GROUND-LEVEL OZONE: EVIDENCE OF INCREASING SERIAL DEPENDENCE IN THE EXTREMES

BY DEBBIE J. DUPUIS[1] AND LUCA TRAPIN

HEC Montréal and Università Cattolica del Sacro Cuore

As exposure to successive episodes of high ground-level ozone concentrations can result in larger changes in respiratory function than occasional exposure buffered by lengthy recovery periods, the analysis of extreme values in a series of ozone concentrations requires careful consideration of not only the levels of the extremes but also of any dependence appearing in the extremes of the series. Increased dependence represents increased health risks and it is thus important to detect any changes in the temporal dependence of extreme values. In this paper we establish the first test for a change point in the extremal dependence of a stationary time series. The test is flexible, easy to use and can be extended along several lines. The asymptotic distributions of our estimators and our test are established. A large simulation study verifies the good finite sample properties. The test allows us to show that there has been a significant increase in the serial dependence of the extreme levels of ground-level ozone concentrations in Bloomsbury (UK) in recent years.

1. Introduction. Ground-level ozone is a secondary pollutant that is formed in the atmosphere due to chemical reactions that depend on meteorological conditions, the latter being most favorable for synthesis in the summer. The ozone precursor compounds of relevance are methane, nonmethane volatile organic compounds (VOC), nitrogen oxides (NO_x) and carbon monoxide. Controls of NO_x and VOC emissions in the United Kingdom (UK) and Europe have generally led to decreases in the intensity of summer ozone episodes [Air Quality Expert Group (2009)].

We study ground-level ozone concentrations in Bloomsbury (UK) and examine the daily maximum of the 8-hour running mean, measured in units of micrograms per cubic metre, from January 1993 to December 2014. Data are obtained from the Environmental Research Group, King's College London (2015). The upper panel of Figure 1 shows decreases in the intensity of summer ozone episodes. We also see that there is a substantial year-to-year variability in summer ozone concentrations because of the variability in the weather, and that the warmest summers (1995, 2003, 2006) led to large concentrations.

Extreme values of ground-level ozone have been studied in the literature. For example, Smith (1989) seeks to detect trends in extreme ozone levels, Eastoe and
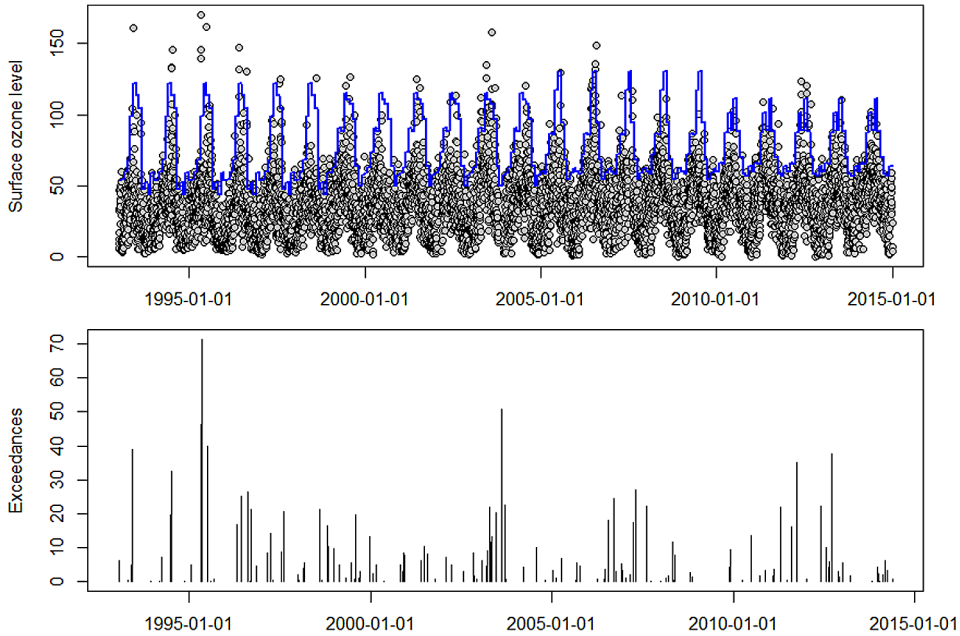
FIG. 1. *Ozone levels in Bloomsbury. Upper panel: Ground-level ozone in* $\mu g/m^3$ *(dots) and the dynamic threshold (blue line). Threshold is set to the* 98*th quantile of the detrended and deseasonalized ozone levels over the* 1993–2014 *time period. Detrending and deseasonalization is detailed in Section* 6. *Lower panel: detrended and deseasonalized exceedances.*

Tawn (2009) seek to explain the changes in extreme ozone levels conditionally on the covariates relating to the precursor concentrations and meteorological conditions, while Dupuis (2005) studies the spatial dependence in extreme behavior. In the first paper the author considers cluster maxima so as to assume temporal independence. In the second paper the authors assume that extreme events of ozone levels, or of ozone levels given the covariates, are temporally independent, while in the third the author works with slightly less correlated weekly maxima under an independence assumption. Noven, Veraart and Gandy (2015a) are the first to consider the temporal dependence of a time series of extreme ozone levels.

When studying extreme values in a time series context both the tail decay and the dependence between consecutive extreme observations are of interest. The first quantity is informative of whether the occurrence of an extreme event is more or less probable than under a Gaussian paradigm and is determined by the marginal distribution of the time series. The second provides information regarding the persistence of a sequence of extreme events and is determined by the extremal dependence structure of the time series.

Figure 2 shows empirical estimates of the conditional probability that the ozone level in Bloomsbury on day t exceeds a high threshold given that this threshold was exceeded on day $t - 1$. Estimates of the conditional probability in the latter part of
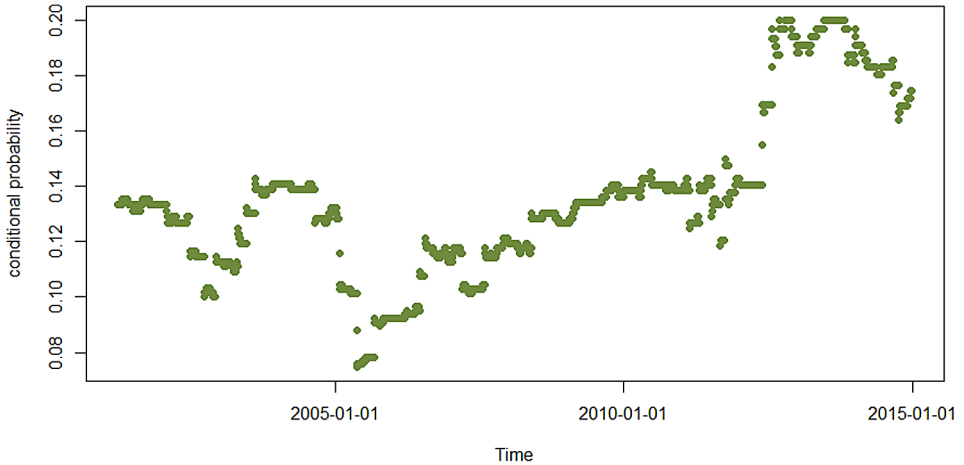
FIG. 2. *Exploring serial dependence in the extremes. Empirical estimates of the conditional probability that the ozone level in Bloomsbury on day t exceeds a high threshold given that this threshold was exceeded on day $t-1$. Estimate for day t is based on detrended and deseasonalized ozone levels in Bloomsbury from day $t-3000$ to day t. Threshold is set to the 98th quantile of the detrended and deseasonalized ozone levels over the 1993–2014 time period.*

the sample period are more than twice the size of some of the estimates in the first half of the sample period, indicating that the serial dependence of extreme ozone levels in Bloomsbury may have changed over the period studied. Our interest lies in detecting a possible change in the extremal dependence structure over the 1993–2014 time period. There is no methodology currently available to carry out this task.

Clustering of extremes is ubiquitous in natural sciences and social-economic phenomena, and extensive efforts have been made in the extreme value literature to measure the degree of dependence appearing in the extremes of a stationary time series [Davis and Mikosch (2009), Ferro and Segers (2003), Ledford and Tawn (2003)]. Investigations directed at determining whether the behavior of this dependence changes over time, however, are currently missing.

Detection of change points in the extremes typically focuses on the tail decay [Dierckx and Teugels (2010), Kim and Lee (2009), Quintos, Fan and Phillips (2001)] and tries to understand whether extreme events have become more or less likely over time. In contrast this paper focuses on the dependence aspect and seeks to determine whether the persistence of extreme events has increased or decreased over time.

Given a stationary time series $\{Y_t\}$, we consider a high threshold u and obtain a censored sequence of extreme observations $\{X_t\}$, where $X_t = \max\{Y_t - u, 0\}$. Standard asymptotic arguments suggest that the distribution of the exceedances $Y_t - u$ given that $Y_t > u$ converges to a Generalized Pareto (GP) distribution [Pickands (1975)]. Following Bortot and Gaetan (2014), we use a decomposition

of the GP distribution to construct a hierarchical model for exceedances that preserves the GP distribution marginally. Rather than using a Markov chain as the latent process as in Bortot and Gaetan (2014), we proceed as in Noven, Veraart and Gandy (2017b) and use a trawl process as the latent process. Trawl processes are stationary, infinitely divisible stochastic processes that have been recently introduced by Barndorff-Nielsen, Benth and Veraart (2011) and that have already shown wide applicability; see, for example, Shephard and Yang (2017). Applying a subsequent marginal transformation model as suggested by Bortot and Gaetan (2014) and explored in Noven, Veraart and Gandy (2017b), we manage to separate the marginal parameters and the dependence parameter. A change-point test for the *extremal dependence* can then be established testing the stability of the dependence parameter over time.

Our model presents intractable likelihoods, therefore, standard inference and testing strategies are not feasible in this setting. We rely on multistage composite likelihood to estimate the model parameters. We then extend the Wald test based on generalized method of moments (GMM) estimators of Andrews (1993) to the class of composite likelihood estimators to test the stability of the dependence parameter.

Throughout the paper we develop the methodology around the time series model for extremes of Noven, Veraart and Gandy (2017b); however, our approach is fully applicable to a more general class of models that can be estimated by composite likelihood.

This paper makes three main contributions. First, to the best of our knowledge, it presents the first change-point test for the serial dependence of a time series of extremes. Second, we extend the test for parameter stability of Andrews (1993) based on the GMM to the case of composite likelihood. Finally, our analysis of ground-level ozone data from Bloomsbury (UK) shows that, although extreme levels of ozone have decreased since 1993, there has been a significant upward shift in their serial dependence over the same period.

The remainder of the paper is organized as follows. Section 2 presents the time series model for the extremes. Section 3 presents the two-step pairwise likelihood approach to estimate the model. Section 4 presents the test for structural changes in the extremal dependence. Section 5 studies via simulation the finite sample properties of the two-step pairwise likelihood estimator and the size and the power of the test. The ozone data are analyzed in Section 6. A complete discussion of our results appears in Section 7. We outline possibilities for future methodological research in Section 8. More technical results are gathered in the Appendix.

2. Modelling the extremes of a time series. Let $\{Y_t\}$ be a stationary sequence and define the extremes as the observations of $\{Y_t\}$ exceeding a high threshold u. Define the censored sequence of excesses $\{X_t\}$ as

$$X_t = (Y_t - u)\mathbb{I}_{\{Y_t > u\}},$$

where $\mathbb{I}_{\{\cdot\}}$ denotes the indicator variable of the set $\{\cdot\}$. If $\{Y_t\}$ is *iid*, then the event of exceeding the threshold u, $\mathbb{I}_{\{Y_t > u\}}$, is a Bernoulli random variable, and the distribution of the excesses, $\Pr(Y_t - u < y | Y_t > u)$, converges to a GP distribution as $u \to \infty$ [Pickands (1975)].

2.1. *A hierarchical model.* When the sequence $\{Y_t\}$ presents dependence, a model for the joint distribution of the extremes is required. Bortot and Gaetan (2014) exploit a useful decomposition of the GP distribution into an exponential-gamma mixture and propose a hierarchical model for the extremes. Like Bortot and Gaetan (2014), we assume that the distribution of X_t depends on the value of a latent process Λ at time t, denoted by Λ_t, and that conditionally on the latent process Λ the X_t are independent and $X_t | (X_t > 0, \Lambda_t) \sim \text{Exp}(\Lambda_t)$. The latent process determines both the probability of observing an exceedance and the distribution of the exceedances. Like Bortot and Gaetan (2014), we assume that $\Pr(X_t > 0 | \Lambda_t = \lambda) = \exp(-\kappa\lambda)$, where κ is a parameter controlling the exceedance rate of the threshold u. Choosing any stationary stochastic process Λ with Gamma marginal distribution will lead to threshold exceedances $\{X_t : X_t > 0\}$ that follow a GP distribution so that the classical Pickands extreme-value paradigm is respected, and more specifically choosing a latent process Λ_t with marginal Gamma$(1/\alpha, \beta/\alpha)$ distribution leads to a $\{X_t : X_t > 0\}$ process that is marginally distributed as a GP distribution with scale $\beta + \alpha\kappa$ and tail parameter α, which we denote GP$(\beta + \alpha\kappa, \alpha)$ [Bortot and Gaetan (2014)].

One of the limits of the latter model is that α is tied to both the marginal distribution and the dependence structure. To obtain more flexibility, Bortot and Gaetan (2016) and Noven, Veraart and Gandy (2017b) propose an alternative formulation where the parameter controlling the dependence in the latent process is the only parameter controlling the dependence in the extremes, and this parameter is not tied to the marginal distribution. To achieve this separation, they transform the sequence $\{X_t\}$ into a sequence $\{Z_t\}$ using a standard probability transform argument. Fixing the parameters $\alpha = \beta = 1$ and applying the transformation $Z_t = g(Y_t)\mathbb{I}_{\{X_t > 0\}}$ with

$$(1) \qquad\qquad g(y) = \nu/\xi \left\{ \left(1 + \frac{y}{1+\kappa}\right)^\xi - 1 \right\},$$

they obtain a process for the extremes with GP(ν, ξ) marginal distribution but with dependence structure fully characterized by the dependence parameter of the latent process (parameter ρ in the models outlined in Section 2.2). The parameter κ is only tied to the probability of exceedance of the threshold, more specifically, $\kappa = 1/p - 1$ where $p = \Pr(X_t > 0)$. The alternative parametrization also allows the shape parameter of the GP distribution to be negative. A negative shape parameter is necessary to capture the tail behavior of many environmental processes.

2.2. *Latent trawl process.* Any latent process with a marginal Gamma distribution will yield a resulting $\{X_t : X_t > 0\}$ process with marginal distribution that respects the classical extreme-value paradigm, but we should choose the latent process wisely. As the conditional independence assumption of the hierarchical model means that any dependence between observations comes from the latent process, the latter process should have a very flexible dependence structure. Having a Markovian structure [Bortot and Gaetan (2014), Bortot and Gaetan (2016)] on the dependence in the extremes is limiting. Noven, Veraart and Gandy (2017b) propose a *trawl process* for $\{\Lambda_t\}$. Trawl processes are a new class of Lévy-driven processes defined in Barndorff-Nielsen, Benth and Veraart (2011). They are stationary infinitely divisible stochastic processes which are made up of two components, a Lévy basis and a trawl set. Let the set $A \subseteq [0, 1] \times (-\infty, 0]$ be the *trawl* and drag it through time without changing its shape to create a collection of *trawl sets* $\{A_t\}$:

$$A_t = A + (0, t) := \{(a_1, a_2 + t) : (a_1, a_2) \in A\} \qquad t > 0.$$

A trawl process is defined as $L(A_t)$ where $L(\cdot)$ is a homogeneous Lévy basis. The definition of a homogeneous Lévy basis appears in the Appendix. The Lévy basis $L(\cdot)$ can be associated to a Lévy seed L', and the law of $L(A)$ is fully determined by L' and the Lebesgue measure on the set A. The trawl process $\{\Lambda_t\}$ is defined by evaluating the homogeneous Lévy basis over the trawl set: $\Lambda_t = L(A_t)$ for $t \in \mathbb{R}$. Noven, Veraart and Gandy (2017b) consider an exponential trawl set with parameter ρ, that is, $A = \{(x, s) : 0 \le x < d(s), s \le 0\}$ with $d(s) = \exp(\rho s)$. By considering a Lévy seed that has a normalised Gamma distribution, they obtain a $\{\Lambda_t\}$ process with the required Gamma marginal distribution. See Noven, Veraart and Gandy (2017b) for a more complete discussion of trawl processes and their properties. The use of a kernel mixture of a Lévy process in the second stage of a hierarchical model has been previously suggested by Wolpert and Ickstadt (1998).

In both the framework of Bortot and Gaetan (2016) and that of Noven, Veraart and Gandy (2017b), the parameter ρ controls the temporal dependence in the time series of extremes. The main difference is that the parameter ρ in Bortot and Gaetan (2016) determines the transition probability of the Markov process while in Noven, Veraart and Gandy (2017b) the parameter ρ defines the trawl set characterizing the nonMarkovian stationary trawl process. The autocorrelation function of the exponential trawl process has the same shape as the trawl set. Figure 3 shows the autocorrelation function for different values of ρ and that larger values of ρ correspond to *weaker* dependence. Since the parameter ρ determines the level of dependence in the extremes, the problem of testing for structural changes in the dependence reduces to testing for changes in ρ.

3. Estimation and inference. The exponential trawl model for the extremes has four parameters to be estimated κ, ξ, ν and ρ. The first three parameters define the marginal GP distribution. The parameter ρ controls the dependence in the extremes, and we want to test whether it stays constant over time.
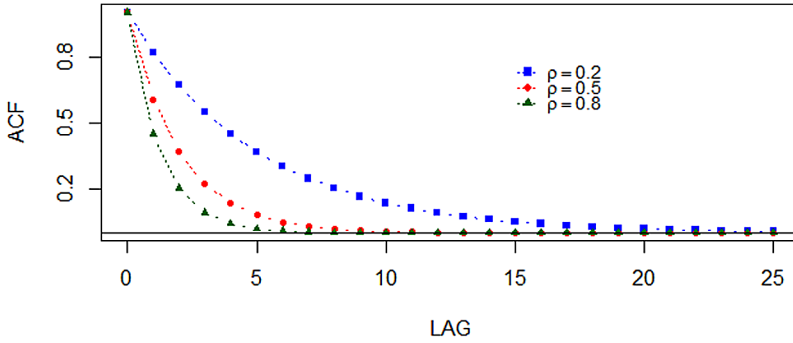
FIG. 3.   *Autocorrelation function of latent exponential trawl model.*

Recall that $\{X_t\}_{t=1}^n$ is the censored sequence of excesses of a high threshold u. If there are k nonzero X_t in $\{X_t\}_{t=1}^n$, then κ is estimated as $n/k - 1$. Letting $\theta = (\nu, \xi, \rho)' \in \Theta \subset \mathbb{R}^3$, the likelihood of the sequence of excesses under the latent exponential trawl model can be written as

$$\ell_{\text{ML}}(\theta | x_1, \dots, x_n) = \int_{\mathbb{R}_+^n} \prod_{i=1}^k J(x_i^+) \lambda_i \exp\left(-\lambda_i \left(\widehat{\kappa} + g^{-1}(x_i^+)\right)\right)$$
$$\times \prod_{j=1}^{n-k} \left(1 - \exp(-\widehat{\kappa}\lambda_j)\right) \mathrm{d}F(\lambda_1, \dots, \lambda_n),$$

where F is the joint density of the trawl process variables $\Lambda_1, \dots, \Lambda_n$, x^+ is the vector containing the k observations exceeding the threshold u, g is as in equation (1), and $J(x) = f_{\text{GP}}(x|\nu, \xi)/f_{\text{GP}}(x|1 + \widehat{\kappa}, \widehat{\kappa})$ is the Jacobian necessary to apply the marginal transformation in equation (1). Although the likelihood of the model is available analytically, it is difficult to compute as it requires multiple numerical partial derivatives to be performed. Noven, Veraart and Gandy (2017b) consider a pairwise likelihood using only observations up to Δ time steps apart, that is,

$$(2) \qquad \ell_{\text{PL}}^{\Delta}(\theta | x_1, \dots, x_n) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{\min(i+\Delta, n)} f(x_i, x_j),$$

where $f(x_i, x_j)$ is the bivariate density of the model and closed-form expressions for $f(x_i, x_j)$ can be found in the Appendix of Noven, Veraart and Gandy (2017b). Define the maximum pairwise likelihood (MPL) estimator as

$$\widehat{\theta}_{\text{MPL}}^{\Delta} = \arg\max_{\theta} \ell_{\text{PL}}^{\Delta}(\theta | x_1, \dots, x_n).$$

We conjecture that the proofs in Davis and Yau (2011) could be adapted to show that

$$\sqrt{n}(\widehat{\theta}_{\text{MPL}} - \theta_0) \xrightarrow{d} (0, H_0^{-1} J_0 H_0^{-1}),$$

where $H_0 = -E[\frac{\partial^2}{\partial\theta\partial\theta'}\ell_{\mathrm{PL}}^{\Delta}(\theta_0|x_1,\ldots,x_n)]$, $J_0 = \mathrm{Var}[\frac{\partial}{\partial\theta}\ell_{\mathrm{PL}}^{\Delta}(\theta_0|x_1,\ldots,x_n)]$ and $\theta_0 = (\nu_0, \xi_0, \rho_0) \in \Theta$ is the true parameter vector. The latter is the standard asymptotic result for *iid* sequences provided by Cox and Reid (2004) where a general discussion on pairwise likelihood can also be found.

As we want to test for changes in the dependence parameter ρ, the parameters ξ and ν are nuisance parameters in our test hypothesis. We consider a two-step procedure to further simplify the estimation strategy. The marginal distribution of the extremes is GP with parameters ξ and ν. We first estimate these parameters by maximizing the likelihood of the GP distribution and obtain the corresponding ML estimators,

$$(\widehat{\nu}_{\mathrm{ML}}, \widehat{\xi}_{\mathrm{ML}}) = \arg\max_{\nu,\xi} \ell_{\mathrm{GP}}(\nu, \xi|x_1,\ldots,x_k),$$

where $\ell_{\mathrm{GP}}(\nu, \xi) = \sum_{i=1}^{k} 1/\nu(1 + \xi x_i^+/\nu)^{(-1/\xi-1)}$. Under standard regularity conditions (satisfied only for $\xi > -0.5$ [Smith (1985)]), the estimators $(\widehat{\nu}_{\mathrm{ML}}, \widehat{\xi}_{\mathrm{ML}})$ are consistent and jointly normal, but the variance needs to be adjusted for the dependence in the observations, using, for example, the HAC estimator of Newey and West (1987). Probability plots and quantile plots are useful for assessing the quality of the fitted GP distribution; see Coles (2001).

In a second step we estimate the parameter ρ with the pairwise likelihood procedure as

$$\widehat{\rho}_{\mathrm{MPL}}^{\Delta} = \arg\max_{\rho} \ell_{\mathrm{PL}}^{\Delta}(\rho, \widehat{\kappa}, \widehat{\nu}_{\mathrm{ML}}, \widehat{\xi}_{\mathrm{ML}}|x_1,\ldots,x_n),$$

where $\ell_{\mathrm{PL}}^{\Delta}$ is as in (2). We denote the estimator obtained from these two steps, $(\widehat{\nu}_{\mathrm{ML}}, \widehat{\xi}_{\mathrm{ML}}, \widehat{\rho}_{\mathrm{MPL}}^{\Delta})$, as $\widehat{\theta}_{\mathrm{TSML}}^{\Delta}$, the two-step maximum likelihood (TSML) estimator. The second stage of the estimation strategy clearly depends on parameter estimates from the first step, and the accumulation of estimation error must be properly incorporated into standard error calculations for $\widehat{\theta}_{\mathrm{TSML}}^{\Delta}$. The two-step estimation that includes a composite likelihood stage can be viewed as a form of multistage GMM estimation, and, under standard regularity conditions, the results in Newey and McFadden (1994) mean that $\widehat{\theta}_{\mathrm{TSML}}^{\Delta}$ is asymptotically normal and

$$\sqrt{n}(\widehat{\rho}_{\mathrm{MPL}}^{\Delta} - \rho_0) \xrightarrow{d} N(0, \widetilde{\sigma}_\rho) \qquad \text{as } n \to \infty,$$

where

$$(3) \qquad \widetilde{\sigma}_\rho = \Gamma_\rho^{-1} E\big[(s_{\mathrm{PL}}^{\Delta} - \Gamma_{\nu,\xi}^{-1}\Upsilon^{-1}s_{\mathrm{GP}})(s_{\mathrm{PL}}^{\Delta} - \Gamma_{\nu,\xi}^{-1}\Upsilon^{-1}s_{\mathrm{GP}})'\big]\Gamma_\rho^{-1'}$$

with $s_{\mathrm{PL}}^{\Delta} = s_{\mathrm{PL}}^{\Delta}(x|\nu_0, \xi_0, \rho_0)$ and $s_{\mathrm{GP}} = s_{\mathrm{GP}}^{\Delta}(x|\nu_0, \xi_0)$ the scores of the pairwise and GP likelihood respectively, and $\Gamma_\rho^{-1} = E[\nabla_\rho s_{\mathrm{PL}}^{\Delta}(\nu_0, \xi_0, \rho_0)]$, $\Gamma_{\nu,\xi}^{-1} = E[\nabla_{\nu,\xi} s_{\mathrm{PL}}^{\Delta}(\nu_0, \xi_0, \rho_0)]$, and $\Upsilon = E[\nabla_{\nu,\xi} s_{\mathrm{GP}}(\nu_0, \xi_0)]$ where $\nabla_{\{\cdot\}}$ denotes the gradient with respect to $\{\cdot\}$. The parameter κ is set to $\widehat{\kappa}$ in the above variance calculations. The asymptotic variance of $\widehat{\theta}_{\mathrm{TSML}}^{\Delta}$ incorporates the uncertainty of the first

step and is greater than or equal to the asymptotic variance of $\widehat{\theta}_{\mathrm{MPL}}^{\Delta}$. We investigate the extent of the efficiency loss of TSML relative to MPL in a simulation study in Section 5.

To assess the goodness of fit of our model for serial dependence, we need to go beyond our estimates for ρ. We compare empirical and model-implied extremal index [Leadbetter (1983)] estimates. The extremal index can be loosely defined as (limiting mean cluster size)$^{-1}$ where limiting is in the sense of clusters of exceedances of increasingly high thresholds [Coles (2001)]. Similarly, model estimates of $\Pr(Y_{t+k} > u | Y_t > u)$ can be compared to empirical estimates. The model-implied probabilities, along with confidence bands, can be computed via Monte Carlo simulation. Data-based estimates should be within the bands if the model for serial dependence is appropriate.

4. Detection of a structural break in the dependence. We want to perform a test on the stability of the parameter ρ against the alternative of an unknown change point. We use our GMM estimator and rely on the Wald-type test for a structural break of Andrews (1993). More precisely, we consider the exponential-trawl model for the extremes indexed by parameters ν_0, ξ_0, and ρ_t for $t = 1, \ldots, n$. Letting $\mathcal{P} \subset \mathbb{R}$ be the parameter space of ρ, the null hypothesis of interest is

$$\mathcal{H}_0 : \rho_t = \rho_0 \qquad \text{for all } t \geq 1 \text{ for some } \rho_0 \in \mathcal{P}.$$

The alternative hypothesis is a one-time structural change with change point $\pi \in (0, 1)$,

$$\mathcal{H}_1(\pi) : \rho_t = \begin{cases} \rho_1(\pi) & \text{for } t = 1, \ldots, \lfloor \pi n \rfloor, \\ \rho_2(\pi) & \text{for } t = \lfloor \pi n \rfloor + 1, \ldots \end{cases}$$

for some constants $\rho_1(\pi), \rho_2(\pi) \in \mathcal{P} \subset \mathbb{R}$. The parameters ν_0 and ξ_0 are taken to be constant under the null hypothesis and the alternative, and are considered nuisance parameters in the estimation and in the test statistic. We focus on a one-time change in the value of the parameter ρ at an unknown break time π_0. A discussion of a possible extension to the case of multiple changes appears in Section 8.

As we do not know π_0, we need to test $\mathcal{H}_0$ against a composite alternative $\bigcup_{\pi \in (0,1)} \mathcal{H}_1(\pi)$. As discussed in Andrews (1993), the standard Wald-, LM- and LR-type tests do not possess their usual asymptotic distributions when π is treated as a parameter as it appears only under the alternative. As a consequence a Wald-test statistic of the form $\sup_{\pi \in \Pi} W_n(\pi)$, where Π is some prespecified subset of $(0, 1)$, is considered.

Define the *partial-sample* estimator $\widehat{\rho}_{\mathrm{PS}}^{\Delta}(\pi) = (\widehat{\rho}_1^{\Delta}(\pi), \widehat{\rho}_2^{\Delta}(\pi))'$ as

(4)
$$\arg\max_{\rho_1, \rho_2} \left\{ \sum_{i=1}^{n\pi} \sum_{j=i+1}^{i+\Delta} f(x_i, x_j | \rho_1, \widehat{\nu}_{\mathrm{ML}}, \widehat{\xi}_{\mathrm{ML}}, \widehat{\kappa}) \right. $$
$$\left. + \sum_{i=n\pi+1}^{n-1} \sum_{j=i+1}^{\min(i+\Delta, n)} f(x_i, x_j | \rho_2, \widehat{\nu}_{\mathrm{ML}}, \widehat{\xi}_{\mathrm{ML}}, \widehat{\kappa}) \right\},$$

where $\widehat{v}_{\mathrm{ML}}, \widehat{\overline{\xi}}_{\mathrm{ML}}$ and $\widehat{\kappa}$ are obtained from the full sample. As $\widehat{\rho}_{\mathrm{PS}}^{\Delta}(\pi)$ is a pairwise likelihood estimator, it is included in the partial-sample GMM estimator definition of Andrews (1993) and under mild regularity conditions it follows from Theorem 1 of the latter that

$$\sqrt{n}\big(\widehat{\rho}_1^{\Delta}(\pi) - \widehat{\rho}_2^{\Delta}(\pi)\big) \overset{d}{\to} N\left(0, \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\widetilde{\sigma}_{\rho}\right),$$

where $\widetilde{\sigma}_{\rho}$ is given in (3). According to equations (3.11)–(3.13) in Andrews (1993), consistent estimators of $\widetilde{\sigma}_{\rho}$ can be obtained from the full sample, and we rely on the $\widehat{\widetilde{\sigma}}_{\rho}$ estimator, based on equation (6.12) of Newey and McFadden (1994). Our estimator takes the form in equation (5) in the Appendix. It is easy to see from (5) that uncertainty in the first step estimation adds to the "sandwich" variance of the ML estimator in the second step. The Wald statistic for testing $\mathcal{H}_0$ against $\bigcup_{\pi \in \Pi} \mathcal{H}_1(\pi)$ is given by

$$W_n(\pi) = n \frac{(\widehat{\rho}_1^{\Delta}(\pi) - \widehat{\rho}_2^{\Delta}(\pi))^2}{(\frac{1}{\pi} + \frac{1}{1-\pi})\widehat{\widetilde{\sigma}}_{\rho}},$$

and from Theorem 3 of Andrews (1993), we have that, under the null hypothesis,

$$\sup_{\pi \in \Pi} W_n(\pi) \overset{d}{\to} \sup_{\pi \in \Pi} \frac{D(\pi)^2}{\pi(1-\pi)},$$

where $D(\pi)$ is a Brownian bridge on the interval [0, 1]. As evaluation of the statistic $W_n(\pi)$ requires a minimum number of observations in each of the two terms in (4), we look for changes over the subinterval $\Pi = (\bar{\pi}, 1 - \bar{\pi})$ where $\bar{\pi} > 0$ is large enough to assure this minimum. The minimum value of $\bar{\pi}$ that may be considered is thus dependent on the threshold u, with higher thresholds demanding larger values of $\bar{\pi}$ and limiting the time period over which a change point can be identified. Asymptotic critical values are tabulated for different values of $\bar{\pi}$ in Andrews (1993).

We also obtain consistency results under the alternative hypothesis following Andrews (1993) where the limiting distribution of the test under a set of local alternative hypotheses is found. These results are outlined in the Appendix, and the asymptotic power can be obtained by simulation. Finite sample performance is examined in the following section.

5. Simulation study. We design simulation studies to assess the finite sample properties of the proposed two-step estimator and of the test for the null hypothesis of constancy of ρ. We consider both the heavy-tailed case assuming a marginal GP distribution with parameters $v = 15$ and $\xi = 0.15$ and the light-tailed case where the GP parameters take values $v = 4$ and $\xi = -0.1$. The values of ξ have been set according to the estimates obtained in the analysis of ozone data in Noven, Veraart and Gandy (2017b) and are realistic values for the tail decay in general. The values

of ν are chosen to explore two different scales of the GP distribution. The value of κ is set such that the number of exceedances is k where k is specified below. To generate the dependence in the extremes, we consider latent exponential trawl processes with $\rho = \{0.2, 0.5, 0.8\}$. Figure 3 shows the corresponding autocorrelation functions.

5.1. *Performance of the two-step estimator.* In this section, we use simulations to compare the finite sample properties of the pairwise likelihood estimator $\widehat{\theta}^{\Delta}_{\text{MPL}}$ of Noven, Veraart and Gandy (2017b) and our two-step estimator $\widehat{\theta}^{\Delta}_{\text{TSML}}$, most notably assessing efficiency, any small-sample biases and the possible impact of the value of Δ.

We generate censored samples of n observations fixing the number of exceedances at $k = 0.1 \cdot n$. The value of k is set to match the number of exceedances considered in our ozone data. We estimate the parameter ρ with both estimators using only observations up to $\Delta = \{4, 10, 20\}$ time steps apart. As there exists a trade-off between efficiency of the estimator and computational burden in the choice of Δ, it is important to understand to what extent it is worth saving computing time at the expense of efficiency. We examine $B = 1000$ replications under heavy- and light-tail scenarios with sample sizes $n = \{2000, 5000\}$.

Figure 4 shows the bias and the standard deviation of the estimates. Estimates for the dependence parameter ρ are nonbiased and show similar variability under both heavy-tailed and light-tailed margins. For both MPL and TSML estimators, the performance is almost the same for $\Delta = 4$, 10 and 20, with more numerous time steps improving only slightly the %bias of the MPL estimates of the marginal scale parameter ν when $n = 2000$. We conclude that the values of the marginal tail parameters do not affect the performance and that $\Delta = 4$ represents a good choice for both estimators. This value of Δ was also suggested in Noven, Veraart and Gandy (2017b). Both estimators show small biases and root mean square errors. As the two-step estimator performs as well as $\widehat{\theta}^{\Delta}_{\text{MPL}}$, losing relatively little efficiency, we move forward confidently using $\widehat{\theta}^{\Delta}_{\text{TSML}}$.

Table 1 reports the rejection probabilities from the different scenarios. With $n = 2000$, the test is slightly oversized when $\rho = 0.2$ and slightly undersized when $\rho = 0.8$. Note however that there is little information available for the estimation as the partial-sample spans from 800 $(\bar{\pi}n)$ to 1200 $(1 - \bar{\pi}n)$ observations only. With $n = 5000$ the size is better respected. When ρ equals 0.2 and 0.5 the test presents the correct size, but it is still slightly undersized when $\rho = 0.8$, that is, for the least dependent of our three latent trawl processes. The thickness of the marginal tail has no impact on the size.

5.2. *Power of the test.* We assess the power of the test assuming: (i) an increase in the strength of the dependence, $\rho_1 = 0.5 \rightarrow \rho_2 = 0.2$, and (ii) a decrease in the strength of the dependence, $\rho_1 = 0.5 \rightarrow \rho_2 = 0.8$. True location of the
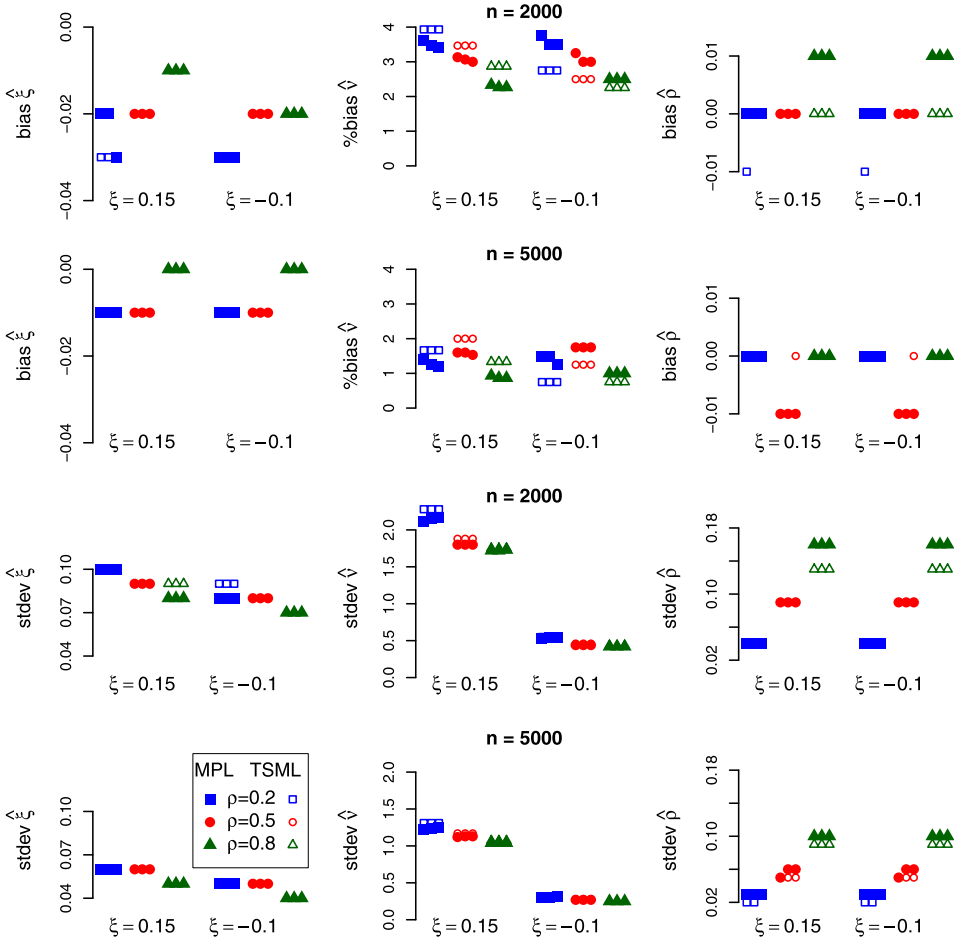
FIG. 4. *Performance of estimators. Bias and standard deviation of the estimates computed over B = 1000 replications. Results are for heavy-tails ($\xi = 0.15$, $v = 15$) and light-tails ($\xi = -0.10$, $v = 4$). %bias is shown for estimators of v. For each combination of tail thickness and dependence parameter ρ, triplets show results for estimates based on observations only up to $\Delta = 4$, 10 and 20 time steps apart respectively. Symbols for MPL are shown when results are overlapping.*

change point is $\pi_0 = 0.5$. We generate censored samples of n observations fixing the number of exceedances at $k = 0.1 \cdot n$. We consider $n = \{2000, 5000\}$ and perform the study under both the heavy- and light-tail scenarios. We use $\Delta = 4$ in the partial-sample two-step estimator $\widehat{\rho}_{PS}^{\Delta}(\pi)$ and consider $\bar{\pi} = 0.4$, that is, subinterval $\Pi = [0.4, 0.6]$.

Table 2 reports the rejection probabilities under each alternative hypothesis. Power, like size, appears unaffected by the decay rate of the marginal tail. Figure 3 shows that a change in the parameter ρ from 0.5 to 0.2 represents a larger change in dependence in the latent process than that from ρ equal 0.5 to ρ equal

TABLE 1

Size of the test. Rejection probabilities under the null hypothesis computed over $B = 1000$ replications using $\Delta = 4$ in the estimation. Heavy *and* Light *correspond respectively to the scenarios ($\xi = 0.15$, $\nu = 15$) and ($\xi = -0.1$, $\nu = 4$). Level of significance of the test is α*

| | $n = 2000$ | | | | | | $n = 5000$ | | | | | |
| | Heavy | | | Light | | | Heavy | | | Light | | |
$\rho \setminus \alpha$	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
0.2	0.12	0.07	0.02	0.12	0.07	0.02	0.11	0.05	0.01	0.11	0.05	0.01
0.5	0.11	0.05	0.01	0.11	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
0.8	0.07	0.03	0.004	0.07	0.03	0.004	0.08	0.04	0.01	0.08	0.04	0.01

0.8. In the former case our test already exhibits good power when $n = 2000$, while for the latter good power is attained when $n = 5000$. The size of the change in the dependence, and not the direction of the change, is important.

6. Data analysis. Our interest lies in detecting a possible change in the extremal dependence structure of the ozone levels in Bloomsbury over the 1993–2014 time period, and we use the test developed in Section 4 to do so. Exceedances over 81 μg/m^3 still show considerable seasonality and cannot be the entrants for our test for a structural break in the dependence. Recall that our test is consistent if the parameters of the GP are constant over time. To achieve the latter constancy, we consider exceedances over a carefully selected, although easily computed, time-varying threshold. More precisely, we divide the 22 years of data into four bins of length six, six, five and five years respectively. We choose a threshold for each bin-month combination so as to simultaneously take care of the seasonal effect and the downward trend in time. The threshold for each bin-month combination is set to the 98th quantile of the data therein. Inspection of quantile-quantile plots (not shown) for the fitted GP models reveals that the latter offer a good fit when using

TABLE 2

Power of the test. Rejection probabilities under the alternative hypothesis ($\rho_1 \to \rho_2$) computed over $B = 1000$ replications using $\Delta = 4$ in the estimation. Heavy *and* Light *correspond respectively to the scenarios ($\xi = 0.15$, $\nu = 15$) and ($\xi = -0.1$, $\nu = 4$). Level of significance of the test is α*

| | $n = 2000$ | | | | | | $n = 5000$ | | | | | |
| | Heavy | | | Light | | | Heavy | | | Light | | |
$\rho_1 \to \rho_2 \setminus \alpha$	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
$0.5 \to 0.2$	0.75	0.64	0.40	0.75	0.63	0.41	0.98	0.95	0.87	0.98	0.96	0.89
$0.5 \to 0.8$	0.27	0.16	0.05	0.28	0.16	0.06	0.60	0.44	0.21	0.59	0.44	0.21

TABLE 3
Full sample estimates. The standard errors (s.e.) for the GP
parameters are corrected to account for the dependence.
The standard error for ρ accounts for the noise induced by
the first-step estimator

	ν	ξ	ρ
est.	5.63	0.37	0.41
s.e.	0.86	0.14	0.05

a threshold set to the 98th quantile and beyond. We use the 98th quantile so as to include the maximum number of points in the analysis. The time-varying threshold appears as the solid line in the upper panel of Figure 1, and the corresponding exceedances are reported in the lower panel of Figure 1.

We start by estimating the model parameters over the whole sample using the two-step estimator $\widehat{\theta}^{\Delta}_{\text{TSML}}$ and only observations $\Delta = 4$ time steps apart. The estimates reported in Table 3 reveal that the marginal GP distribution presents a heavy-tail after detrending and deseasonalizing.

To assess the goodness of fit of our model, we compare empirical and model-implied estimates of the extremal index. Figure 5 shows data-based estimates of the extremal index along with estimates of the extremal index for the latent trawl model and both latent Markov chain models of Bortot and Gaetan (2016). Time series of length 100,000 are simulated from the fitted models. The extremal index is estimated using the intervals estimator of Ferro and Segers (2003) as implemented in the R package `extRemes` [Gilleland and Katz (2016)]. Both latent Markov chain models underestimate the extremal dependence, and the extremal dependence structure for the deseasonalized data is better captured by the trawl-based model. In order to investigate possible changes in ρ, we apply the test statistic $W_n(\pi)$ over $\Pi = [0.25, 0.75]$, still using $\Delta = 4$ for the computation of the partial-sample estimator. Figure 6 shows the value of the test for each $\pi \in [0.25, 0.75]$ along with the asymptotic critical values for the tests at the 10% and 5% levels of significance. The null hypothesis of constancy of the ρ parameter is rejected at the 10% level on 2007/09/23–24 and from 2008/03/11 to 2009/07/08 (except for on 2008/04/20), with the statistic reaching its maximum on 2008/11/16. The partial-sample estimates on 2008/11/16 are $\widehat{\rho}_1 = 0.51$ and $\widehat{\rho}_2 = 0.24$, indicating that the strength of the dependence is greater in more recent years.

To assure the consistency of our test, we use the change-point algorithm of Dierckx and Teugels (2010) to test for the null of constancy of the GP parameters over the sample. We find a p-value of 0.41, and the null hypothesis of constant GP parameters cannot be rejected.[2]

---

[2]The Dierckx and Teugels (2010) test is designed for *iid* data. We also apply the test to declustered data in order to account for the dependence, but the result does not change. Since the presence of
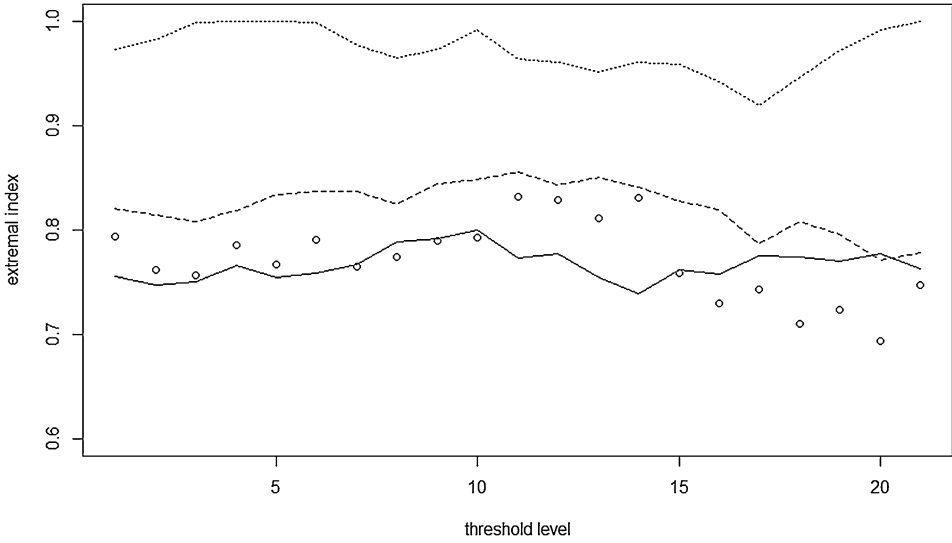
FIG. 5. *Estimated extremal index. For deseasonalized ozone data* (*circles*), *latent trawl model* (*solid*), *latent Markov chain model with Walker process* (*dashed*) *and latent Markov chain model with Warren process* (*dotted*). *Extremal index* $=$(*limiting mean cluster size*)$^{-1}$.
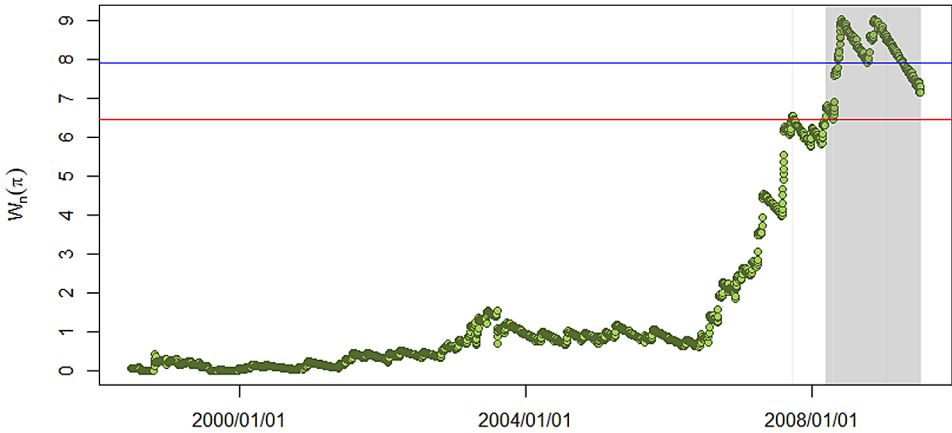

FIG. 6. *Test statistics. The bullets represent the value of the test statistic* $W_n(\pi)$ *for each* $\pi \in [0.25, 0.75]$, *that is, for each day between* 1998/06/24 *and* 2009/07/09, *for the ozone data in Bloomsbury* (*UK*). *The red* (*blue*) *line represents the critical value of the test at the* 10% (5%) *level of significance when* $\bar{\pi} = 0.25$. *A value of the test statistic beyond these lines supports rejection of the null hypothesis at the corresponding level of significance. The grey-shaded area marks days for which the null hypothesis is rejected at the* 10% *level.*

dependence in the data induces more false positives, nonrejection of the null for our data can really be seen as a substantial lack of evidence in favor of non constant GP parameters.
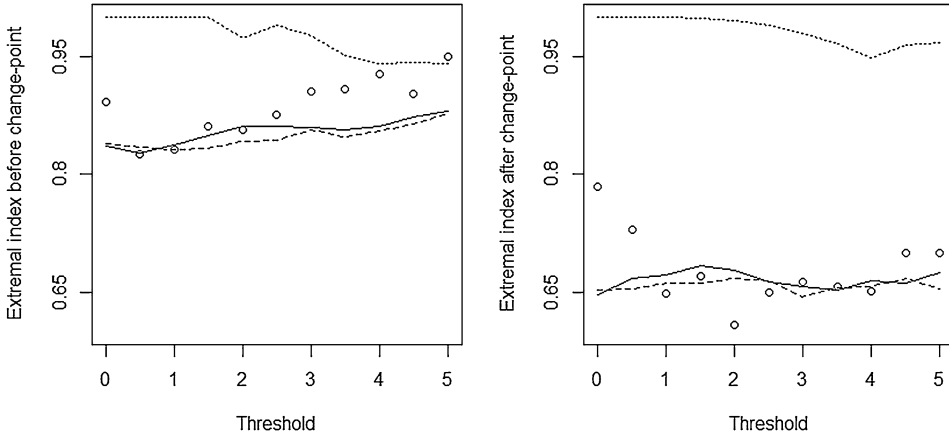
FIG. 7. *Estimated extremal index. For deseasonalized ozone data (circles), latent trawl model (solid), latent Markov chain model with Walker process (dashed) and latent Markov chain model with Warren process (dotted). Models are fitted to data before (left) and after (right) the estimated structural break of 2008/11/16. Extremal index =(limiting mean cluster size)$^{-1}$.*

Having established a structural break on 2008/11/16 using our trawl-based test, we can also consider the fit of our trawl-based model and both latent Markov models of Bortot and Gaetan (2016) to the two partial samples. Proceeding as we did for Figure 5, we calculate estimates of the extremal index under all models for each partial sample. Figure 7 shows that over these partial samples, the asymptotically independent trawl-based process and the asymptoticaly dependent latent Markov chain model with Walker process, demonstrate similarly good fits. The asymptotically independent latent Markov chain model with Warren process is inadequate, especially in the second partial sample where more recent ozone levels show greater serial dependence in their extremes. While also asymptotically independent, the trawl-based process is showing good flexibility and adapting well to the serial dependence observed at high thresholds.

We carry out some robustness checks. There are three extremely large observations in May 1995, visible in the lower panel of Figure 1, which cannot be explained by unusually warm temperatures. To ensure that our findings do not rest on these possibly influential observations, we repeat the analysis excluding the records for May 5–7, 1995. Full sample estimates are $\widehat{\nu} = 5.72$ (0.90), $\widehat{\xi} = 0.28$ (0.11) and $\widehat{\rho} = 0.43$ (0.05). Figure 8 shows the values of the $W_n(\pi)$ statistic computed for the reduced dataset, and we find even greater evidence of a change in dependence over the end of the observation period. The null of constant ρ is rejected at the 10% level over the entire 2007/08/06 to 2009/07/09 period, with the statistic reaching its maximum on 2008/10/20. The partial-sample estimates are $\widehat{\rho}_1 = 0.56$ and $\widehat{\rho}_2 = 0.24$, indicating an even weaker extremal dependence in the first part of the sample if data for May 5–7, 1995 are removed. We find a *p*-value
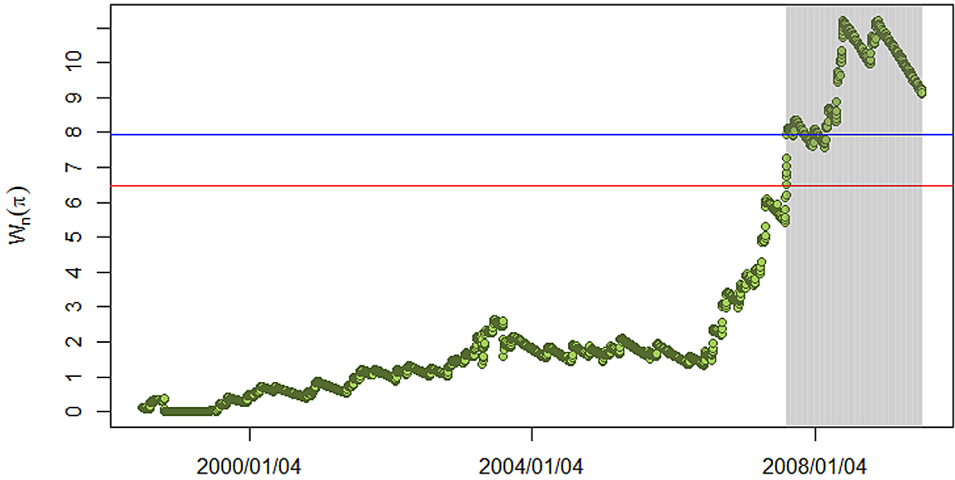
FIG. 8. *Test statistics removing data from May 5–7, 1995. The bullets represent the value of the test statistic $W_n(\pi)$ for each $\pi \in [0.25, 0.75]$ for the ozone data in Bloomsbury (UK). The red (blue) line represents the critical value of the test at the 10% (5%) level of significance when $\bar{\pi} = 0.25$. A value of the test statistic beyond these lines supports rejection of the null hypothesis at the corresponding level of significance. The grey-shaded area marks days for which the null hypothesis is rejected at the 10% level.*

of 0.57 for the test of constant GP parameters. All analyses corroborate our initial findings.

Figure 9 reports $\Pr(Y_{t+k} > u | Y_t > u)$ estimates for the pre and poststructural break exceedances respectively, along with the confidence bands of the model-implied estimates obtained with Monte Carlo simulations. We can see that splitting the sample allows the model to nicely capture the extremal dependence observed in the data. It is clear that the model-implied estimates from the prestructural break period (blue-shaded area) do not reproduce the larger dependence observed in the data at small lags during the poststructural break period as the estimates at the first four lags for the poststructural break period (red dots) lie clearly outside the blue band. Moreover, these red dots lie almost exactly in the middle of the red band representing model implied estimates of our post structural break model which nicely captures $\widehat{\rho}_2 \ll \widehat{\rho}_1$. We see much of the same in Figure 10 where the analysis is repeated without data for May 5–7, 1995.

7. Discussion. High levels of ground-level ozone concentrations can have an adverse effect on both human health and vegetation. Along with the number of episodes in a season, the length of the recovery period between successive episodes of high ground-level ozone concentrations can be an important factor in the nature and magnitude of health impacts [WHO (1987)]. Similarly, the impact of the exposure on plants depends on levels, duration, frequency, the time of day, the season and the interval between exposures [Ackermann et al. (1999)]. Any changes in the
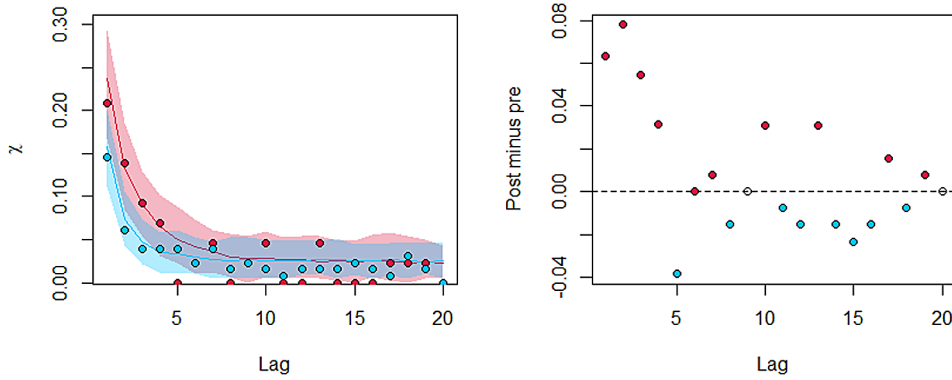
FIG. 9. *Extremal dependence. Empirical estimates of* $\Pr(Y_{t+k} > u | Y_t > u)$ *and 95% confidence interval of the model implied estimates, computed on the exceedances of ozone data in Bloomsbury (UK), pre (blue) and poststructural (red) break (left panel). Right panel shows the difference between the pre and poststructural break empirical estimates of* $\Pr(Y_{t+k} > u | Y_t > u)$.

temporal dependence of extreme ozone levels should be detected as studies seeking to evaluate present risks should only be compiling data over comparable and relevant time periods, for example.

Working within a hierarchical trawl-based framework, we propose a test for a change point in the extremal dependence of a stationary time series. The test is flexible and easy to use. The strength of our approach lies in its ability to reduce the dependence modeling to a one-dimensional problem that can be cast in the GMM framework of Andrews (1993). Our test allows us to go beyond a suspected increased dependence based on empirical checks as in Figure 2 to establish a sta-
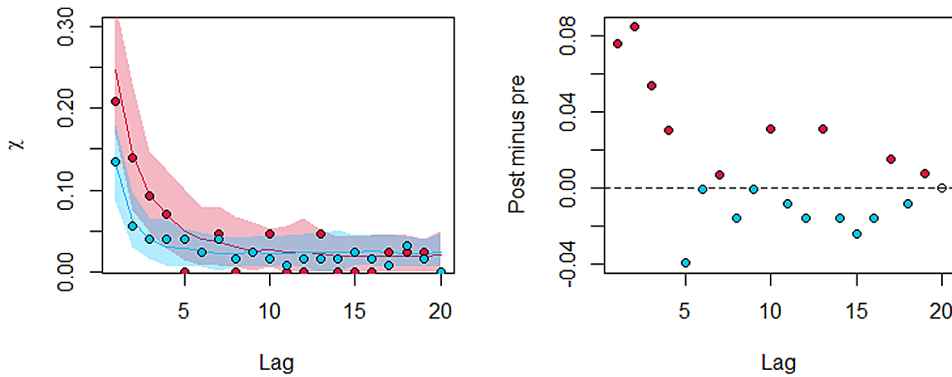

FIG. 10. *Extremal dependence removing data from May 5–7, 1995. Empirical estimates of* $\Pr(Y_{t+k} > u | Y_t > u)$ *and 95% confidence interval of the model implied estimates, computed on the exceedances of ozone data in Bloomsbury (UK), pre (blue) and poststructural (red) break (left panel). Right panel shows the difference between the pre and poststructural break empirical estimates of* $\Pr(Y_{t+k} > u | Y_t > u)$.
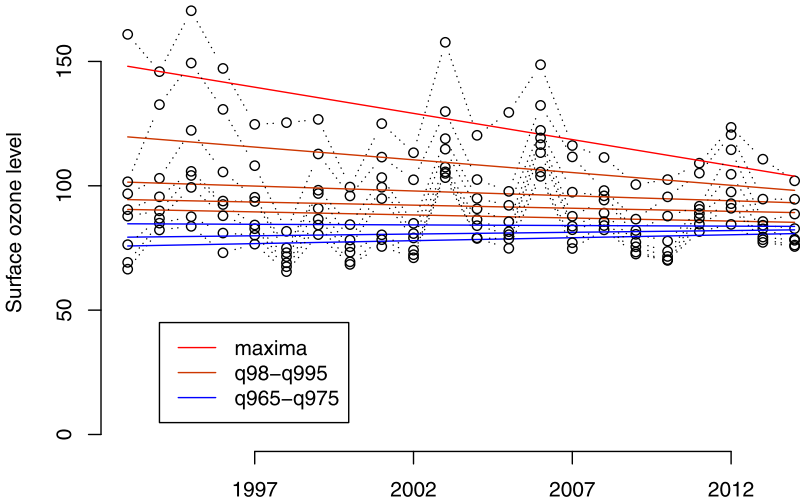
FIG. 11.   *Upper tail of ozone levels in Bloomsbury. Open circles show annual maxima and upper quantiles* (99.5*th down to* 96.5*th in steps of* 0.5) *of the daily maximum* 8*-hour running mean of ground-level ozone concentration. Solid lines show fitted linear regression lines through each of annual maxima* (*top*) *to* 96.5*th quantiles* (*bottom*).

tistically significant change. Our analyses show that extreme episodes have become increasingly clustered. Controls of NO_x and VOC emissions in the UK and Europe have generally led to decreases in the intensity of summer ozone episodes [Air Quality Expert Group (2009)], and Figure 11 shows that the annual maximum of the daily maximum 8-hour running mean has in fact decreased in Bloomsbury over the 1993–2014 period. The figure also shows, however, that the lesser extreme 96.5th, 97th and 97.5th quantiles have actually increased slightly over the same period. This tightening of the upper tail can have an effect on the temporal dependence, but other factors are also likely contributing to extreme episodes becoming increasingly clustered. It is well established that in many polluted regions, high surface ozone levels correlate strongly with temperature; see, for example, Fiore et al. (2012) and references therein. It is also well established that heat waves are getting longer and more intense; see, for example, Dupuis (2012) and references therein. Figure 12 shows empirical estimates of the conditional probability that the daily maximum temperature at Heathrow Airport on day t exceeds a high threshold given that this threshold was exceeded on day $t - 1$. Data at Heathrow Airport are obtained from https://www7.ncdc.noaa.gov/CDO/cdo and are also deseasonalized and detrended following the procedure described in Section 6. Estimates of the conditional probability are trending upwards since 2005. This, in combination with the tightening of the upper tail seen in Figure 11, could explain the increased clustering of extremes. The formation and accumulation of air pollutants is also known to correlate strongly with local meteorological variables such as wind; see, for example, Fiore et al. (2012) and references therein. Figure 13 shows the distribution
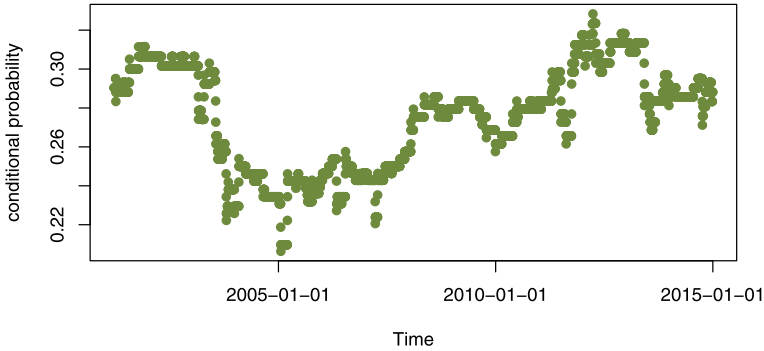
FIG. 12. *Exploring serial dependence in the temperature extremes. Empirical estimates of the conditional probability that the temperature at Heathrow Airport on day t exceeds a high threshold given that this threshold was exceeded on day t − 1. Estimate for day t is based on detrended and deseasonalized daily maximum temperature at Heathrow Airport from day t − 3000 to day t. Threshold is set to the 98th quantile of the detrended and deseasonalized daily maximum temperatures over the 1993–2014 time period.*

of hourly wind speed and direction at Heathrow Airport, along with the maximum daily maximum of the 8-hour running mean of ground-level ozone concentrations in Bloomsbury at these winds, over the period studied. Wind data are obtained from https://www7.ncdc.noaa.gov/CDO/cdo and observations at 0000, 0100, ..., 2300 hours are retained. While there were fewer strong southwesterly winds over the 2009–2014 period than previously, this reduction does not seem to have adversely affected ozone concentrations. Easterly winds are not more frequent over
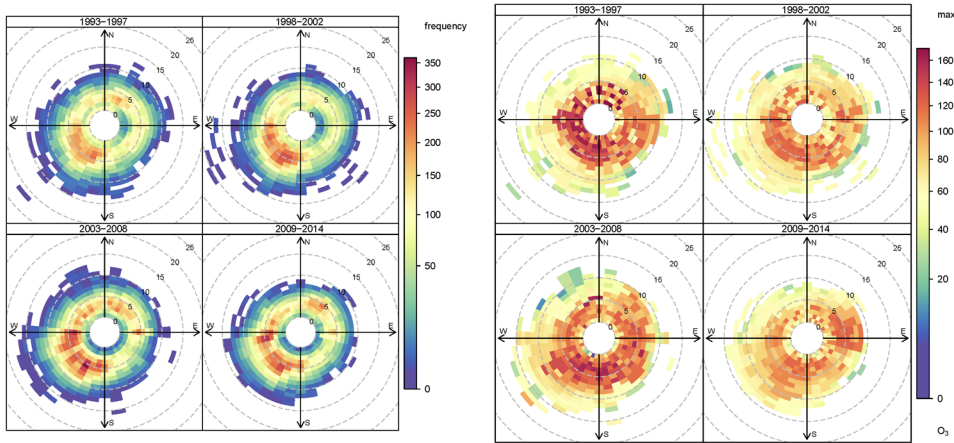

FIG. 13. *Hourly winds. Frequency of hourly wind speed and direction at Heathrow Airport (left) and maximum daily maximum of the 8-hour running mean of ground-level ozone concentrations in Bloomsbury (right) at these winds from January 1993 to December 2014. Wind speeds are in meters per second. Ozone concentration is measured in units of micrograms per cubic metre.*

the 2009–2014 period than previously but are now resulting in the largest ozone concentrations. This is consistent with the work of Pope et al. (2016) who show that anticyclonic conditions and easterly flows significantly enhance ozone concentrations over the UK relative to summer-time average values. While the intensity of the largest ozone concentrations under easterly winds has not increased from the 2003–2008 period, these concentrations are now the most extreme, and their temporal dependence could differ from that of the extreme concentrations over the 1993–2008 period which were mostly occurring under southwest winds.

Jacob and Winner (2009) observe correlation between surface ozone and temperature in polluted regions and establish the detrimental effect of warming, however the analysis is restricted to marginal considerations. Studies on current and future impacts of increasing heat waves on ozone episodes, for example, Schell and Prather (2017) and Shen, Mickley and Gilleland (2016), also only examine intensity. Our findings suggest that changes to the clustering and persistence of extreme ozone levels should be studied. Epidemiological studies that compare ozone-exposure related health problems before and after our estimated change point (late 2008) could also be pursued. Finally, health guidelines for ozone are set assuming 8-hour exposure [WHO (2000)], however further consideration should be given to acceptable levels over longer exposure periods in light of the increased clustering in the last 10 years.

8. Future work. Despite its simple form, our approach is very general and can be extended along several lines. The Markov models of Bortot and Gaetan (2014) could be used for the dependence instead of our trawl-based models. We opted for the exponential-trawl model of Noven, Veraart and Gandy (2017b) as it offered a superior fit for the ozone data of interest. Trawl processes in their most general form are quite flexible and can be constructed to have any monotonically decreasing autocovariance function; see Section 2.4.1 of Noven, Veraart and Gandy (2015a). More general models involving more parameters to control the dependence are easily fitted using the same partial-sample estimator. For instance, one could consider a more general trawl set obtained by superimposing several exponential trawl sets. This approach would give even greater flexibility to capture large serial dependence at very high thresholds but with an asymptotically independent process. We opt for a simple exponential trawl set as it is easier to use and flexible enough for our data; see Figures 7 and 9. We leave these dependence model extensions for future research. Our approach could also be used to test for structural changes in the spatiotemporal correlation function. It would be interesting to see if there has been a structural change due to climate change, for example, in the extremal dependence among the hourly rainfalls analyzed in Huser and Davison (2014).

Finally, the identification of multiple structural changes could be pursued. The generalization could be investigated along two different paths. First, one could adapt the test of Andrews (1993) to a null assumption of no breaks against an

alternative of $k > 1$ breaks. Bai and Perron (1998) succeeded in the linear regression context where they exploited the analogy between the F-test and the Wald test to test for differences in the residual sum of squares of the model with k breaks and those of the model without breaks. In our setting this approach does not seem a likewise viable solution. Alternatively, one could try to develop a similar test using the likelihoods as a criterion instead of the residual sum of squares, but likelihood ratio tests for composite likelihoods do not have the standard limiting distribution, and calculations might quickly become cumbersome. A third more promising strategy might be a sequential procedure. Let m be the number of change points, and let $\pi = \{\pi_1, \ldots, \pi_m\} \in (0, 1)^m$ be their location in the unit interval.[3] An iterative algorithm may be obtained as follows:

(i) Compute the test statistic W_n over the entire sample, $t = 1, \ldots, n$.

(ii) If the null hypothesis is rejected at the level α_1, record the location of the break: $\widehat{t} = n\pi : \sup_{\pi \in \Pi} W_n > \mathcal{B}_{\alpha_1}$ where $\mathcal{B}_{\alpha_1}$ is the $(1 - \alpha_1)$ quantile of the limiting distribution of the test statistic. Otherwise the algorithm stops.

(iii) The algorithm proceeds iteratively. Let $\widehat{t_1}, \ldots, \widehat{t_k}$ be the locations of the breaks collected up to the $(i - 1)$th iteration. Let $\{[1, \widehat{t_1}], \ldots, [\widehat{t_{k-1}} + 1, \widehat{t_k}], [\widehat{t_k} + 1, n]\}$ be the $k + 1$ time intervals considered at the ith iteration, and compute the test statistics $W_{t_j - t_{j-1}}$ over the corresponding subsamples, $j = 1, \ldots, k + 1$, where we use the notation $t_0 = 0$ and $t_{k+1} = n$ for convenience.

(iv) Repeat steps (ii)–(iii) until no further rejections occur. At each iteration the level α_i of the tests needs to be adjusted to account for the reduced sample size and the multiple $k + 1$ tests. In general α_i needs to decrease toward zero at an appropriate rate to keep the test consistent.

This kind of procedure generally yields desirable results, but its effective power strongly depends on the sample size and the distance between the change points. See Dupuis, Sun and Wang (2015) for a large simulation study in the context of change-points in the tail index. We leave an analogous investigation of the sequential approach to the context herein for future work.

APPENDIX

Definition of a homogeneous Lévy basis. Let S be a Borel set in $\mathbb{R}^2$ with associated Borel σ-algebra $\mathcal{S} = \mathcal{B}(S)$ and Lebesgue measure λ^{Leb}. Let $\mathcal{B}_b(S)$ be the subsets of S with finite Lebesgue measure, that is, $\mathcal{B}_b(S) = \{A \in \mathcal{S} : \lambda^{Leb}(A) < \infty\}$.

---

[3]Similarly to the one change-point case, the search is effectively carried out over $(\bar{\pi}, 1 - \bar{\pi})$ where $\bar{\pi}$ depends on the threshold. The threshold also has an impact on the number m of change points that can be identified as evaluation of test statistics requires a minimum number of observations in each term like those in (4).

Definitions [*Noven, Veraart and Gandy* (2017)].

1. A *random measure on* $(S, \mathcal{B}(S))$ is a collection of $\mathbb{R}$-valued random variables $\{M(A) : A \in \mathcal{B}_b(S)\}$ such that for any sequence $A_1, A_2, \ldots$ of disjoint elements of $\mathcal{B}_b(S)$ with $\bigcup_{j=1}^{\infty} A_j \in \mathcal{B}_b(S)$, we have $M(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} M(A_j)$ a.s.

2. A random measure M on $(S, \mathcal{S})$ is *independently scattered* if for any sequence $A_1, A_2, \ldots$ of disjoint elements of $\mathcal{B}_b(S)$, the random variables $M(A_1)$, $M(A_2), \ldots$ are independent.

3. A random measure M on $(S, \mathcal{S})$ is called *infinitely divisible* if for each $n \in \mathbb{N}$ there exist n independent, identically distributed random measures $Z_1^n, \ldots, Z_n^n$ such that $M \overset{d}{=} Z_1^n + \cdots + Z_n^n$. In particular, infinite divisibility implies that for any finite collection $A_1, \ldots, A_n$ of elements of $\mathcal{B}_b(S)$, the random vector $(M(A_1), \ldots, M(A_n))$ is infinitely divisible in $\mathbb{R}^n$.

4. A random measure on $(S, \mathcal{S})$ is called *stationary* if for any point $s \in S$ and finite collection $A_1, A_2, \ldots, A_n \in \mathcal{B}_b(S)$ such that $A_i + s \subset S$, we have that $(M(A_1 + s), M(A_2 + s), \ldots, M(A_n + s)) \overset{d}{=} (M(A_1), M(A_2), \ldots, M(A_n))$.

5. A *homogeneous Lévy basis L on* $(S, \mathcal{S})$ is a random measure that is independently scattered, infinitely divisible and stationary.

Consistent estimator of $\tilde{\sigma}_\rho$. We use

(5) $$\widehat{\tilde{\sigma}}_\rho = \widehat{\Gamma}_\rho^{-1} \widehat{\tilde{\Sigma}}_\rho \widehat{\Gamma}_\rho^{-1'} + \widehat{\Gamma}_\rho^{-1} \widehat{\Gamma}_{\nu,\xi}^{-1} \widehat{\tilde{\Sigma}}_{\nu,\xi} \widehat{\Gamma}_{\nu,\xi}^{-1'} \widehat{\Gamma}_\rho^{-1'},$$

where $\widehat{\Gamma}_\rho^{-1} = \frac{1}{n} \sum_{i=1}^{n} \nabla_\rho s_{\mathrm{PL}}^\Delta(x_i | \widehat{\nu}, \widehat{\xi}, \widehat{\rho})$, $\widehat{\Gamma}_{\nu,\xi}^{-1} = \frac{1}{n} \sum_{i=1}^{n} [\nabla_{\nu,\xi} s_{\mathrm{PL}}^\Delta(x_i | \widehat{\nu}, \widehat{\xi}, \widehat{\rho})]'$, $\widehat{\tilde{\Sigma}}_\rho = \frac{1}{n} \sum_{i=1}^{n} s_{\mathrm{PL}}^\Delta(x_i | \widehat{\nu}, \widehat{\xi}, \widehat{\rho}) s_{\mathrm{PL}}^{\Delta'}(x_i | \widehat{\nu}, \widehat{\xi}, \widehat{\rho})$ and $\widehat{\tilde{\Sigma}}_{\nu,\xi}$ is the covariance matrix of the ML estimator for the GP distribution obtained from the first step of the estimation.

Asymptotic power. Local alternatives are obtained substituting the assumption that the expected value of the score of the partial-sample estimator at the optimum must be equal to zero, $\mathbb{E}(s_{\mathrm{PS}}^\Delta(\pi, \rho_0, \widehat{\nu}, \widehat{\xi})) = 0$, with $\sup_{\pi \in \Pi} \| s_{\mathrm{PS}}^\Delta(\pi, \rho_0, \widehat{\nu}, \widehat{\xi}) - \mu(\pi) \| = o_p(1)$, where $\mu(\pi) = (\mu_1(\pi), \mu_2(\pi))' \in \mathbb{R}^2$ is a nonrandom function on Π. See Assumption 1-LP in Section 5.4 of Andrews (1993) for details. In short, $\mu(\pi)$ acts as a displacement factor on the value of the objective function at the optimum. The greater the value of $\mu(\pi)$ the further the alternative hypothesis is from the null. Then, from Theorem 4 of Andrews (1993), we have that under the alternative hypothesis,

$$\sup_{\pi \in \Pi} W_n(\pi) \overset{d}{\to} \sup_{\pi \in \Pi} \left(\frac{D(\pi)}{\sqrt{\pi(1-\pi)}} - S^{-1/2} \left[\mu_1(\pi) \left(\frac{1-\pi}{\pi} \right)^{1/2} - \mu_2(\pi) \left(\frac{\pi}{1-\pi} \right)^{1/2} \right] \right)^2,$$

where S is the asymptotic variance of the score of the pairwise likelihood estimator $\widehat{\rho}_{\mathrm{MPL}}^{\Delta}$. From Equation (3), we have that $S = e_1' E[(s_{\mathrm{PL}}^{\Delta} - \Gamma_{\nu,\xi}^{-1} \Upsilon^{-1} s_{\mathrm{GP}})(s_{\mathrm{PL}}^{\Delta} - \Gamma_{\nu,\xi}^{-1} \Upsilon^{-1} s_{\mathrm{GP}})'] e_1$ where e_1 is the unitary vector. We can thus easily obtain asymptotic power by simulation.

REFERENCES

ACKERMANN, R., HUGHES, G., HANRAHAN, D., SOMANI, A., AGGARWAL, S., FITZGERALD, A., DIXON, J., KUNTE, A., LOVEI, M. and LVOVSKY, K. (1999). *Pollution Prevention and Abatement Handbook* 1998: *Toward Cleaner Production*. World Bank Group, Washington, DC. Available at http://documents.worldbank.org/curated/en/758631468314701365/Pollution-prevention-and-abatement-handbook-1998-toward-cleaner-production.

AIR QUALITY EXPERT GROUP (2009). Ozone in the United Kingdom. Available at http://www.defra.gov.uk/environment/airquality/aqeg.

ANDREWS, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678

BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121

BARNDORFF-NIELSEN, O. E., BENTH, F. E. and VERAART, A. E. D. (2011). Ambit processes and stochastic partial differential equations. In *Advanced Mathematical Methods for Finance* 35–74. Springer, Heidelberg. MR2752540

BORTOT, P. and GAETAN, C. (2014). A latent process model for temporal extremes. *Scand. J. Stat.* **41** 606–621. MR3249419

BORTOT, P. and GAETAN, C. (2016). Latent process modelling of threshold exceedances in hourly rainfall series. *J. Agric. Biol. Environ. Stat.* **21** 531–547. MR3542085

COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics*. Springer, London. MR1932132

COX, D. R. and REID, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91** 729–737. MR2090633

DAVIS, R. A. and MIKOSCH, T. (2009). The extremogram: A correlogram for extreme events. *Bernoulli* **15** 977–1009. MR2597580

DAVIS, R. A. and YAU, C. Y. (2011). Comments on pairwise likelihood in time series models. *Statist. Sinica* **21** 255–277. MR2796862

DIERCKX, G. and TEUGELS, J. L. (2010). Change point analysis of extreme values. *Environmetrics* **21** 661–686. MR2838438

DUPUIS, D. J. (2005). Ozone concentrations: A robust analysis of multivariate extremes. *Technometrics* **47** 191–201. MR2188080

DUPUIS, D. J. (2012). Modeling waves of extreme temperature: The changing tails of four cities. *J. Amer. Statist. Assoc.* **107** 24–39. MR2949339

DUPUIS, D. J., SUN, Y. and WANG, H. J. (2015). Detecting change-points in extremes. *Stat. Interface* **8** 19–31. MR3320386

EASTOE, E. F. and TAWN, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **58** 25–45. MR2662232

ENVIRONMENTAL RESEARCH GROUP, KING'S COLLEGE LONDON (2015). *London Air Quality Network Database*. King's College, London. Available at http://www.londonair.org.uk

FERRO, C. A. T. and SEGERS, J. (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 545–556. MR1983763

FIORE, A. M., NAIK, V., SPRACKLEN, D. V., STEINER, A., UNGER, N., PRATHER, M., BERGMANN, D., CAMERON-SMITH, P. J., CIONNI, I. et al. (2012). Global air quality and climate. *Chem. Soc. Rev.* **41** 6663–6683.

GILLELAND, E. and KATZ, R. W. (2016). extRemes 2.0: An extreme value analysis package in R. *J. Stat. Softw.* **72** 1–39.

HUSER, R. and DAVISON, A. C. (2014). Space-time modelling of extreme events. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 439–461. MR3164873

JACOB, D. J. and WINNER, D. A. (2009). Effect of climate change on air quality. *Atmos. Environ.* **43** 51–63.

KIM, M. and LEE, S. (2009). Test for tail index change in stationary time series with Pareto-type marginal distribution. *Bernoulli* **15** 325–356. MR2543865

LEADBETTER, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **65** 291–306. MR0722133

LEDFORD, A. W. and TAWN, J. A. (2003). Diagnostics for dependence within time series extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 521–543. MR1983762

NEWEY, W. K. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics*, *Vol. IV. Handbooks in Econom.* **2** 2111–2245. North-Holland, Amsterdam. MR1315971

NEWEY, W. K. and WEST, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708. MR0890864

NOVEN, R. C., VERAART, A. E. and GANDY, A. (2015a). A latent trawl process model for extreme values. Available at http://arxiv.org/abs/1511.08190.

NOVEN, R. C., VERAART, A. E. and GANDY, A. (2017b). A latent trawl process model for extreme values. Available at http://arxiv.org/abs/1511.08190.

PICKANDS, J. III (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3** 119–131. MR0423667

POPE, R. J. et al. (2016). The impact of synoptic weather on UK surface ozone and implications for premature mortality. *Environ. Res. Lett.* **11** 124004.

QUINTOS, C., FAN, Z. and PHILLIPS, P. C. B. (2001). Structural change tests in tail behaviour and the Asian crisis. *Rev. Econ. Stud.* **68** 633–663. MR1855475

SCHELL, J. L. and PRATHER, M. J. (2017). Co-occurrence of extremes in surface ozone, particulate matter, and temperature over eastern North America. *Proc. Natl. Acad. Sci. USA* **114** 2854–2859.

SHEN, L., MICKLEY, L. J. and GILLELAND, E. (2016). Impact of increasing heat waves on US ozone episodes in the 2050s: Results from a multimodel analysis using extreme value theory. *Geophys. Res. Lett.* **43** 4017–4025.

SHEPHARD, N. and YANG, J. J. (2017). Continuous time analysis of fleeting discrete price moves. *J. Amer. Statist. Assoc.* **112** 1090–1106. MR3735362

SMITH, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72** 67–90. MR0790201

SMITH, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statist. Sci.* **4** 367–393. MR1041763

WHO (WORLD HEALTH ORGANIZATION) (1987). *Air Quality Guidelines for Europe*. WHO Regional Office for Europe, Copenhagen.

WHO (WORLD HEALTH ORGANIZATION) (2000). *Air Quality Guidelines for Europe*, Second ed. WHO Regional Office for Europe, Copenhagen.

WOLPERT, R. L. and ICKSTADT, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika* **85** 251–267. MR1649114

DEPARTMENT OF DECISION SCIENCES
HEC MONTRÉAL
3000, CHEMIN DE LA CÔTE-SAINTE-CATHERINE
MONTRÉAL, QUÉBEC
CANADA H3T 2A7
E-MAIL: debbie.dupuis@hec.ca

DEPARTMENT OF MATHEMATICAL SCIENCES
MATHEMATICAL FINANCE AND ECONOMETRICS
UNIVERSITÀ CATTOLICA DEL SACRO CUORE
LARGO GEMELLI 1
20123 MILANO
ITALY
E-MAIL: luca.trapin@gmail.com