\eta}_{i+1} - \hat{\eta}_i}{\hat{\eta}_{i+1} - \hat{\eta}_{i-1}} < \beta \qquad \text{or} \qquad \frac{\hat{\eta}_{i+1} - \hat{\eta}_i}{\hat{\eta}_{i+1} - \hat{\eta}_{i-1}} > 1 - \beta$$

where $\beta \in (0, 1/2)$ is a small user-specified constant. If $\beta \leq \frac{c\,b_T}{T}$, this pruning will not affect the theoretical consistency result from Theorem 3.3 of the main article.

(B) As with any change-point estimator which is consistent for the number and location of change-points, the accuracy of the estimated change-point locations in $\hat{f}$ can be further improved by employing the location re-estimation procedure described in the final paragraph of Section 3.1 in Fryzlewicz (2014).

3. Accuracy of TGUH in estimating change-point locations. To further investigate the performance of the TGUH method in estimating the locations η_i of the true change-points in the signals tested, we plot in Figures 1 – 5 those estimates for which $\hat{N} = N$, for models (1), (2a), (3), (4a), (5a) from Section 4.2 of the main article. In an ideal estimation procedure, the locations of change-points in the estimates should align exactly with the change-points in the true signal. The result for `blocks` (model (1)) appears satisfactory, except a small number of estimates which contain spurious spikes. The relatively low rate of correct estimation of N for this signal (for all methods) comes from the difficulty, at this signal-to-noise ratio, of correctly estimating the first feature after $t = 1500$; Figure 1 shows that given that this feature has been picked up, the locations η_i are in a large part estimated correctly. The `fms` signal (model (2a)) is relatively straightforward to estimate. The `mix` signal (model (3)) is possibly the most challenging one to estimate among models (1), (2a), (3), (4a), (5a), and the number of spurious spikes in Figure 3 reflects this, although given the degree of difficulty, we are pleased with the result for TGUH. The result for `teeth10` (model

(4a)) is satisfactory, and `stairs10` (model (5a)) is a relatively easy signal to estimate.

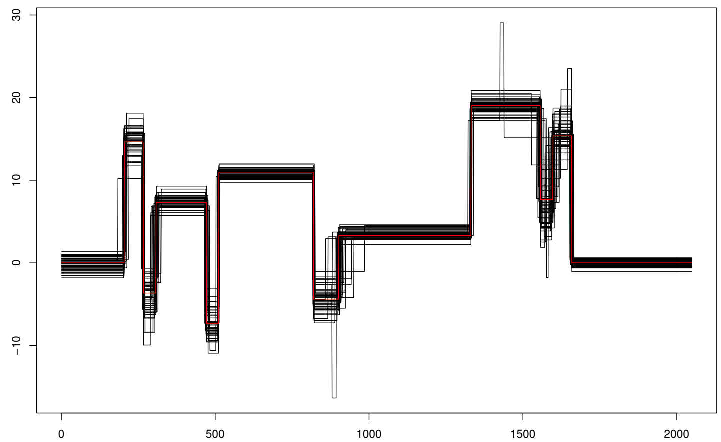

FIG 1. *Black: ensemble of those TGUH estimates of the* `blocks` *function (model (1)) from the simulation study described in Section 4.2 of the main article for which* $\hat{N} = N$. *Red: the true* `blocks` *function.*

4. Additional figure for the analysis of Section 4.3 of the main article. This is provided in Figure 6.

References.

BOSQ, D. (1998). *Nonparametric Statistics for Stochastic Processes*, 2 ed. Springer-Verlag, New York.

FRYZLEWICZ, P. (2014). Wild Binary Segmentation for multiple change-point detection. *Annals of Statistics* **42** 2243–2281.

DEPARTMENT OF STATISTICS
LONDON SCHOOL OF ECONOMICS
HOUGHTON STREET
LONDON WC2A 2AE
UK E-MAIL: p.fryzlewicz@lse.ac.uk

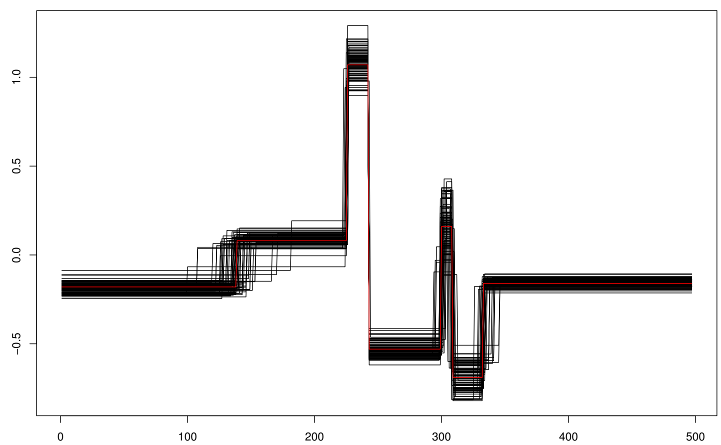

Fig 2. *Black: ensemble of those TGUH estimates of the* `fms` *function (model (2a)) from the simulation study described in Section 4.2 of the main article for which* $\hat{N} = N$. *Red: the true* `fms` *function.*

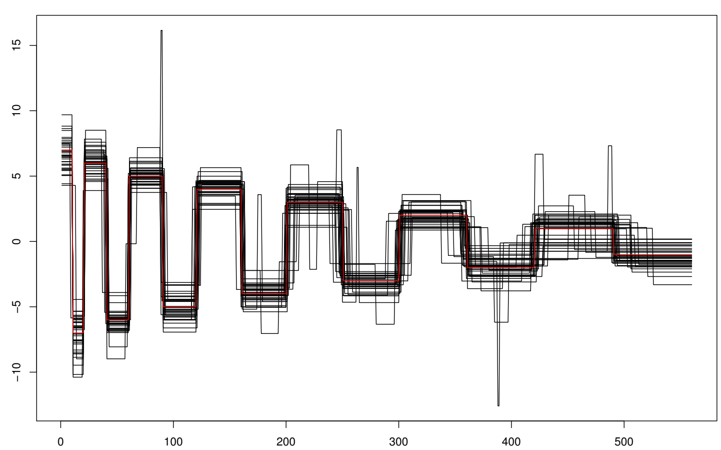

Fig 3. *Black: ensemble of those TGUH estimates of the* `mix` *function (model (3)) from the simulation study described in Section 4.2 of the main article for which* $\hat{N} = N$. *Red: the true* `mix` *function.*

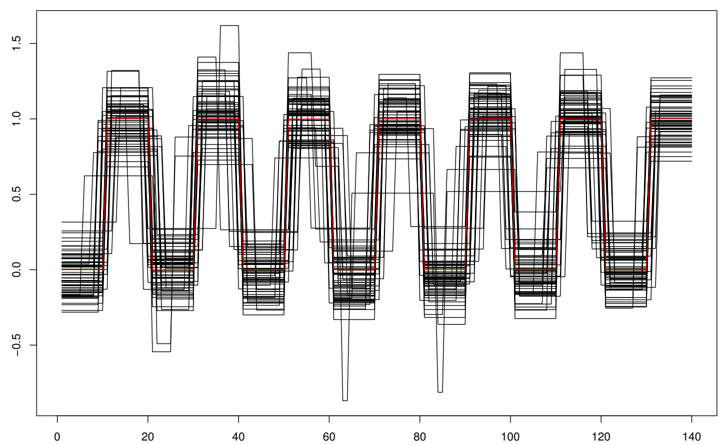

FIG 4. *Black: ensemble of those TGUH estimates of the* `teeth10` *function (model (4a)) from the simulation study described in Section 4.2 of the main article for which* $\hat{N} = N$*. Red: the true* `teeth10` *function.*

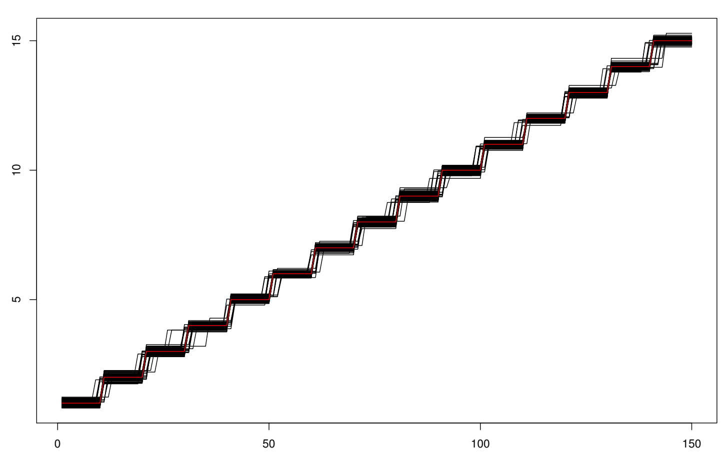

FIG 5. *Black: ensemble of those TGUH estimates of the* `stairs10` *function (model (5a)) from the simulation study described in Section 4.2 of the main article for which* $\hat{N} = N$*. Red: the true* `stairs10` *function.*

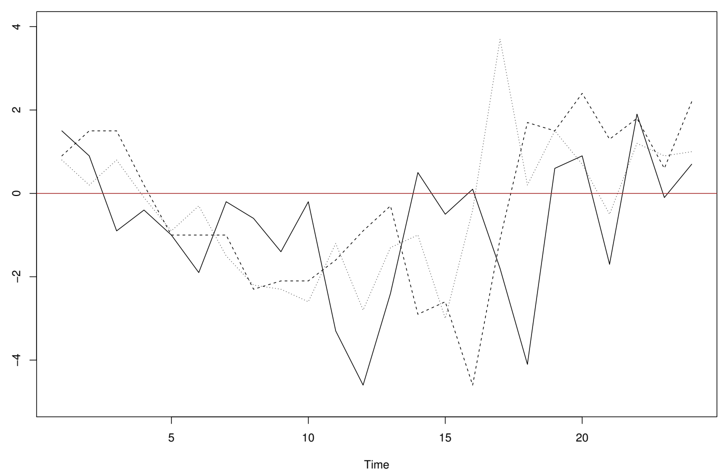

Fig 6. *The HPI increases in the boroughs of Newham (solid), Hackney (dashed) and Tower Hamlets (dotted), from January 2008 to December 2010. Brown line: zero.*