

Asymptotic theory for the inference of the latent trawl model for extreme values

Valentin Courgeau  | Almut E.D. Veraart 

Department of Mathematics, Imperial College London, London, UK

Correspondence

Valentin Courgeau, Department of Mathematics, Imperial College London, 180 Queen's Gate, London, UK.

Email:

valentin.courgeau15@imperial.ac.uk

Abstract

This article develops statistical inference methods and their asymptotic theory for the latent trawl model for extremes, which captures serial dependence in the time series of exceedances above a threshold. We review two methods based on pairwise likelihood and show that they underestimate the serial dependence in the extremes. We propose two generalized method of moments procedures based on auto-covariance matching to overcome this shortcoming. Out of those four inference approaches, two are single-stage strategies while the others have two stages, and we provide central limit theorems in the sense of weakly approaching sequences of distributions for all of them. This additional flexibility ensures good behavior between the estimators and estimates of the limiting distribution. In an empirical illustration using London air pollution data, we find that the two-stage auto-covariance matching scheme yields a high-quality inference. It comprises two interpretable steps and correctly captures the serial dependence structure of extremes while performing on par with other methods in terms of marginal fit.

KEYWORDS

central limit theorem, generalized method of moments, generalized Pareto distribution, pairwise likelihood estimation, trawl process

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Scandinavian Journal of Statistics* published by John Wiley & Sons Ltd on behalf of The Board of the Foundation of the Scandinavian Journal of Statistics.

1 | INTRODUCTION

Modeling and estimating serial dependence in extremes is of great importance in many applications, including the environmental science, hydrology, oceanography, and financial risk management. In this article, we consider the framework of extremes as values above a high threshold—the peaks-over-threshold methodology. Under standard assumptions and necessary rescaling, the Pickands–Balkema–de Haan theorem (Balkema & de Haan, 1974; Davison & Smith, 1990; Pickands III, 1975) yields that the extreme values behave asymptotically as a Generalized Pareto Distribution (GPD) as the threshold gets larger. We denote the GPD cumulative distribution function (CDF) by $F_{\text{GPD}(\xi, \sigma)}$ for a shape parameter $\xi \in \mathbb{R}$ and a scale parameter $\sigma > 0$.

Distributional properties of extremes and standard inference approaches in the i.i.d. setting have been extensively studied (Davison & Smith, 1990; Smith, 1985). Somewhat recent developments allowing for general serial dependencies (Francq & Zakořan, 2013), involving Markov chains (Bortot & Gaetan, 2014; Bortot & Gaetan, 2016), or parametrizable latent serial dependence (Dupuis & Trapin, 2019; Noven et al., 2018) have rendered those approaches more flexible and reliable for applications, notably through a mixture representation of the GPD as an exponential distribution with a Gamma-distributed process as the rate (Hogg & Klugman, 1983; Smith & Miller, 1986).

In this article, we will focus on the latent trawl model for extremes (LTME) introduced by Noven et al. (2018), which is also based on the mixture representation of the GPD. By having a time-varying latent rate process, here modeled as Gamma-distributed trawl process (Barndorff-Nielsen, 2011), serial dependence arises in the extremes. While classical mixture-based approaches are restricted to a positive shape parameter ξ in the GPD, our model is designed to cover the general case when $\xi \in \mathbb{R}$.

So far, two inference approaches have been considered for the LTME: Noven et al. (2018) used a (single-stage) pairwise likelihood (PL) method (M1), whereas Dupuis and Trapin (2019) proposed a three-stage approach consisting of first estimating the exceedance rate parameter, then the marginal GPD parameters, followed by the parameters associated with the trawl process using PL. We will discuss a two-stage variant of this estimation method, where the exceedance rate and the GPD parameters are estimated jointly, followed by the estimation of the trawl parameters using PL (M2). However, the corresponding asymptotic theory for these methods was still missing, and our article fills this gap in the literature. Moreover, in our numerical and empirical work, we identified a shortcoming of the existing PL-based inference procedures resulting in an underestimation of the serial dependence in the extremes. In order to tackle this problem, we also propose two alternative inference procedures in this article: a single-stage (M3) and a two-stage (M4) generalized-method-of-moments-based approach which combines marginal likelihood maximization with auto-covariance matching (AM) of the extremes. By design, those methods capture the second-order properties of the data. We establish a suitable asymptotic theory for M3 and M4 and prove the asymptotic normality of the corresponding estimators. Adopting a two-stage approach provides interpretable steps with a comparable inference quality to the single-stage formulations while also maintaining similar theoretical guarantees. This is particularly essential for risk-oriented applications permitted by the extreme value theory.

In terms of theoretical contributions, we prove that the exceedance process in the LTME is strongly mixing. Moreover, we provide central limit theorems for each of the *four* inference approaches. The single-stage methods exploit *weakly approaching sequences of distributions* (Belyaev & Luna, 2000) and a central limit theorem for strongly mixing sequences

(Ekström, 2014), whereas the two-stage approaches are considered as two-stage generalized method of moments (GMM) (Newey & McFadden, 1994, Section 6) for clustered data (Hwang, 2021).

In an empirical illustration, we verify the above-mentioned claims on a dataset consisting of six air pollutants concentration hourly time series in London, United Kingdom. With the *exponential trawl* as a latent process, we show the superiority of the AM approach in estimating the dependence structure as well as the equivalence of the single-stage and two-stage procedures on the marginal fit.

The remainder of this article is structured as follows. We recall the definition of the marginally transformed model from Noven et al. (2018) in Section 2 with a primer on the strong mixing property and an overview of trawl processes. In Section 3, we describe the four inference methods mentioned above in detail. Section 4 comprises the asymptotic convergence analysis of the four inference schemes for a strongly mixing exceedance process with the same square-root convergence rate. We present a short overview of why we cannot simulate from the model in Section 5. Finally, Section 6 showcases the application of the inference approaches on six air pollutants concentration dataset. We present the qualitative comparison of the four schemes on the marginal fit, the capacity to capture the second-order dependence structure, and the quantification of the estimation uncertainty across approaches. We highlight some empirical evidence showing that the two-stage auto-covariance method (M4) is the best-performing and most interpretable approach available and, hence, should be preferred.

2 | THE LATENT TRAWL MODEL FOR EXTREMES

Let us recall the LTMEs, first introduced in Noven et al. (2018), where we focus on the marginally transformed formulation in Section 2.4 to allow for general shape parameter $\xi \in \mathbb{R}$ in the GPD.

2.1 | The peaks-over-threshold methodology

Let $(t_j, j \in \mathbb{N})$ be a sequence of increasing observation times and consider the time series $Y = (Y_{t_j}, j \in \mathbb{N})$ which we denote by $Y = (Y_j, j \in \mathbb{N})$ for simplicity. We define the exceedance process corresponding to an original time series and a threshold as follows:

Definition 1. Let $Y := (Y_j, j \in \mathbb{N})$ be a strictly stationary time series. We define the exceedance process $X^u := (X_j^u, j \in \mathbb{N})$ of Y with respect to a threshold $u \in \mathbb{R}$ by $X_j^u := \max(Y_j - u, 0)$, for all $j \in \mathbb{N}$.

The *peaks-over-threshold* methodology provides an asymptotic distributional framework for the exceedance process above a high threshold u , which we review briefly.

Notation 1. The GPD with, respectively, shape and scale parameters $(\xi, \sigma) \in \mathbb{R} \times (0, \infty)$ has a probability density function given by

$$f_{\text{GPD}(\xi, \sigma)}(x) = \begin{cases} \sigma^{-1}(1 + \xi x/\sigma)^{-1/\xi-1}, & \text{if } \xi \neq 0, \\ \sigma^{-1} \exp(-x/\sigma), & \text{if } \xi = 0, \end{cases}$$

with support $(0, x_F)$ where $x_F = \infty$ if $\xi \geq 0$ and $x_F = -\sigma/\xi$ if $\xi < 0$. Note that x_F is also called the right endpoint of the distribution. The CDF is given by $F_{\text{GPD}(\xi, \sigma)}(x) = 1 - (1 + \xi x/\sigma)^{-1/\xi}$ if $\xi \neq 0$ and $F_{\text{GPD}(\xi, \sigma)}(x) = 1 - \exp(-x/\sigma)$ if $\xi = 0$.

We denote the *conditional excess distribution function defined with respect to a threshold u* as:

$$F_u(x) := P(X_t^u \leq x | X_t^u > 0) = \frac{F(u+x) - F(u)}{1 - F(u)},$$

for $0 \leq x \leq x_F - u$ where $x_F \in (0, +\infty]$ is the right endpoint of the distribution F and the time $t \in \mathbb{N}$ is removed from F_u by stationarity. Let G be an extremal distribution function, that is, Fréchet, Weibull or Gumbel (Gnedenko, 1943; Gumbel, 2004). If $x \mapsto F(x)$ is continuous, F belongs to the domain of attraction of some distribution G if and only if

$$\lim_{u \rightarrow x_F} \inf_{0 < \sigma < \infty} \sup_{0 < x < x_F - u} |F_u(x) - F_{\text{GPD}(\xi, \sigma)}(x)| = 0,$$

for some $\xi \in \mathbb{R}$ (Pickands III, 1975, theorem 7). That is, for a large fixed u , one may approximate F_u by $F_{\text{GPD}(\xi, \sigma)}$ for some $\xi \in \mathbb{R}, \sigma > 0$. Formally, F belongs to the domain of attraction of G if there are two sequences $(a_n, n \in \mathbb{N}) \subset (0, +\infty)$ and $(b_n, n \in \mathbb{N}) \subset \mathbb{R}$, such that $P((X_{(n)} - b_n)/a_n \leq x) \xrightarrow{n \rightarrow +\infty} G(x)$ for all $x \in \mathbb{R}$.

2.2 | Model definition

In the following model description, we consider a time series Y with a continuous marginal CDF which belongs to the domain of attraction of an *extremal distribution function*. We will then fix a high threshold u , typically a high quantile value (Bader et al., 2018), and assume that the GPD approximation to the CDF of the corresponding exceedance process X^u holds exactly. We then set $\sigma := \sigma(u)$ and $F_u = F_{\text{GPD}(\xi, \sigma)}$. We also set $X := X^u$ for ease of exposition.

2.2.1 | Hierarchical model construction

Consider a stochastic process $Z := (Z_j, j \in \mathbb{N})$ taking values in $[0, \infty)$ and let $\Psi := (\Psi_j, j \in \mathbb{N})$ be a strictly stationary latent process such that

$$\Psi_j \sim \text{Gamma}(1, 1) = \text{Exp}(1), \quad \text{for } j \in \mathbb{N}.$$

Suppose that, conditional on Ψ , the random variables Z_j are independent with

$$Z_j | \{Z_j > 0, \Psi_j\} \sim \text{Exp}(\Psi_j),$$

and, conditional on Ψ , the events $\{Z_j > 0\}$ are independent, and for $\kappa > 0$,

$$P(Z_j > 0 | \Psi_j) = \exp(-\kappa \Psi_j).$$

Then

$$Z_j | \{Z_j > 0\} \sim \text{GPD}(1, 1 + \kappa), \quad \text{and} \quad P(Z_j > 0) = (1 + \kappa)^{-1}.$$

Using the probability integral transform, we define a suitable transformation function as follows: Let $\xi \in \mathbb{R}$ and $\sigma > 0$ and define

$$h(z) := F_{\text{GPD}(\xi, \sigma)}^{-1}(F_{\text{GPD}(1, 1 + \kappa)}(z)) = \frac{\sigma}{\xi} \left[\left(1 + \frac{z}{1 + \kappa} \right)^{\xi} - 1 \right].$$

We model the exceedance process X by setting $\{X_j = 0\} = \{Z_j = 0\}$, and if $X_j > 0$, we write $X_j = h(Z_j)$.

From this definition, we have $X_j | X_j > 0 \sim \text{GPD}(\xi, \sigma)$ and $P(X_j > 0) = (1 + \kappa)^{-1}$, that is, κ controls the probability of having an exceedance and the parameters (ξ, σ) characterize the distribution of extremes. Also, given that Ψ is strictly stationary, X is also strictly stationary.

Using the transformation function h in the above construction allows us to go beyond the restrictive assumption of having a *positive* shape parameter $\xi > 0$ when using the mixture representation of the GPD, see Bortot and Gaetan (2016) for more details. We can summarize the model construction via the following diagram:

$$\Psi \sim \text{Gamma}(1, 1) \xrightarrow{\text{Exp}} \text{GPD}(1, 1 + \kappa) \xrightarrow{h(\cdot)} \text{GPD}(\xi, \sigma) \sim X | X > 0.$$

Note that we can express the conditional density of X_j given Ψ_j , with respect to the measure $\mu(dx_j) = \delta_0(dx_j) + dx_j$, as

$$f(x_j | \psi_j) = \begin{cases} 1 - \exp(-\kappa \psi_j), & \text{if } x_j = 0, \\ J_h(x_j) \psi_j \exp\{-(\kappa + h^{-1}(x_j)) \psi_j\}, & \text{if } x_j > 0, \end{cases} \quad (1)$$

where the Jacobian scalar is given by

$$J_h(x) := \frac{f_{\text{GPD}(\xi, \sigma)}(x)}{f_{\text{GPD}(1, 1 + \kappa)}(h^{-1}(x))} = \frac{1 + \kappa}{\sigma} \left(1 + \frac{\xi x}{\sigma} \right)^{\xi - 1}.$$

Note that the inverse function h^{-1} is defined as

$$h^{-1}(x) = (1 + \kappa) \left[\left(1 + \frac{\xi x}{\sigma} \right)^{1/\xi} - 1 \right].$$

Notation 2. We denote by f_{PL} the PL of the model with $\text{GPD}(1, 1 + \kappa)$ marginals while the model likelihood with $\text{GPD}(\xi, \sigma)$ marginals is denoted by f .

We refer to X as the exceedance process and Ψ as the latent process.

2.2.2 | Dependence structure

We assume the independence of observations X_j conditional on the latent process Ψ_j . This shifts the dependence structure of the joint distribution $(X_1, \dots, X_l)$ given $(\Psi_1, \dots, \Psi_l)$ onto the latent process itself and it no longer relies on the original exceedance process. Formally, we suppose that

$$f(x_{a_1}, \dots, x_{a_l} | \psi_{a_1}, \dots, \psi_{a_l}) = \prod_{k=1}^l f(x_{a_k} | \psi_{a_k}), \quad (2)$$

for any integer $l > 0$ and $(a_j, j \in \{1, \dots, l\}) \subseteq \mathbb{Z}$, that is, any finite-dimensional density $(X_1, \dots, X_l)$ for $l \in \mathbb{N}$ of (X_j) depends on the joint distribution of the latent process $(\Psi_1, \dots, \Psi_l)$ and the conditional contributions of $X_i | \Psi_i$ for $i \in \{1, \dots, l\}$ which are all identically distributed.

Note that the marginal transformation $h(\cdot)$ does not impact the auto-covariance dependence structure. Only the interactions between the positive *transformed* values are taken into account and those can be transformed back to the GPD(1, 1 + κ) marginals without additional compensating term (e.g., see Appendix C).

2.2.3 | Mixing properties

We recall the definition of the *strong mixing coefficient*: for a sequence of random variables $Z := (Z_i, i \in \mathbb{Z})$ on a probability space $(\Omega, \mathcal{F}, P)$ and $n \leq m \in \mathbb{Z}$, denote by $\mathcal{F}_n^m(Z)$ the σ -algebra generated by $\{Z_i, n \leq i \leq m\}$ and define the strong mixing coefficient at time n with a time lag $k \in \mathbb{N}$ by

$$\alpha_n^*(Z, k) := \sup_{\substack{A \in \mathcal{F}_{-\infty}^n(Z) \\ B \in \mathcal{F}_{n+k}^{+\infty}(Z)}} |P(A \cap B) - P(A)P(B)|,$$

and the overall strong mixing coefficient by $\alpha^*(Z, k) := \sup_{n \in \mathbb{N}} \alpha_n^*(Z, k)$,¹ which is redundant by stationarity.

The dependence structure defined in Section 2.2.2 allows to connect the properties of the *unobserved* latent process to the ones of the exceedance process itself such as ergodicity as follows:

Proposition 1. *If Ψ is mixing then X is ergodic.*

Proof. We argue as follows: from Equation (1), one can characterize the distribution of $(X_j, j \in \mathbb{N})$ using $(\Psi_j, j \in \mathbb{N})$ and an additional independent sequence of i.i.d uniformly distributed random variables on $[0, 1]$, denoted (U_j) . Since $(\Psi_j, j \in \mathbb{N})$ is mixing and $(U_j, j \in \mathbb{N})$ is ergodic (and independent by definition), $((\Psi_j, U_j), j \in \mathbb{N})$ is also ergodic. Applying the measurable transformation $(\Psi_j, U_j) \mapsto X_j$ using Eq. 1 for each $j \in \mathbb{N}$ yields an ergodic sequence which concludes the proof. ■

As suggested above, this framework is defined for any strictly stationary Gamma-distributed latent process. In this work, we focus on the particular case of trawl processes whose definition we recall below. Note that trawl processes are a particular case of Mixed Moving-Average processes (Fuchs & Stelzer, 2013) hence mixing.

2.3 | Latent trawl process

Although any strictly stationary Gamma-distributed process can be taken as the latent process, we use trawl processes (Barndorff-Nielsen, 2011) since they allow for great flexibility in modeling serial dependence.

Notation 3. Suppose $p \in \mathbb{N}$ and let $S \subseteq \mathbb{R}^p$ be a Borel set and we denote $\mathcal{B}(S)$ the associated Borel σ -algebra. Consider the set of bounded borel sets with respect to the Lebesgue measure μ^{leb} , that is, $\mathcal{B}_b(S) := \{A \in \mathcal{B}(S) : \mu^{\text{leb}}(A) < \infty\}$. We denote trawl processes by Λ as opposed to a general latent process denoted by Ψ .

2.3.1 | A characterization of trawls

Trawl processes are a class of stationary and infinitely divisible stochastic processes introduced in Barndorff-Nielsen (2011). For the definitions of random measures and Lévy basis as well as the independent scattering and infinite divisibility properties, see Appendix A.

Notation 4. For a set $A \subseteq \mathbb{R}^d$ where $d \in \mathbb{N}$, the indicator function $\mathbb{I}_A : \mathbb{R}^d \rightarrow \{0, 1\}$ is defined such that $\mathbb{I}_A(\mathbf{x}) = 1$ if $\mathbf{x} \in A$ and $\mathbb{I}_A(\mathbf{x}) = 0$ otherwise.

For $u \in \mathbb{R}$, we denote the characteristic function by $\Phi(u; L(A)) := \mathbb{E}[\exp(iuL(A))]$. Let L be an homogeneous Lévy basis on $(S, \mathcal{B}(S))$. Then, the characteristic function can be written as

$$\Phi(u; L(A)) = \exp\{\mu^{\text{leb}}(A)\mathcal{K}(u)\}, \text{ for any } A \in \mathcal{B}_b(S),$$

where $u \mapsto \mathcal{K}(u)$ is the corresponding cumulant generating function that satisfies

$$\mathcal{K}(u) = iu\mu - \frac{1}{2}u^2\eta^2 + \int_{\mathbb{R}} (e^{iux} - 1 - iu\mathbb{I}_{\{|y| \leq 1\}}(x)) \nu(dx), \quad (3)$$

for some constants $\mu, \eta^2 \geq 0$ and ν a Lévy measure (Rajput & Rosinski, 1989, proposition 2.4). Since L is infinitely divisible, there is always a corresponding random variable L' —the Lévy seed—whose law is characterized by (3), such that

$$\Phi(u; L(A)) = \exp\{\mu^{\text{leb}}(A)C(u; L')\}, \quad \forall A \in \mathcal{B}_b(S),$$

where we write $C(u; L') = \mathcal{K}(u)$ for the L' cumulant function (Barndorff-Nielsen, 2011).

Summary 1. The law of $L(A)$ is therefore characterized by the Lebesgue measure of the set A , $\mu^{\text{leb}}(A)$, and the law of the associated Lévy seed L' , for any fixed $A \in \mathcal{B}_b(S)$. In practice, we use this characterization and choose a set A and a law for the Lévy seed L' to define a trawl process.

2.3.2 | Trawl-related definitions and core properties

Barndorff-Nielsen (2011, section 7.3) and Barndorff-Nielsen et al. (2014) provide the foundations of trawl processes.

Given Remark 1, the set A which partially characterizes the trawl process (along with the Lévy seed L') is called a *trawl set* and defined as follows:

Definition 2 (Trawl set, definition 5 of Barndorff-Nielsen et al. (2014)). A trawl set is a Borel set $A \subset [0, \infty) \times (-\infty, 0]$ such that $\mu^{\text{leb}}(A) < \infty$. Then, we set $A_t := A + (0, t)$, which gives the position of the trawl at time t and such that $\mu^{\text{leb}}(A_t) = \mu^{\text{leb}}(A)$ for all $t \in \mathbb{R}$.

The trawl set A is given using an intermediate function, called the *trawl function*, defined as follows:

Definition 3 (Trawl function, Barndorff-Nielsen et al. (2014)). A trawl function is a continuous map $g : (-\infty, 0] \rightarrow [0, +\infty)$ such that

$$A := \{(x, s) : 0 \leq x \leq g(s), s \leq 0\} \subset \mathbb{R}^2 \quad \text{is a trawl set.}$$

Then, one has

$$A_t := A + (0, t) = \{(x, s) : 0 \leq x \leq g(s - t), s \leq t\}, \quad t \in \mathbb{R}.$$

In particular, we say that a trawl set is *monotonic* if the underlying trawl function g is itself monotonic.

Note that one often identifies A with the trawl set at time 0, namely A_0 . We parametrize the trawl function with a *trawl parameter* $\rho \in \mathbb{R}^p$ for some $p \in \mathbb{N}$ such that $A_t(\rho) = \{(x, s) : 0 \leq x \leq g(s - t; \rho), s \leq t\}$. For brevity, we write $A_t = A_t(\rho)$.

Examples of trawl functions include the *exponential function*, that is, $g(u) = \exp(\rho u)$ for some fixed $\rho > 0$ (Dupuis & Trapin, 2019; Noven et al., 2018), or the *weighted sum of exponentials* such that $g(u) = \sum_{i=1}^d w_i \exp(\rho_i u)$ for some integer $d \geq 2$, $\rho_1 > \dots > \rho_d > 0$ and nonnegative weights such that $\sum_{i=1}^d w_i = 1$ (Barndorff-Nielsen et al., 2014, example 5). See Barndorff-Nielsen et al. (2014) for additional trawl sets. We now give the definition of trawl processes:

Definition 4 (Trawl processes, section 7.3 of Barndorff-Nielsen (2011)). Let $A \in \mathcal{B}_b(\mathbb{R}^2)$ be a trawl set and consider the collection of sets $(A_t, t \in \mathbb{R})$ defined as $A_t := A + (0, t) = \{(x, s) : (x, s - t) \in A\}$ for any $t \in \mathbb{R}$. Let L be an homogeneous Lévy basis on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. The trawl process $(\Lambda_t, t \in \mathbb{R})$ associated with the trawl set A with respect to the Lévy basis L is defined by setting $\Lambda_t := L(A_t)$ for $t \in \mathbb{R}$.

Barndorff-Nielsen (2011) also showed that trawl processes and ID laws are linked to one another in the following sense:

Proposition 2. *For any infinitely divisible law, one can find a trawl process with the same marginal law (but potentially featuring some serial dependence).*

From the definition, it is clear that we have the following stochastic integral formulation:

$$\Lambda_t = \int_{\mathbb{R}^2} \mathbb{I}_{A_t}(x, s) L(dx, ds) = \int_{\mathbb{R}^2} \mathbb{I}_A(x, s - t) L(dx, ds),$$

demonstrating that we have a mixed moving average representation. Therefore, one has $\text{Cor}(\Lambda_t, \Lambda_{t+h}) = \mu^{\text{leb}}(A \cap A_h) / \mu^{\text{leb}}(A)$. In a framework similar to Section 2.2 and using (2), one obtains in general for $\text{Gamma}(\alpha, \beta)$ -distributed trawls that

$$\mathbb{E}(X_0 X_h) = \int_{\kappa}^{\infty} \int_{\kappa}^{\infty} (1 + u_0/\beta)^{b_{0 \setminus h}} (1 + (u_0 + u_h)/\beta)^{b_{0, h}} (1 + u_h/\beta)^{b_{h \setminus 0}} du_0 du_h, \quad (4)$$

where

$$\begin{cases} b_{h \setminus 0} = b_{0 \setminus h} := -\alpha(1 - \text{Cor}(\Lambda_0, \Lambda_h)) = -\alpha\mu^{\text{leb}}(A \setminus A_h)/\mu^{\text{leb}}(A), \\ b_{0,h} := -\alpha\text{Cor}(\Lambda_0, \Lambda_h) = -\alpha\mu^{\text{leb}}(A \cap A_h)/\mu^{\text{leb}}(A). \end{cases} \quad (5)$$

For instance, a $\text{Gamma}(1, 1)$ -distributed latent trawl process yields:

$$\mathbb{E}(X_0 X_h) = \int_{\kappa}^{\infty} \int_{\kappa}^{\infty} (1 + u_0)^{\text{Cor}(\Lambda_0, \Lambda_h)-1} (1 + u_0 + u_h)^{-\text{Cor}(\Lambda_0, \Lambda_h)} (1 + u_h)^{\text{Cor}(\Lambda_0, \Lambda_h)-1} du_0 du_h.$$

The $\text{Gamma}(\alpha, \beta)$ distribution has a probability density function given by $f(x; \alpha, \beta) = \Gamma(\alpha)^{-1} \beta^{\alpha} x^{\alpha-1} e^{-\beta x} \mathbb{I}_{(0, \infty)}(x)$ with $\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx$ for $z > 0$. We refer to α (resp. β) as the Gamma shape (resp. scale) parameter. Note that α impacts the integrability properties of the LTME process as hinted in the proofs of Section 4.

Remark 1. In this work, we consider the discretized version of the corresponding continuous-time processes $(X_t, t \in \mathbb{R})$. We write $(X_{t_j}, j \in \mathbb{Z})$, or more succinctly $(X_j, j \in \mathbb{Z})$ where the times (t_j) are increasing with $i \in \mathbb{Z}$ such that $\lim_{j \rightarrow \pm\infty} t_j = \pm\infty$.

It is known that trawl processes are ergodic and mixing (Fuchs & Stelzer, 2013). We prove that the LTME process X with a latent trawl process is strongly mixing:

Proposition 3. *If (Ψ_j) is a trawl process, say (Λ_j) , then the unconditional exceedance process (X_j) has a strong mixing coefficient dominated by the correlation of (Λ_j) , that is,*

$$\alpha^*(X, k) = O(\text{Cor}(\Lambda_0, \Lambda_k)).$$

If its trawl set satisfies $\mu^{\text{leb}}(A \cap A_k) \rightarrow 0$ as $k \rightarrow \infty$, then (X_j) is strongly mixing.

Proof. See Appendix C. ■

3 | INFERENCE

In this section, we introduce the four inference strategies of interest: for a target parameter vector denoted $\theta = (\xi, \sigma, \kappa, \rho)^T \in \mathbb{R}^{3+p}$, we either estimate the full parameter vector at once (Noven et al., 2018) or take a two-stage approach where one infers the marginal parameters (ξ, σ, κ) and then estimates the trawl parameter ρ (Dupuis & Trapin, 2019). We present two corresponding schemes based on the PL function first, and then introduce the novel AM methods formalized as a GMM approach as listed below:

- Method 1 (M1): Use the PL to estimate the full parameter vector in a single stage as in Noven et al. (2018).
- Method 2 (M2): Use the marginal likelihood to estimate the marginal parameters first, and proceed to a PL-based estimation of the trawl parameter in the second stage, similarly to Dupuis and Trapin (2019).
- Method 3 (M3): Use both the marginal likelihood and AM to estimate the full parameter vector in a single stage subdivided into two substeps.
- Method 4 (M4): Use the marginal likelihood to estimate the marginal parameters first, and proceed to an AM-based estimation of the trawl parameter in the second stage.

Note that Methods 2 and 4 differ slightly from the approach taken in Dupuis and Trapin (2019) when it comes to the estimation of the parameter κ . They take advantage of the fact that $X|X > 0 \sim \text{GPD}(\xi, \sigma)$ and estimate $P(X_t > 0) = (1 + \kappa)^{-1}$ empirically in a first step (out of three steps). On the other hand, we use the unconditional marginal likelihood that includes the likelihood of the event $\{X_t = 0\}$ to estimate κ along with ξ and σ . This minor change makes the scheme more theoretically consistent by adopting a likelihood-based optimization to the cost of a slight underestimation of $P(X_t > 0)$ (see Section 6.2).

3.1 | Notations

We denote the true parameter value by θ_0 . We also denote by $\nabla_{\{\cdot\}}$ (resp. $\nabla_{\{\cdot\}}^2$) the gradient (resp. the Hessian matrix) with respect to some variable(s) $\{\cdot\}$. Let G be a positive integer defining the number of clusters, or groups of pairs, and consider the triangular set $\mathcal{G} := \{(X_i, X_j) : 1 \leq i < j \leq N\}$. For $g \in \{1, \dots, G\}$, we define

$$\mathcal{G}_g := \{(X_i, X_{i+g}) : 1 \leq i \leq N - g\},$$

and $\mathcal{G}_{G+1} := \{(X_i, X_j) : j - i > G, 1 \leq i < j \leq N\}$ such that $\mathcal{G} = \cup_{g \in \{1, \dots, G+1\}} \mathcal{G}_g$ and $\mathcal{G}_g \cap \mathcal{G}_h = \emptyset$ for $g \neq h$. We denote by $N_g := N - g$ the cardinality of the cluster $\mathcal{G}_g$ for $g \in \{1, \dots, G\}$ and we have $N_{G+1} = (N - G)(N - G + 1)/2$. More generally, we use $|\cdot|$ to indicate the cardinality of a set, for example, $|\mathcal{G}_g| = N_g$.

Next, depending on the context, we denote by $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^d$, or on $\mathbb{R}^{d \times d}$, which is the reference metric for the convergence in probability of d -dimensional, or $d \times d$ -dimensional, sequences.

We introduce the PL approach (Dupuis & Trapin, 2019; Noven et al., 2018) for M1 and M2 and the novel AM scheme for M3 and M4. Since we may infer the marginal parameters separately, we first formalize the unconditional marginal likelihood in the following section.

3.2 | Unconditional marginal likelihood

Dupuis and Trapin (2019) leverage that $X|X > 0 \sim \text{GPD}(\xi, \sigma)$ and $P(X_t > 0) = (1 + \kappa)^{-1}$ to estimate the marginal parameters (ξ, σ) and κ , respectively, and then estimate the trawl parameters in a second stage (see Sections 3.3–3.6). We augment this scheme by harnessing the full unconditional marginal likelihood of X to estimate ξ , σ and κ at once. Note that by *unconditional* we refer to being unconditional on the latent process as opposed to the model definition. From (1), we define the Unconditional Likelihood (UL) as follows

$$f_{\text{UL}}(x; \xi, \sigma, \kappa) := (1 + \kappa)^{-1} \sigma^{-1} \left(1 + \frac{\xi x}{\sigma} \right)^{-1/\xi - 1} \mathbb{I}_{(0, x_F)}(x) + (1 - (1 + \kappa)^{-1}) \mathbb{I}_{\{0\}}(x),$$

where $(\xi, \sigma, \kappa) \in \mathbb{R} \times (0, \infty) \times (0, \infty)$ and we define

$$\text{UL}(\xi, \sigma, \kappa) = \text{UL}(\xi, \sigma, \kappa; X_1, \dots, X_N) := \sum_{i=1}^N \ln f_{\text{UL}}(X_i; \xi, \sigma, \kappa).$$

We denote by $(\hat{\xi}_{UL,N}, \hat{\sigma}_{UL,N}, \hat{\kappa}_{UL,N})$, or $(\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL})$ for brevity, the estimators given by UL maximization of the true values $(\xi_0, \sigma_0, \kappa_0) \in \mathbb{R} \times (0, \infty) \times (0, \infty)$.

3.3 | Method 1: single-stage PL estimator

Noven et al. (2018) applied the PL to infer the marginal and trawl parameters as a single-stage inference procedure. Suppose $\Theta \subset \mathbb{R} \times (0, \infty) \times (0, \infty) \times \mathbb{R}^p$ is a compact parameter subspace. We maximize the PL which is defined as follows:

$$PL_N(\theta; G) = G^{-1} \sum_{g=1}^G N_g^{-1} \sum_{i=1}^{N_g} \ln f(X_i, X_{i+g}; \theta), \quad (6)$$

for an integer $G \geq 2$ where $\theta = (\xi, \sigma, \kappa, \rho) \in \Theta$ and we call G the *block depth*.

We introduce Method 1 (M1), the single-stage PL-based inference strategy (Noven et al., 2018), studied in this article: the maximum PL likelihood estimator $\hat{\theta}_{M1}$ as follows

$$\hat{\theta}_{M1} := \arg \max_{\theta \in \Theta} PL_N(\theta; G).$$

3.4 | Method 2: two-stage PL estimator

In the spirit of Dupuis and Trapin (2019), we consider PL solely to estimate the trawl parameters ρ given some marginal estimates (ξ, σ, κ) . Method 2 (M2) is a two-stage inference scheme where ones first estimates (ξ, σ, κ) by UL maximization yielding the marginal estimates

$$(\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL}) = \arg \max_{(\xi, \sigma, \kappa)} UL(\xi, \sigma, \kappa),$$

and we obtain $\hat{\rho}_{M2}$ via PL maximization using the UL estimates, that is,

$$\hat{\rho}_{M2} := \arg \max_{\rho \in \mathbb{R}^p} PL(\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL}, \rho; G),$$

such that $\hat{\theta}_{M2} := (\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL}, \hat{\rho}_{M2})^\top$.

3.5 | Method 3: single-stage AM

Empirically, the PL-based approaches underestimate the serial dependence (Section 6.2). To circumvent this issue, we propose an AM-based procedure to estimate ρ and use the unconditional marginal distribution for the marginal parameters (ξ, σ, κ) .

3.5.1 | Generalized moments condition

M3 is a single-stage approach in the sense that we estimate all parameters at once. However, it is formulated as a two-step GMM approach on clustered data with a fixed number of clusters

(Hwang, 2021)—namely, the two-step centered GMM estimation. Additionally, define for any $g \in \{1, \dots, G\}$ and $i \in \{1, \dots, N_g\}$

$$r_i^g(\theta) := \left[\nabla_{(\xi, \sigma, \kappa)} (\ln f_{\text{UL}}(X_i) + \ln f_{\text{UL}}(X_{i+g}))^\top, X_i X_{i+g} - \mathbb{E}_\theta(X_0 X_g) \right]^\top, \quad (7)$$

as the moments condition for this GMM problem. This vector is defined as the score function with respect to the UL likelihood (which should have zero expectation under the assumption of a well-specified model and $\theta = \theta_0$). The average value over the G clusters is defined as

$$\bar{r}_N(\theta; G) := G^{-1} \sum_{g=1}^G N_g^{-1} \sum_{i=1}^{N_g} r_i^g(\theta).$$

3.5.2 | Procedure

Recall that $p \in \mathbb{N}$ is the dimension of the trawl parameter ρ . We first obtain the first-step estimate

$$\hat{\theta}_{\text{M3},1} := \arg \min_{\theta \in \Theta} \bar{r}_N(\theta)^\top W_N^{-1} \bar{r}_N(\theta),$$

where W_N is a real-valued $(3+p) \times (3+p)$ positive definite matrix such that $W_N \rightarrow W > 0$ in probability as $N \rightarrow \infty$. We only consider the identity matrix $W_N = \mathbf{I}_{3+p}$. Then, one proceeds to compute

$$\hat{\theta}_{\text{M3},2} := \arg \min_{\theta \in \Theta} \bar{r}_N(\theta)^\top \hat{\Omega}(\hat{\theta}_{\text{M3},1})^{-1} \bar{r}_N(\theta),$$

where the total variance estimator $\hat{\Omega}(\theta)$ is defined as

$$\hat{\Omega}(\theta) := G^{-1} \sum_{g=1}^G \left(N_g^{-1/2} \sum_{i=1}^{N_g} [r_i^g(\theta)] \right) \left(N_g^{-1/2} \sum_{i=1}^{N_g} [r_i^g(\theta)] \right)^\top. \quad (8)$$

As noted in Hwang (2021), the estimate defined in (8) can be recentered in the following way:

$$\hat{\Omega}^c(\theta) := G^{-1} \sum_{g=1}^G \left(N_g^{-1/2} \sum_{i=1}^{N_g} [r_i^g(\theta) - \bar{r}_N(\theta)] \right) \left(N_g^{-1/2} \sum_{i=1}^{N_g} [r_i^g(\theta) - \bar{r}_N(\theta)] \right)^\top.$$

Under some technical assumptions, the inference using our noncentered estimator leads to a Wald statistic that is not asymptotically pivotal as opposed to the centered one, meaning that it will depend on the estimate of the first stage. However, the recentering requires the homogeneity of clusters: that for $\Omega_g = \Omega$ and $\Gamma_g = \Gamma$ for any $g \in \{1, \dots, G\}$ which can be restrictive and does not hold in general in our framework.

Remark 2. This two-step procedure can be extended to an iterative one (Hansen & Lee, 2021) where one considers obtaining $\hat{\theta}_{s+1}$ from $\hat{\theta}_s$ and the updated weighting matrix $\hat{\Omega}(\hat{\theta}_s)$ sequentially for $s = 0, 1, \dots$ such the GMM estimate is $\lim_{s \rightarrow \infty} \hat{\theta}_s$. Note that the asymptotic theory for

clustered data is then explored in Hansen and Lee (2019). Given the differences in the assumptions and nature of the inference strategy, we leave the assessment of this approach for future research although looking very promising to remove the arbitrary choice of the first weighting matrix (Hansen & Lee, 2021, section 5).

3.6 | Method 4: two-stage AM

In Method 4 (M4), we define

$$\hat{\theta}_{M4} := (\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL}, \hat{\rho}_{M4})^\top,$$

where one first estimates the marginal parameters by UL maximization $(\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL})$ and then, given those estimates, infers the trawl parameter as follows

$$\hat{\rho}_{M4} := \arg \min_{\rho \in \mathbb{R}^p} \left(\sum_{g=1}^G \mathbb{E}_{(\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL}, \rho)} (X_0 X_g) - \sum_{g=1}^G N_g^{-1} \sum_{i=1}^{N_g} X_i X_{i+g} \right)^2, \quad (9)$$

and $(\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL}, \rho)$ indicates the model parameters used to compute the cross-moment.

4 | ASYMPTOTICS

In this section, we provide different convergence guarantees for the estimators defined in Section 3: the strong consistency of the UL estimator and asymptotic normality of the UL and M1–M4 estimators. The convergence results for M1–M4 are presented in the sense of weakly approaching sequences of distributions (Belyaev & Luna, 2000, definition 1) from which the convergence in distribution is implied. It is a stronger result and yields the convergence in the limit but also as sequences, that is, between the estimators and a potential sequence that estimates the limiting distribution.

Denote by $(x, y) \mapsto \zeta(x, y; \theta)$ a d -dimensional measurable (and differentiable where necessary) mapping of $(X_i, X_j) \in \mathcal{G}$ with parameter $\theta \in \mathbb{R}^d$; for simplicity, we sometimes write $\zeta_i^g(\theta)$ for $\zeta(X_i, X_{i+g}; \theta)$ and we use $d = 3 + p$ where p is such that $\rho \in \mathbb{R}^p$. This mapping is further referred to as the *score*. When discussing the regularity of $\theta \mapsto \zeta_i^g(\theta)$, it is an abuse of notation as we actually discuss the regularity of $\theta \mapsto \mathbb{E}(\zeta_i^g(\theta) | X_i, X_{i+g})$. We define the following matrices

$$\begin{aligned} \Gamma_g(\theta) &:= \lim_{N_g \rightarrow \infty} \mathbb{E} \left(N_g^{-1} \sum_{i=1}^{N_g} \nabla_{\theta} \zeta_i^g(\theta) \right), & \Gamma(\theta) &:= G^{-1} \sum_{g=1}^G \Gamma_g(\theta), \\ \Omega_g(\theta) &:= \lim_{N_g \rightarrow \infty} \text{Var} \left(N_g^{-1/2} \sum_{i=1}^{N_g} \zeta_i^g(\theta) \right), & \Omega(\theta) &:= \lim_{N \rightarrow \infty} \text{Var} \left(G^{-1/2} \sum_{g=1}^G N_g^{-1/2} \sum_{i=1}^{N_g} \zeta_i^g(\theta) \right). \end{aligned}$$

Note that we simplify notation by writing Ω , Ω_g , Γ , and Γ_g when evaluated at θ_0 . By strict stationarity, one has $\Gamma_g(\theta) = \mathbb{E}(\nabla_{\theta} \zeta_i^g(\theta))$ for any $i \in \mathbb{N}$.

4.1 | Asymptotics for UL

We denote by $\mathcal{N}_0$ a neighborhood of the true parameter values $(\xi_0, \sigma_0, \kappa_0)$ in $(-1/2, \infty) \times (0, \infty) \times (0, \infty)$. The GPD has infinite variance if $\xi \geq 1/2$ and the corresponding MLE is only asymptotically consistent if $\xi > -1/2$ (Smith, 1985) and the same bounds for ξ arise for the asymptotic convergence to occur as given in the proposition below.

Proposition 4. *For any $\xi_0 > -1/2$, $\sigma_0 > 0$ and $\kappa_0 > 0$, UL is infinitely continuously differentiable in (ξ, σ, κ) such that $\mathbb{E}[|\ln f_{\text{UL}}(X_i; \xi, \sigma, \kappa)|] < \infty$ for any $i \in \mathbb{N}$ and $(\xi, \sigma, \kappa) \in \mathcal{N}_0$ as well as*

$$\mathbb{E}(\|s_{\text{UL}}(\xi_0, \sigma_0, \kappa_0)\|^2) < \infty \quad \text{and} \quad \mathbb{E}\left(\sup_{(\xi, \sigma, \kappa) \in \mathcal{N}_0} \|\nabla_{(\xi, \sigma, \kappa)} s_{\text{UL}}(\xi, \sigma, \kappa)\|\right) < \infty,$$

where $s_{\text{UL}}(\xi, \sigma, \kappa) := s_{\text{UL}}(x; \xi, \sigma, \kappa) = \nabla_{(\xi, \sigma, \kappa)} \ln f_{\text{UL}}(x; \xi, \sigma, \kappa)$.

Proof. See Appendix D. ■

Proposition 5. *Under the conditions of Proposition 4 and for strictly increasing trawl functions, we have for some $\delta > 0$*

$$\mathbb{E}(\|s_{\text{UL}}(\xi_0, \sigma_0, \kappa_0)\|^{2+\delta}) < \infty \quad \text{and} \quad \sum_{k=0}^{\infty} k^2 \alpha(X, k)^{\delta/(2+\delta)} < \infty.$$

Proof. The first point is proved similarly to the second point of Proposition 4. We have that $\alpha(X, k) = O(\text{Cor}(\Lambda_0, \Lambda_k))$ as given by Proposition 3 where $\text{Cor}(\Lambda_0, \Lambda_k) \propto \mu^{(\text{leb})}(A \cap A_k) / \mu^{(\text{leb})}(A)$ in k therefore if $\mu^{(\text{leb})}(A \cap A_{k+1}) < \mu^{(\text{leb})}(A \cap A_k)$, D'Alembert's criterion shows the summability of $(k^2 \alpha(X, k))^{\delta/(2+\delta)}$, $k \in \mathbb{N}$ for any $\delta > 0$ ■

We obtain the consistency and joint normality of the estimators $(\hat{\xi}_{\text{UL}}, \hat{\sigma}_{\text{UL}}, \hat{\kappa}_{\text{UL}})$ as follows:

Proposition 6. *Suppose that $\xi_0 > -1/3$. For strictly increasing trawl functions and $(\xi_0, \sigma_0, \kappa_0)$ in a compact subset of $(-1/3, \infty) \times \mathbb{R} \times \mathbb{R}$, the UL estimators are strongly consistent and we obtain*

$$\sqrt{N} \begin{bmatrix} \hat{\xi}_{\text{UL}, N} \\ \hat{\sigma}_{\text{UL}, N} \\ \hat{\kappa}_{\text{UL}, N} \end{bmatrix} - \begin{bmatrix} \xi_0 \\ \sigma_0 \\ \kappa_0 \end{bmatrix} \xrightarrow{D} \mathcal{N}(\mathbf{0}, S_{\text{UL}}^{-1} \cdot \text{Var}_{\text{UL}} \cdot S_{\text{UL}}^{-1}), \quad \text{as } N \rightarrow \infty,$$

where $S_{\text{UL}} := \mathbb{E}[s_{\text{UL}}(\xi_0, \sigma_0, \kappa_0) \cdot s_{\text{UL}}(\xi_0, \sigma_0, \kappa_0)^{\top}]$ and $\text{Var}_{\text{UL}} := -\mathbb{E}[\nabla_{(\xi, \sigma, \kappa)}^2 \text{UL}(\xi_0, \sigma_0, \kappa_0)]$.

Proof. See Appendix D.0.2. ■

Given the dependencies in the observations, we adjust the estimation of the variance matrix using the HAC estimator (Newey & West, 1987) in Section 6.

4.2 | Breakdown of the following steps

We extend the asymptotic convergence in distribution of single-stage and two-stage estimators from Newey and McFadden (1994) and Hwang (2021) to weakly approaching sequences of distributions. As explained in Section 3, p. 2142, Newey and McFadden (1994)

As described above, asymptotic normality results from convergence in probability of the Hessian, convergence in distribution of the average score, and the Slutsky theorem.

Therefore, we first introduce the assumptions necessary for the asymptotic study and provide a Slutsky-type result for weakly approaching sequences of distributions. We then prove the weakly approaching convergence of the pairwise mapping ζ . Next, we prove the strong mixing property of the score which gives, in particular, the ergodicity of the LTME process. From that, we obtain the convergence in probability of the Hessian directly for single-stage approaches M1 and M3 and under a consistency assumption of the first stage in two-stage methods M2 and M4.

4.3 | Assumptions for the single-stage asymptotics

To prove the convergence of the estimation schemes M1 and M3, we consider the following assumptions

Assumption 1. Suppose that: (i) $\mathbb{E}(\zeta_i^g(\theta)) = \mathbf{0}$ for any $g \in \{1, \dots, G\}$ and any $1 \leq i \leq N_g$ if and only if $\theta = \theta_0$. (ii) We assume that the impact of the $(G + 1)$ th cluster on the inference is negligible.

The first point is equivalent to maximizing the likelihood function or gives the moments condition for the AM approach. The second point means that the following asymptotic results are sufficient as Ω contains all the sample variability: no additional bias study nor adjustment ought to be performed.

We introduce a regularity assumption on the latent trawl set to allow the application of the central limit theorem from Hwang (2021). It unfolds as follows

Assumption 2. Consider a trawl process $(\Lambda_t, t \geq 0)$ with corresponding trawl sets (A_t) parametrized by ρ . One supposes that $\rho \mapsto \mu^{\text{leb}}(A \cap A_t)$ is twice continuously differentiable.

We then clarify our clustering setup through the following assumption:

Assumption 3 (Adapted from assumption 1 of Hwang (2021)). Suppose that (i) the number of clusters G is fixed and (ii) the cardinality of the clusters grows with the total sample size N and are approximately equal, that is, for $\bar{N} := G^{-1} \sum_{h=1}^G N_h$ we have $N_g/\bar{N} \rightarrow 1$ as $N \rightarrow \infty$.

In particular, for the clusters $(G_g, g \in \{1, \dots, G\})$ defined above, Assumption 1 is sufficient to show that Assumption 3 (ii) holds since $N_g = N - g$. We relax the assumption of a *cross-sectional CLT* across clusters given in Assumption 1, (ii), Hwang (2021) by showing a similar result can be deduced from Assumptions 1 and 3 and a couple of standard conditions, see Section 4.4. In particular, our moments condition is cluster-wise asymptotically uncorrelated. This result is necessary to show the convergence of the M1 and M3 estimators but does not require the strict independence between clusters—which is more flexible to the case where G grows with the sample size (Carter et al., 2017; Hansen & Lee, 2019).

4.4 | The convergence of a pairwise measurable mapping

In this section, we are interested in obtaining a Slutsky-type result for weakly approaching sequences of distributions and the convergence of $G^{1/2}N_g^{1/2}(\hat{\theta} - \theta_0)$ where $\hat{\theta} = \hat{\theta}_{N,G}$ is the estimator of θ_0 given by minimizing a function of

$$\bar{\zeta}_N(\theta) := G^{-1} \sum_{g=1}^G N_g^{-1} \sum_{i=1}^{N_g} \zeta_i^g(\theta),$$

where $\zeta_i^g(\theta)$ is a mapping of the pair (X_i, X_{i+g}) as defined in the introduction of Section 4. Under some standard derivability conditions, we obtain the asymptotic convergence of the PL estimator $\hat{\theta}_{M1}$ and the first step of the M3 estimation $\hat{\theta}_{M3,1}$ —see Sections 4.6 and 4.8, respectively.

Since we are dealing with strongly mixing sequences, the literature uses the concept of *weakly approaching sequences of distributions* introduced in definition 1 of Belyaev and Luna (2000):

Definition 5. Two sequences $(K_n, n \in \mathbb{N})$ and $(L_n, n \in \mathbb{N})$ are said to have weakly approaching distributions, respectively denoted $(\mathcal{L}(K_n), n \in \mathbb{N})$ and $(\mathcal{L}(L_n), n \in \mathbb{N})$, if for any real-valued bounded continuous function $f(\cdot)$, $\mathbb{E}[f(K_n)] - \mathbb{E}[f(L_n)] \rightarrow 0$ as $n \rightarrow \infty$ and we write $\mathcal{L}(K_n) \xleftrightarrow{\text{wa}} \mathcal{L}(L_n)$ as $n \rightarrow \infty$.

4.4.1 | A Slutsky-type result

To obtain a central limit theorem, we extend lemma 7 of Belyaev and Luna (2000) and prove an additional Slutsky-type result for the product of weakly approaching sequences of distributions:

Proposition 7. Consider three sequences $(K_n, n \in \mathbb{N}) \subseteq \mathbb{R}^d$, $(L_n, n \in \mathbb{N}) \subseteq \mathbb{R}^d$ and $(M_n, n \in \mathbb{N}) \subseteq \mathbb{R}^d$ defined on the same probability space. Suppose that (M_n) is tight, that is, if for any $\epsilon > 0$ there is a compact set $C_\epsilon \subset \mathbb{R}^d$ such that $P(M_n \notin C_\epsilon) < \epsilon$ for all $n \in \mathbb{N}$, and converges in probability to M . If $\mathcal{L}(K_n) \xleftrightarrow{\text{wa}} \mathcal{L}(L_n)$ then, $\mathcal{L}(M_n K_n) \xleftrightarrow{\text{wa}} \mathcal{L}(M K_n)$, $\mathcal{L}(M_n L_n) \xleftrightarrow{\text{wa}} \mathcal{L}(M L_n)$ and $\mathcal{L}(M_n L_n) \xleftrightarrow{\text{wa}} \mathcal{L}(M_n K_n)$ as $n \rightarrow \infty$.

Proof. See Appendix E. ■

We prove the tightness of $(\zeta_i^g(\theta_0), i \in \mathbb{N})$ and $(\nabla_{\theta} \bar{\zeta}_N(\theta_0), N \in \mathbb{N})$ as well as the strong consistency of the latter:

Proposition 8. Let $g \in \{1, \dots, G\}$ and suppose $(\zeta_i^g(\theta_0), i \in \mathbb{N})$ is strongly mixing, square-integrable and continuously differentiable in θ . Then, the sequence $(\nabla_{\theta} \bar{\zeta}_N(\theta_0), N \in \mathbb{N})$, converges almost surely to its limit $\Gamma(\theta_0)$. In addition, these sequences are tight.

Proof. See Appendix F.1. ■

4.4.2 | Convergence of the score

Recall that the g th cluster cardinality is given by $N_g = N - g$. We obtain a fundamental lemma towards obtaining a central limit theorem for $\hat{\theta}$:

Lemma 1. Suppose Assumptions 1 and 3 hold for the LTME. If for some $\delta > 0$, one has:

- (A1) $\sum_{k=0}^{\infty} (k+1)^2 \alpha^*(\zeta^g(\theta_0), k)^{\delta/(4+\delta)} < \infty$ and where $\zeta(\theta) := (\zeta_i^g(\theta), i \in \mathbb{N}, g \in \{1, \dots, G\})$;
 (A2) $\mathbb{E} \left(\left\| \zeta_i^g(\theta_0) - \mathbb{E}[\zeta_i^g(\theta_0)] \right\|^{2+\delta} \right) < \infty$ for all $g \in \{1, \dots, G\}, i \in \{1, \dots, N_g\}$;

Then, for the cluster-wise scaled vector of partial sums given by

$$\mathbf{W}_N^G(\theta_0) := \left(N_1^{-1/2} \sum_{i=1}^{N_1} \zeta_i^1(\theta_0)^\top, \dots, N_G^{-1/2} \sum_{i=1}^{N_G} \zeta_i^G(\theta_0)^\top \right)^\top,$$

we obtain the following convergence of sequences

$$\mathcal{L}(\mathbf{W}_N^G(\theta_0)) \xrightarrow{\text{wa}} \mathcal{N}(\mathbf{0}, \text{Var}(\mathbf{W}_N^G(\theta_0))), \quad \text{as } N \rightarrow \infty. \quad (10)$$

which simplifies into

$$\mathcal{L}(\mathbf{W}_N^G(\theta_0)) \xrightarrow{\text{wa}} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \Omega_1(\theta_0) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Omega_G(\theta_0) \end{bmatrix}\right), \quad \text{as } N \rightarrow \infty.$$

Note that the block elements are all finite by Condition (A2) and the right-hand side is to be considered as a constant sequence. In particular, we have

$$\mathcal{L}(G^{1/2} \bar{N}^{-1/2} \bar{\zeta}_N(\theta_0)) \xrightarrow{\text{wa}} \mathcal{N}(\mathbf{0}, \text{Var}(G^{1/2} \bar{N}^{-1/2} \bar{\zeta}_N(\theta_0))), \quad \text{as } N \rightarrow \infty, \quad (11)$$

which is then iterated into $\mathcal{L}(G^{1/2} \bar{N}^{-1/2} \bar{\zeta}_N(\theta_0)) \xrightarrow{\text{wa}} \mathcal{N}(\mathbf{0}, \Omega(\theta_0))$ as $N \rightarrow \infty$, with the decomposition $\Omega(\theta_0) = G^{-1} \sum_{g=1}^G \Omega_g(\theta_0)$.

Proof. See Appendix F.2. ■

Remark 3. From the definition of weakly approaching distributions along with the Portmanteau theorem, the results of Lemma 1 yield that

$$\left(N_1^{-1/2} \sum_{i=1}^{N_1} \zeta_i^1(\theta_0)^\top, \dots, N_G^{-1/2} \sum_{i=1}^{N_G} \zeta_i^G(\theta_0)^\top \right)^\top \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \Omega_1(\theta_0) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Omega_G(\theta_0) \end{bmatrix}\right),$$

as $N \rightarrow \infty$. This is an analogue to assumption 1 (ii) in Hwang (2021) but this time implied by the model formulation.

To obtain the asymptotics of $\hat{\theta}$, one needs to link the result from Lemma 1 to $\hat{\theta}$ itself using one of the sequences from Proposition 8. Using Proposition 7 along with the Portmanteau theorem, we obtain an essential reformulation

Proposition 9. Suppose the assumptions of Lemma 1 hold and, in addition, that $\theta \mapsto \zeta_i^g(\theta)$ is continuously differentiable in $N_b(\theta_0)$, an open neighbourhood of θ_0 . Then, one obtains

$$\mathcal{L}\left(G^{1/2}\bar{N}^{1/2}(\hat{\theta} - \theta_0)\right) \xrightarrow{\text{wa}} \mathcal{N}\left(\mathbf{0}, \Gamma(\theta_0)^{-1} \cdot \Omega(\theta_0) \cdot (\Gamma(\theta_0)^{-1})^\top\right), \quad \text{as } N \rightarrow \infty.$$

Proof. See Appendix F.3. ■

Although standard results (Newey & McFadden, 1994, theorem 6.1) provide the *weaker* convergence in distribution, we can derive it from the weakly approaching convergence of the estimator as follows

Corollary 1. Suppose the assumptions of Lemma 1 hold and, in addition, that $\theta \mapsto \zeta_i^g(\theta)$ is continuously differentiable in $N_b(\theta_0)$, an open neighbourhood of θ_0 . The estimator is asymptotically normal as follows

$$G^{1/2}\bar{N}^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{D} \mathcal{N}\left(\mathbf{0}, \Gamma(\theta_0)^{-1} \cdot \Omega(\theta_0) \cdot (\Gamma(\theta_0)^{-1})^\top\right), \quad \text{as } N \rightarrow \infty.$$

Proof. We apply Proposition 9 and noticing that, by definition of weakly approaching sequences, this corresponds to the setting of the Portmanteau theorem (as in Remark 3) and we conclude then. ■

Remark 4. The limiting distribution in (11) is given as a sequence in N . For ease of notation, we present the M1–M4 convergence results for with a constant sequence (e.g. Equation 10) but they also hold for sequences of variance matrices $(\text{Var}\left(G^{1/2}\bar{N}^{1/2}\bar{\zeta}_N(\theta_0)\right), N \in \mathbb{N})$.

4.5 | Strong mixing property of the score

We check that the score function ζ is strongly mixing and that Condition (A1) can be verified for any measurable pairwise mapping ζ as defined above under the LTME if the trawl function vanishes quickly enough as $t \rightarrow -\infty$. One formulates the following proposition

Proposition 10. Let $g \in \{1, \dots, G\}$. For the LTME with a latent trawl process (Λ_j) , then the process $\zeta^g(\theta) := (\zeta(X_i, X_{i+g}; \theta), i \in \mathbb{N})$ has a strong mixing coefficient dominated by the auto-correlation of (Λ_j) , i.e.

$$\alpha^*(\zeta^g(\theta), k) = O(\text{Cor}(\Lambda_0, \Lambda_k)).$$

Similarly, for $\zeta(\theta_0) = \cup_{g=1}^G \zeta^g(\theta_0)$, we also have

$$\alpha^*(\zeta(\theta), k) = O(\text{Cor}(\Lambda_0, \Lambda_k)).$$

If its trawl set vanishes asymptotically, i.e. $\mu^{\text{leb}}(A \cap A_k) \rightarrow 0$ as $k \rightarrow \infty$, then $\zeta^g(\theta)$ and $\zeta(\theta)$ are strongly mixing. Finally, if the trawl set is asymptotically strictly increasing, that is, if $\lim_{k \rightarrow \infty} \mu^{\text{leb}}(A \cap A_{k+1}) / \mu^{\text{leb}}(A \cap A_k) < 1$, Condition (A1) holds.

Proof. See Appendix F.4. ■

Although Proposition 10 yields that Condition (A1) holds independently of the score function, Condition (A2) should be verified individually for different score function ζ_i^g which we do in both the PL and GMM cases.

4.6 | Asymptotics for Method 1

In this section, we show that one can apply Lemma 1 with strongly mixing observations for M1, the single-stage PL scheme (M1) if latent process (Ψ_j) is a trawl process and we assume that Assumptions 1, 2, and 3 hold with an asymptotically vanishing and strictly increasing trawl set.

In particular, we define ζ_i^g from Section 4.4 as the PL score function, that is,

$$\zeta_i^g(\theta) = \zeta(X_i, X_{i+g}; \theta) := \nabla_{\theta} \ln f(X_i, X_{i+g}; \theta), \quad 1 \leq g \leq G, \quad 1 \leq i \leq N_g,$$

First, we show that Condition (A2) is satisfied:

Lemma 2. *For the LTME in the PL case, Condition (A2) is satisfied by any parameter vector $\theta_0 \in \mathbb{R} \times (0, \infty)^2 \times \mathbb{R}^p$.*

Proof. See Appendix F.5. ■

The reasoning behind this lemma is that the parameter that actually matters is the shape parameter α of the transformed marginal, that is, Condition (A2) holds if $\alpha \geq 1$ which is the case since we set $\alpha = 1$.

We now formulate the convergence result as follows

Proposition 11. *Suppose that the latent process (Ψ_j) is a trawl process and assume that Assumptions 1, 3, and 2 hold with an asymptotically vanishing and strictly increasing trawl set. Suppose that the θ_0 is in the interior of a compact set $\Theta \subset \mathbb{R} \times (0, \infty) \times (0, \infty) \times \mathbb{R}^p$, then we have that*

$$\mathcal{L} \left(G^{1/2} \bar{N}^{1/2} (\hat{\theta}_{M1} - \theta_0) \right) \overset{\text{wa}}{\rightsquigarrow} \mathcal{N} \left(\mathbf{0}, \Gamma(\theta_0)^{-1} \cdot \Omega(\theta_0) \cdot (\Gamma(\theta_0)^{-1})^{\top} \right), \text{ as } N \rightarrow \infty,$$

where in particular $\Gamma(\theta_0) := -\mathbb{E} [\nabla_{\theta}^2 \text{PL}_N(\theta_0; G)]$.

Proof. Since the PL function is a proper likelihood function, the expectation of the score function is zero for $\theta = \theta_0$, that is, $\mathbb{E} [\bar{\zeta}_N(\theta_0)] = 0$. By Assumption 2 and the infinite differentiability w.r.t to (ξ, σ, κ) of PL_N , $\theta \mapsto \text{PL}_N(\theta; G)$ is twice continuously differentiable on Θ . Proposition 10 and Lemma 2 yield that both Conditions (A1) and (A2) are satisfied, respectively. Therefore, since we assume that Assumptions 1 and 3 hold, then Proposition 9 yields the required asymptotic normality of the M1 estimator. ■

4.7 | Asymptotics for Method 2

We are interested in incorporating the estimation error from the estimation of marginal parameters in the first stage into the second stage when estimating the trawl parameter ρ . As in

Section 4.6, this holds if we suppose Assumptions 1, 2, and 3 hold for the LTME with an asymptotically vanishing and strictly increasing trawl set but with the additional restriction that $\xi_0 > -1/3$.

Recall that, by Proposition 6, the marginal parameters estimated by UL maximization are strongly consistent and asymptotically normal. We obtain the following asymptotic convergence:

Proposition 12. *Suppose Assumptions 1, 2, and 3 hold for the LTME with an asymptotically vanishing and strictly increasing trawl set and $\theta_0 = (\xi_0, \sigma_0, \kappa_0, \rho_0)^\top$ such that $\xi_0 > -1/3$ and that θ_0 is in the interior of a compact set $\Theta \subset (-1/3, \infty) \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^p$.*

Then, $\hat{\rho}_{M2}$ is consistent. Also, the vector $\hat{\theta}_{M2} = (\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL}, \hat{\rho}_{M2})$ is asymptotically normal in the weakly approaching sense where $(\sum_{g=1}^G N_g)^{1/2} \hat{\rho}_{M2}$ has the asymptotic variance

$$\Gamma_{PL,\rho} \cdot \left\{ G^{-1} \sum_{g=1}^G \mathbb{E} \left[\left(s_{PL}^g - \Gamma_{(\xi,\sigma,\kappa)}^{-1} \text{Var}_{UL}^{-1} \bar{s}_{UL} \right) \left(s_{PL}^g - \Gamma_{(\xi,\sigma,\kappa)}^{-1} \text{Var}_{UL}^{-1} \bar{s}_{UL} \right)^\top \right] \right\} \cdot \Gamma_{PL,\rho}^\top,$$

where $s_{PL}^g = s_{PL}(X_0, X_g; \theta_0)$ and $\bar{s}_{UL} (= \bar{s}_{UL}^g) = s_{UL}(X_0; \xi_0, \sigma_0, \kappa_0) + s_{UL}(X_g; \xi_0, \sigma_0, \kappa_0)$ the PL and UL score functions, respectively, for a pair (X_0, X_g) . Also, we have $\Gamma_{PL,\rho} = G^{-1} \sum_{g=1}^G \mathbb{E}[\nabla_{\rho} s_{PL}^g(\theta_0)]$, $\Gamma_{PL,(\xi,\sigma,\kappa)} = G^{-1} \sum_{g=1}^G \mathbb{E}[\nabla_{(\xi,\sigma,\kappa)} s_{PL}^g(\theta_0)]$ and finally, we define $\text{Var}_{UL} := -\mathbb{E}[\nabla_{(\xi,\sigma,\kappa)} s_{UL}(\xi_0, \sigma_0, \kappa_0)]$.

Proof. Recall that the UL and PL score functions have a zero expectations at θ_0 and are continuously differentiable. First, we check that conditions (i)–(iv) of theorem 2.6 of Newey and McFadden (1994) are satisfied for the estimation of ρ : (i) the weighting matrix is implicitly $W = \mathbf{I}_{p \times p}$ and the PL score has zero expectation only if $\theta = \theta_0$. By strong consistency of the UL estimates and the continuous mapping theorem, this also holds with probability approaching one when plugging in $(\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL})$ as $N \rightarrow \infty$ if $\rho = \rho_0$; (ii) Θ is compact by definition; (iii) the PL score function is continuous; (iv) since $\xi > -1/3$, then the PL score function is integrable at θ_0 and hence, by (iii), it is also integrable in a compact neighbourhood of θ_0 . Thus, theorem 2.6 of Newey and McFadden (1994) yields the consistency of $\hat{\rho}_{M2}$. Next, we proceed to show the asymptotic normality property: again, since $(\hat{\xi}_{UL,N}, \hat{\sigma}_{UL,N}, \hat{\kappa}_{UL,N})^\top$ are strongly consistent, both equations (6.1) and (6.2) of Newey and McFadden (1994) are satisfied with probability approaching one as $N \rightarrow \infty$. We then check that conditions (i)–(v) of theorem 3.4 of Newey and McFadden (1994) are satisfied: first, condition (i) holds since Θ is a compact set. Then, since UL and PL are twice continuously differentiable, condition (ii) is satisfied. For condition (iii), recall that the score functions have a zero expectations at θ_0 while Lemma 2 yields that (A2) holds. The UL Fisher information matrix is finite in a neighbourhood $\mathcal{N}_0$ of $(\xi_0, \sigma_0, \kappa_0)$ (Proposition 4) as well as invertible. The independence of UL from ρ yields that $\nabla_{\theta} [s_{UL}(\xi_0, \sigma_0, \kappa_0)^\top, s_{PL}(\xi_0, \sigma_0, \kappa_0, \rho_0)^\top]^\top$ is invertible too. Since both UL and PL are twice continuously differentiable, then it is still invertible in $\mathcal{N}_0$ and

$$\mathbb{E} \left(\sup_{(\xi,\sigma,\kappa,\rho) \in \mathcal{N}_0} \nabla_{\theta} [s_{UL}(\xi, \sigma, \kappa)^\top, s_{PL}^g(\xi, \sigma, \kappa, \rho)^\top]^\top \right) < \infty,$$

and, therefore, both conditions (iv) holds. For condition (v), since the moments conditions are twice continuously differentiable then $M := \mathbb{E}(\nabla_{\theta} [s_{UL}(\xi_0, \sigma_0, \kappa_0)^\top, s_{PL}(\xi_0, \sigma_0, \kappa_0, \rho_0)^\top]^\top)$ is symmetric such that $M \cdot M^\top = M^2$ is positive definite hence nonsingular. Next, we leverage theorem 6.1 of Newey and McFadden (1994) where, instead of using the asymptotic normality given by theorem 3.4, we apply Proposition 9 on $[s_{UL}(\xi_0, \sigma_0, \kappa_0)^\top, s_{PL}(\xi_0, \sigma_0, \kappa_0, \rho_0)^\top]^\top$ since Condition (A1)

holds (Proposition 10). Since the sequence is tight (Proposition 8), our Slutsky-type result (Proposition 7) yields the asymptotic normality of $(\sum_{g=1}^G N_g)^{1/2} \hat{\rho}_{M2}$ in the weakly approaching sense with asymptotic variance

$$\Gamma_{PL,\rho} \cdot \left\{ \mathbb{E} \left[\left(G^{-1/2} \sum_{g=1}^G s_{PL}^g - \Gamma_{(\xi,\sigma,\kappa)}^{-1} \text{Var}_{UL}^{-1} \bar{s}_{UL} \right) \left(G^{-1/2} \sum_{g=1}^G s_{PL}^g - \Gamma_{(\xi,\sigma,\kappa)}^{-1} \text{Var}_{UL}^{-1} \bar{s}_{UL} \right)^{\top} \right] \right\} \cdot \Gamma_{PL,\rho}^{\top}.$$

Again, we conclude that cluster-wise decomposition of the variance is given by Lemma 1. ■

Note that theorem 6.1 of Newey and McFadden (1994) can be applied to a strictly stationary and ergodic process (Proposition 3), see Section 2.4 therein.

4.8 | Asymptotics for Method 3

We adapt the result of Hwang (2021) on centered two-step GMM estimators to the two-step AM-based Method 3. This time, we assume that Assumptions 1, 2, and 3 hold for the LTME with an asymptotically vanishing and strictly increasing trawl set but where $-1/3 < \xi_0 < 1/4$ and where θ_0 is in a compact subset of $(-1/3, 1/4) \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^p$.

From (7), we define in particular ζ_i^g to be the element-wise moments condition, that is, $\zeta_i^g(\theta) := r_i^g(\theta)$ and we prove the connection between our framework and Hwang (2021) by showing that Assumptions 2–6 therein defined are satisfied in the following way:

Lemma 3. *If Assumptions 1 and 3 and Conditions (A1) and (A2) are satisfied then the following properties hold:*

(P2) $\hat{\theta}_{M3,1} \rightarrow \theta_0$ in probability as $N \rightarrow \infty$.

(P3) (i) Recall that $\Gamma_g(\theta) := \lim_{N_g \rightarrow \infty} \mathbb{E} \left(N_g^{-1} \sum_{i=1}^{N_g} \nabla_{\theta} \zeta_i^g(\theta) \right)$ for any $g \in \{1, \dots, G\}$. Then, we have that

$$\sup_{\theta \in N_b(\theta_0)} \left\| N_g^{-1} \sum_{i=1}^{N_g} \nabla_{\theta} \zeta_i^g(\theta) - \Gamma_g(\theta) \right\| \rightarrow 0, \quad \text{in probability as } N_g \rightarrow \infty,$$

where $N_b(\theta_0)$ is an open neighborhood of θ_0 and $\|\cdot\|$ is the Euclidean norm. (ii) $\Gamma_g(\theta)$ is continuous at $\theta = \theta_0$, and $\Gamma(\theta_0) := G^{-1} \sum_{g=1}^G \Gamma_g(\theta_0)$ has full rank.

(P4) Let $B_{3+p,g} \stackrel{i.i.d.}{\sim} N(\mathbf{0}, \mathbf{I}_{3+p})$ for $g \in \{1, \dots, G\}$, then

$$P \left(N_g^{-1/2} \sum_{i=1}^{N_g} \zeta_i^g(\theta_0) \leq \mathbf{x} \right) = P \left(\Omega_g^{1/2} B_{3+p,g} \leq \mathbf{x} \right) + o_p(1), \quad \text{as } N \rightarrow \infty,$$

for any $\mathbf{x} \in \mathbb{R}^{3+p}$.

Proof. See Appendix F.6. ■

This lemma proves that we can apply the convergence results from Hwang (2021). Note that we keep the assumption numbering consistent with Hwang (2021): properties (P2) and (P3)

correspond to Assumptions 2 and 3 in Hwang (2021), respectively. Although the property (P4) is not strictly necessary to obtain the asymptotic distribution to the inference scheme, it provides some explanation on the notation using in Proposition 13 below.

Remark that the Wald and t -statistic of the two-step GMM scheme are asymptotically pivotal only if Γ and Ω are homogeneous (i.e., $\Gamma_g(\theta_0) = \Gamma(\theta_0)$ and $\Omega_g(\theta_0) = \Omega(\theta_0)$ for any $g \in \{1, \dots, G\}$; Hwang, 2021, assumptions 4 and 5).

We can prove that Assumption 1 (i) holds directly in the sense that

Lemma 4. *Given the definition of ζ_i^g , Assumption 1 (i) holds if $\xi_0 < 1/2$.*

Proof. Since we consider maximizing the UL likelihood, the zero score function derivative property allows concluding for the first three components. Similarly to the proof of Lemma 5 in Appendix F.7, the cross-moment is finite if $\xi_0 < \alpha/2 = 1/2$. ■

A last intermediary result is a sufficient condition such that Condition (A2) holds

Lemma 5. *For the GMM setting, Condition (A2) holds for any $\delta > 0$ if $\xi_0 < 1/4$.*

Proof. See Appendix F.7. ■

We now formulate the main asymptotic convergence result for the two-step M3 estimator:

Proposition 13. *Suppose Assumptions 1, 2, and 3 hold for the LTME with an asymptotically vanishing and strictly increasing trawl set and $\theta_0 = (\xi_0, \sigma_0, \kappa_0, \rho_0)^\top$ such that $-1/3 < \xi_0 < 1/4$ and that θ_0 is in the interior of a compact set $\Theta \subset (-1/3, 1/4) \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^p$. Then, one obtains the following convergence of the two-step estimator*

$$\mathcal{L}\left((G^{1/2}\bar{N}^{1/2}(\hat{\theta}_{M3,2} - \theta_0))\right) \xrightarrow{\text{wa}} \mathcal{L}\left(-[\Gamma^\top(\Omega_\infty)^{-1}\Gamma]^{-1}\Gamma^\top(\Omega_\infty)^{-1}\Omega^{1/2}G^{1/2}\bar{B}_{3+p}\right), \quad (12)$$

where $\Omega = \Omega(\theta_0)$, $\bar{B}_{3+p} := G^{-1}\sum_{g=1}^G B_{3+p,g}$ with $B_{3+p,g} \stackrel{i.i.d}{\sim} \mathcal{N}(0, I_{3+p})$ for $g \in \{1, \dots, G\}$ and $\Omega_\infty := \Omega^{1/2} \cdot \bar{W}_G \cdot (\Omega^{1/2})^\top$, where

$$\bar{W}_G := G^{-1}\sum_{g=1}^G (\Omega_g^{1/2}B_{3+p,g} - \Gamma_g\Gamma^{-1}\Omega^{1/2}\bar{B}_{3+p}) \cdot (\Omega_g^{1/2}B_{3+p,g} - \Gamma_g\Gamma^{-1}\Omega^{1/2}\bar{B}_{3+p})^\top.$$

Proof. Proposition 10 and Lemma 5 yield that Conditions (A1) and (A2) hold, respectively. Then, by application of Lemma 1, we obtain the convergence in the sense of weakly approaching sequences of distributions given by

$$\mathcal{L}\left((G\bar{N})^{1/2}\bar{r}_N(\theta_0)\right) \xrightarrow{\text{wa}} \mathcal{L}\left(\Omega(\theta_0)^{1/2} \cdot \bar{B}_{3+p}\right),$$

as $N \rightarrow \infty$ since $\bar{N}/N \rightarrow 1$. By performing a Taylor's expansion around θ_0 of the AM objective function, we have

$$G^{1/2}\bar{N}^{1/2}(\hat{\theta}_{M3,2} - \theta_0) = -\left[\Gamma^\top\left[\hat{\Omega}(\hat{\theta}_{M3,1})\right]^{-1}\Gamma\right]^{-1}\Gamma^\top\left[\hat{\Omega}(\hat{\theta}_{M3,1})\right]^{-1}\Omega^{1/2}\bar{N}^{1/2}\bar{r}_N(\theta_0) + o_p(1).$$

Thanks to Lemmas 1 and 3, the cluster-wise convergence and the properties (P2) & (P3) hold. Those results correspond to assumptions 1–3 of Hwang (2021) where Assumption 1-(ii) is shown in the weakly approaching sense and can be used as it is thanks to the Slutsky-type result from Proposition 7. Thus, proposition 1 from Hwang (2021) in its form given in remark 4 that relaxes the cluster homogeneity conditions (since $\Omega_g \neq \Omega$ and $\Gamma_g \neq \Gamma$ in general) yields that $\hat{\Omega}(\hat{\theta}_{M3,1}) \xrightarrow{wa} \Omega_\infty = \Omega^{1/2} \cdot \bar{W}_G \cdot (\Omega^{1/2})^\top$, where

$$\bar{W}_G := G^{-1} \sum_{g=1}^G (\Omega_g^{1/2} B_{3+p,g} - \Gamma_g \Gamma^{-1} \Omega^{1/2} \bar{B}_{3+p}) \cdot (\Omega_g^{1/2} B_{3+p,g} - \Gamma_g \Gamma^{-1} \Omega^{1/2} \bar{B}_{3+p})^\top.$$

Finally, the Slutsky-type result from Proposition 7 yields that, as $N \rightarrow \infty$, we have

$$\mathcal{L} \left((G^{1/2} \bar{N}^{1/2} (\hat{\theta}_{M3,2} - \theta_0)) \right) \xrightarrow{wa} \mathcal{L} \left(-[\Gamma^\top (\Omega_\infty)^{-1} \Gamma]^{-1} \Gamma^\top (\Omega_\infty)^{-1} \Omega^{1/2} G^{1/2} \bar{B}_{3+p} \right).$$

This implies $\hat{\Omega}(\theta_1)$ converges in distribution to Ω_∞ which follows a highly nonstandard distribution based on Gaussian mixtures. In $\bar{W}_G$, the term $\Omega_g^{1/2}$ scales each Gaussian distribution to represent each cluster whilst the term $\Gamma_g \Gamma^{-1} \Omega^{1/2}$ is a *quasi-demeaning* factor that rises from the noncentered variance estimator defined in (8) as the empirical mean is not zero in finite samples, see sections 3.2 (p. 1001) & 6 of Hwang (2021).

4.9 | Asymptotics for Method 4

Analogously to Section 4.8, we assume that latent process (Ψ_j) is a trawl process and that Assumptions 1, 2, and 3 hold with an asymptotically vanishing and strictly increasing trawl set. We also suppose that $-1/3 < \xi_0 < 1/4$ and θ_0 is in a compact subset of $(-1/3, 1/4) \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^p$. The asymptotic convergence result for M4 unfolds as follows:

Corollary 2. Suppose Assumptions 1, 2, and 3 hold for the LTME with an asymptotically vanishing and strictly increasing trawl set and $\theta_0 = (\xi_0, \sigma_0, \kappa_0, \rho_0)^\top$ such that $-1/3 < \xi_0 < 1/4$ and that θ_0 is in the interior of a compact set $\Theta \subset (-1/3, 1/4) \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^p$. Finally, suppose that, when plugging in $(\xi_0, \sigma_0, \kappa_0)^\top$, the AM moments condition given by (9) holds only if $\rho = \rho_0$.

Then, $\hat{\rho}_{M4}$ is consistent. Also, the vector $\hat{\theta}_{M4} = (\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL}, \hat{\rho}_{M4})$ is asymptotically normal in the weakly approaching sense where $(\sum_{g=1}^G N_g)^{1/2} \hat{\rho}_{M4}$ has the asymptotic variance

$$\Gamma_{AM,\rho} \cdot \left\{ G^{-1} \sum_{g=1}^G \mathbb{E} \left[\left(s_{AM}^g - \Gamma_{(\xi,\sigma,\kappa)}^{-1} \text{Var}_{UL}^{-1} \bar{s}_{UL} \right) \left(s_{AM}^g - \Gamma_{(\xi,\sigma,\kappa)}^{-1} \text{Var}_{UL}^{-1} \bar{s}_{UL} \right)^\top \right] \right\} \cdot \Gamma_{AM,\rho}^\top,$$

where $s_{AM}^g (= s_{AM}(X_0, X_g; \theta_0)) := 2\nabla_\rho \mathbb{E}_{\theta_0}(X_0 X_g) [X_0 X_g - \mathbb{E}_{\theta_0}(X_0 X_g)]$, and where $\Gamma_{AM,\rho} = G^{-1} \sum_{g=1}^G \mathbb{E}[\nabla_\rho s_{AM}^g(\theta_0)]$ as well as $\Gamma_{AM,(\xi,\sigma,\kappa)} = G^{-1} \sum_{g=1}^G \mathbb{E}[\nabla_{(\xi,\sigma,\kappa)} s_{AM}^g(\theta_0)]$.

Proof. We check that conditions (i)–(iv) of theorem 2.6 of Newey and McFadden (1994) are satisfied for the estimation of ρ in the second stage: (i) the weighting matrix is implicitly $W = \mathbf{I}_{p \times p}$ and the AM moments condition has zero expectation only if $\theta = \theta_0$ by assumption. By the strong

consistency of the UL estimates and the continuous mapping theorem, this also holds with probability approaching one as $N \rightarrow \infty$ when plugging in $(\hat{\xi}_{UL}, \hat{\sigma}_{UL}, \hat{\kappa}_{UL})$ only if $\rho = \rho_0$; (ii) Θ is compact by definition; (iii) the AM moments condition is continuous in ρ ; (iv) then the AM moments condition is integrable at θ_0 since $-1/3 < \xi < 1/4$ and hence, by (iii), it is also integrable in a compact neighbourhood of θ_0 . Therefore, $\hat{\rho}_{M4}$ is consistent by application of theorem 2.6 of Newey and McFadden (1994). Then, note that s_{AM}^g is the derivative w.r.t. ρ of the AM objective function used in (9). Both the UL score function and s_{AM}^g have zero expectation and are continuously differentiable on a compact set that contains θ_0 . Since UL is independent from ρ , the expectation of the Hessian matrix $\nabla_{\theta}[s_{UL}(\xi, \sigma, \kappa)^T, s_{AM}(\xi, \sigma, \kappa, \rho)^T]^T$ has finite expectation on $\mathcal{N}_0$ and is invertible. In addition, Conditions (A1) and (A2) hold similarly (Proposition 10 and Lemma 5, respectively). We can therefore replicate the rest of the proof of Proposition 12. ■

5 | SIMULATION AND THE LTME PROCESS

In this section, we sketch the reasons why it is not possible to simulate data under the current formulation of the LTME process, and therefore, why we cannot perform a simulation study and will proceed with a block bootstrap in Section 6. From the hierarchical and latent structure of the LTME process (Section 2.2.1), one may suppose that the following algorithm allows to simulate from the LTME model:

Algorithm 1. (Construction via the latent process).

1. Generate a path at times $t_1 < \dots < t_N$ of the latent process $\lambda_1, \dots, \lambda_N$ and a sequence of i.i.d samples $u_1, \dots, u_N \sim U(0, 1)$;
2. If $u_i < 1 - e^{-\kappa\lambda_i}$, set $X_i = 0$; otherwise, $X_i \sim \text{Exp}(\lambda_i)$;

However, this allows to simulate a path of the LTME process $(X_1, \dots, X_N)$ given a path of the latent process $(\Lambda_1, \dots, \Lambda_N)$. The structural assumption implied by the conditional independence assumption of X_j conditional on Ψ_j given in (2) yields that the dynamics of observed process X are independent from any previous sequential information from X itself. For a trawl latent process Λ , Bayes' formula gives

$$P(X_{i+k} > 0 | X_i > 0, \Lambda_{i+k} = \lambda_{i+k}, \Lambda_i = \lambda_i) = e^{-\kappa\lambda_{i+k}} = P(X_{i+k} > 0 | \Lambda_{i+k} = \lambda_{i+k}),$$

for $i \in \mathbb{Z}$ and $k \in \mathbb{N}$. Therefore, the probability of a positive exceedance at time t_{i+k} given that there is one at time t_i only relies on λ_{i+k} . Although λ_{i+k} is related to λ_i through the trawl dependence structure, the information from $\{X_i > 0\}$ is not passed forward: for instance, in Algorithm 1, one draws $u_i \sim U(0, 1)$ and checks that $u_i < e^{-\kappa\lambda_i}$ to assess if the exceedance at time t_i is positive or zero, respectively. The hierarchical nature of the model is particularly useful for inference with closed-form expressions for moments and likelihood; however, it is challenging to simulate data from it.

Furthermore, to simulate unconditionally on the latent process, we ought to average out the contribution of latent path, that is, by using that

$$f(x_1, \dots, x_N) = \int_{(0, \infty)^N} \left[\prod_{i=1}^N f(x_i | \lambda_i) \right] f(\lambda_1, \dots, \lambda_N) d\lambda_1 \dots d\lambda_N.$$

Similarly to the PL derivation in appendix A.1 of Noven et al. (2018) (see Appendix B), this particular expression is not separable with respect to the latent marginals making any exact sequential simulation procedure impractical. It is also intractable and not invertible and approximate schemes are beyond the scope of this article.

6 | CASE STUDY: LONDON AIR POLLUTION

In this section, we apply the LTME to six main air pollutants presented below and we compare the inference strategies M1 to M4 with the exponential trawl set. The data was provided by *King's College London Air Quality Network* which provides air pollution data on different timescales and pollutants for many locations in Greater London. All the chemicals reactions mentioned below are described in detail in the related air pollution literature from the World Health Organization (Krzyzanowski & Cohen, 2008).

Our dataset consists of hourly measurements of the six main air pollutants: Ozone (O₃), Nitrate Oxide (NO), Nitrate Dioxide (NO₂), Carbon Oxide (CO), Particulate Matter under 10 microns (PM₁₀) and Sulphur Dioxide (SO₂), from January 1, 2000 to April 13, 2002 at Bloomsbury, London, UK. Those ground-level measurements correspond to 20,000 entries and are communicated in $\mu\text{g}/\text{m}^3$.

We provide some empirical elements showing that one should primarily use the two-stage GMM approach M4 to correctly fit both the marginal distribution and the serial dependence structure as given by the auto-correlation function.

6.1 | Data preparation

We use the Loess function (R package `stl`) on each pollutant time series individually with a quarterly trend and a daily seasonality component and work on the remainder after removing those two quantities. This removes any seasonality in the auto-correlation function and provides centered time series around zero. Additionally, we observe some rounding present in the data which creates “bumps” in the auto-correlation function. Hence we jitter the data using some white noise with a variance set to 15% of the time series variance. To make the different time series comparable in scale, we standardize the exceedance time series to the unit variance.

Using a goodness-of-fit test for the GPD (Villaseñor-Alva & González-Estrada, 2009) with 2000 bootstrap samples (R package `gpdtest`), we find that 95%-quantiles for all pollutants but one are suitable thresholds as presented in Table 1 using Method M4. Carbon Monoxide (CO) requires to choose the 98%-quantile with a p -value of 28%.

6.2 | Discussion

We compare the performance of Methods M1–M4 on the air pollutant dataset using the exponent trawl, that is, exponential-LTME. Note that in this case, the serial dependence given by the auto-correlation function (ACF) is a decreasing function of the trawl parameter $\rho > 0$ and time given by $(t, \rho) \mapsto e^{-\rho t}$.

TABLE 1 Extreme thresholds and GoF p -values with parameter estimates yielded by M1–M4 with $G = 4$ on the whole dataset. Monte Carlo standard deviations in parenthesis: by a block bootstrap for M1 and M3; two-stage corrected standard deviations for ρ for M2 and M4 (Corollary 2 and HAC estimator)

Agent	Threshold	p -value	Method	ξ	σ	κ	ρ
O3	62.93	90%	M1	−.009 (.007)	.31 (.002)	24.36 (.10)	.111 (.001)
			M2	−.009 (.007)	.31 (.002)	19.96 (.10)	.111 (.002)
			M3	−.003 (.006)	.62 (.004)	16.24 (.13)	.074 (.001)
			M4	−.009 (.004)	.31 (.003)	19.96 (.25)	.072 (.023)
CO	00.82	28%	M1	.073 (.003)	0.59 (.004)	27.85 (.21)	.111 (.002)
			M2	.073 (.003)	0.59 (.004)	22.60 (.21)	.111 (.002)
			M3	.219 (.006)	1.18 (.012)	18.56 (.12)	.105 (.001)
			M4	.073 (.003)	0.59 (.010)	22.60 (.23)	.081 (.126)
NO	84.75	43%	M1	.033 (.010)	0.61 (.012)	78.07 (.62)	.147 (.002)
			M2	.033 (.010)	0.61 (.012)	78.07 (.62)	.147 (10^{-6})
			M3	.100 (.007)	1.23 (.002)	52.05 (.40)	.097 (.002)
			M4	.033 (.010)	0.61 (.001)	78.08 (.62)	.091 (.023)
NO2	92.27	28%	M1	.012 (.003)	0.28 (.001)	26.09 (.15)	.183 (.001)
			M2	.012 (.003)	0.28 (.001)	20.99 (.15)	.183 (.004)
			M3	.036 (.007)	0.53 (.005)	17.39 (.24)	.114 (.001)
			M4	.012 (.003)	0.28 (.009)	20.99 (.18)	.111 (.007)
PM10	52.12	6%	M1	.119 (.003)	0.47 (.004)	28.08 (.21)	.183 (.004)
			M2	.119 (.003)	0.72 (.004)	20.20 (.22)	.183 (.002)
			M3	.356 (.008)	0.77 (.010)	18.72 (.16)	.132 (.005)
			M4	.119 (.003)	0.75 (.012)	20.20 (.26)	.130 (.006)
SO2	13.53	8%	M1	.144 (.004)	0.50 (.007)	29.15 (.30)	.147 (.003)
			M2	.144 (.004)	0.50 (.007)	25.46 (.30)	.147 (10^{-3})
			M3	.289 (.008)	1.01 (.011)	19.43 (.16)	.106 (.002)
			M4	.144 (.005)	0.50 (.017)	25.46 (.32)	.101 (.032)

6.2.1 | Experimental setup and general observations

For each pollutant, we draw 200 bootstrap samples of 5000 observations using a block bootstrap. We then fit the exponential-LTME with a block depth $G = 4$ (Dupuis & Trapin, 2019; Noven et al., 2018) on each sample and present the results as boxplots in Figure 1 where dashed lines correspond to estimates on the whole dataset given by M2 and M4 and black bars correspond to the average of block-bootstrapped values. Figure 2 showcases the empirical ACF compared to the exponential-LTME ACF induced by the fitted parameters for M2 and M4. Finally, numerical values are presented in Table 1. Note that for M4, we take the weighting matrix $W_N = \mathbf{I}_4$.

In this section, we find that M4 provides an inference scheme that is easy to implement and understand, that performs similarly to PL-based approaches on the estimation of GPD marginal

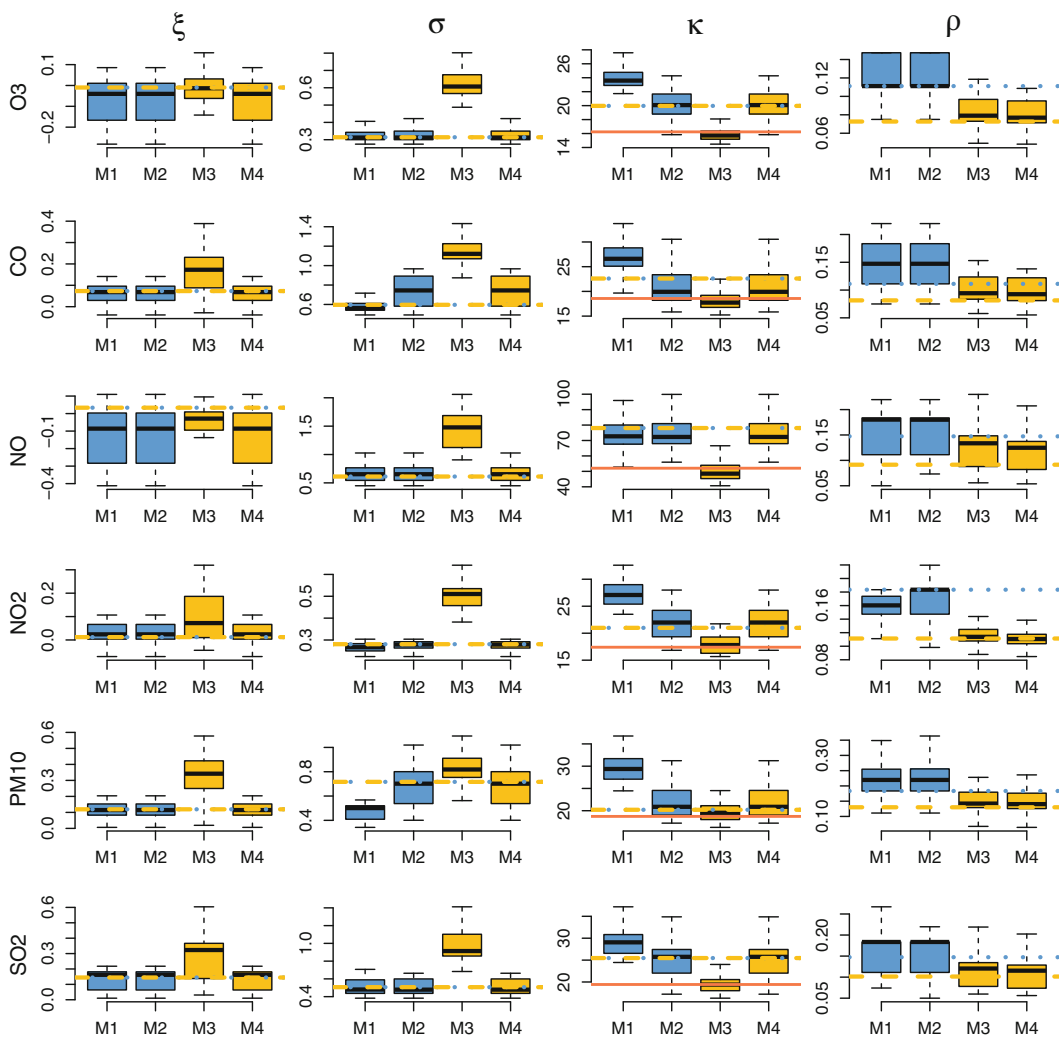


FIGURE 1 Comparing the four methods M1–M4: boxplots of the estimates of ξ , σ , κ and ρ . pairwise likelihood-based schemes (M1–M2) in blue; auto-covariance matching-based (M3–M4) approaches in gold; dashed lines correspond to M2 and M4 estimators on the full dataset. The brown solid line is the value of κ obtained by inverting $P(X_t > 0) = (1 + \kappa)^{-1}$

parameters and captures the serial dependence in the data. More precisely, we show that M4 provides similar marginal parameters compared to M2, as it should by definition, but estimates better the serial dependence present in the data. Although M1 yields GPD parameter estimates similar to M2 and M4, it overestimates κ by almost 50% from the value expected by the inversion of $P(X_t > 0) = (1 + \kappa)^{-1}$ and, hence, underestimates the probability of positive exceedance. On the other hand, M3 estimates κ remarkably well in that sense, and captures the serial dependence as well as M4 but its estimation of ξ and σ appears to be biased in most cases. From this, we conclude that M4 balances better the estimation of all four parameters in a simple estimation scheme.

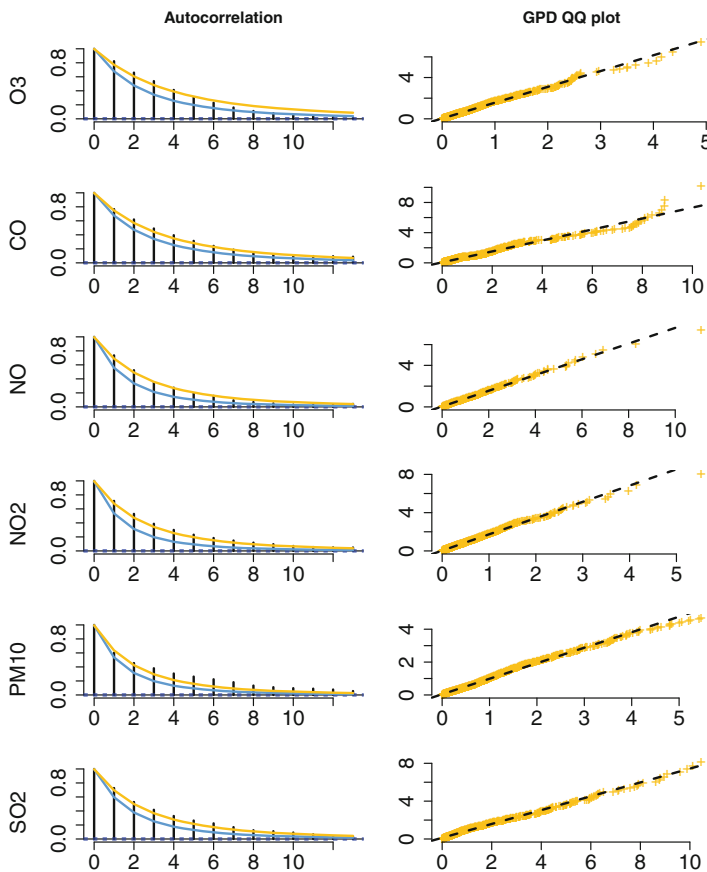


FIGURE 2 (left) Empirical and exponential-latent trawl model for extremes autocorrelation function with the trawl parameter values for M2 (blue) and M4 (gold); (right) generalized pareto distribution Q-Q plots for M4 estimator values

6.2.2 | GPD parameters

Across the six pollutants, M1, M2, and M4 showcase similar performances for the GPD marginal parameters (ξ , σ) on both the entire dataset and also on block-bootstrapped data, both on average and in terms of variability. For instance, the bootstrapped estimates of ξ and σ are closely centered around the estimates obtained on the whole dataset especially for O₃, NO₂, PM₁₀, and SO₂. However, the corresponding M3 estimates are two to three times as large: for example, for NO, ξ is estimated to be 0.033 ± 0.01 by M1, M2, and M4 against $\hat{\xi}_{M3} = 0.1 \pm 0.007$; σ is estimated to be 0.61 ± 0.012 for M1, M2, and M4 against $\hat{\sigma}_{M3} = 1.23 \pm 0.002$. In addition, estimates from M3 are more instable with interquartile ranges two to three times as large compared to the other approaches. In particular, this holds for all pollutants except CO and NO for ξ ; and except PM₁₀ for σ .

Note that M4 fits the GPD of positives exceedances correctly as shown by the M4 Q-Q plots in Figure 2. The Q-Q plots obtained with M1 and M2 are redundant and are not presented.

We remark the M3 bootstrapped estimates appear to be biased if we consider the M2 and M4 full-samples estimates as candidates true values. This potential bias would be positive for all

except the NO shape parameter ξ where it would be slightly negative. We believe that the difference in parameter sensitivity between UL log-likelihood and the AM moments condition could be a driving factor in the sense that the latter would dominate the gradient estimates. As the estimation of κ and ρ is somewhat separated from that of ξ and σ through the marginal transformation trick (see Section 6.2.3 below), the competing and equally weighted moments conditions balance the $h(\cdot)$ Jacobian and the UL log-likelihood. Although the estimates for κ and ρ are satisfactory (see below), this balance seems unreliable for the GPD parameters. A potential solution could be to use a more sophisticated first-stage weighting matrix, which is beyond the scope of this article and is left for future research.

6.2.3 | Inference of κ

Recall that the third parameter κ controls the probability of positive exceedances, and it can be estimated directly by $P(X_t > 0) = (1 + \kappa)^{-1}$ (Dupuis & Trapin, 2019), as shown by the solid brown line in the third column of Figure 1.

We observe a similar behavior as with the GPD parameters: the bootstrapped estimates from M2 and M4 coincide to the full-dataset estimates (except for NO & NO). M1 usually provides higher values: for example, for PM10, $\hat{\kappa}_{M1} = 28.08 \pm 0.21$ against 20.20 ± 0.22 from M2/M4.

However, with M1, M2, and M3, we observe that as $\hat{\sigma}$ grows larger, $\hat{\kappa}$ gets smaller. We explain this behavior by the dividing factor $(1 + \kappa)\sigma$ of the UL likelihood for positive exceedances: to remain at a given level, an increase in κ would certainly imply a decrease in σ and vice versa. We remark $\hat{\kappa}_{M3}$ is the closest of the four approaches to the value obtained by direct inversion. Consequently, M1, M2, and M4 feature a positive bias in κ and subsequently underestimate the probability of having a positive exceedance at any time: for example, for PM10 where $P(X_t > 0) = 5\%$, we have $(1 + \hat{\kappa}_{M3})^{-1} \approx 0.0507$ against $(1 + \hat{\kappa}_{M1})^{-1} \approx 0.0331$. We conjecture that this bias is caused for different reasons: for M1—the scheme for which the bias seems the strongest—we believe it is rooted in the interplay between ρ and κ in the PL. The negative bias in ρ (see Section 6.2.4) directly impacts κ as the marginal transformation trick separates ξ and σ from this conflict. In the two-stage schemes M2 and M4, the aforementioned UL balancing effect is not affected by the PL or AM term in contrary to M1 and M3. This interplay is likely to help settle κ to capture the serial dependence and thus the estimation of κ relies on how well the scheme captures those dependencies. The absence of interplay, as in M2 and M4, exposes a problem arising from a time series in a 19-to-1 ratio between zeros and positive exceedances. As the probability of zero is given by $\kappa/(1 + \kappa)$, overestimating κ increases the relative preponderance of zeros in the likelihood function and diminishes that of nonzero values. Thus, this may maximize the likelihood by simply bypassing the likelihood of positive exceedances. This said, recall that κ only appears as the factor $1/(1 + \kappa)$ in the UL. At this particular scale, a difference of 1% to 2% as suggested above can be attributed to the finite sample estimation error of the marginal distribution.

To conclude for κ , M3 appears to be the best-performing approach and we attribute this to the interaction between s_{UL} and s_{AM} in the moments condition.

6.2.4 | Serial dependence

Concerning the trawl parameter, the bootstrapped estimates of ρ and full-dataset ones match in each case except again, for NO. However, the M3 and M4 yield estimates which replicate the

empirical ACF more closely as opposed to M1 and M2 that both underestimate it (as shown for M2 in the left-hand column, Figure 2). Indeed, M1 and M2 perform very similarly (the fitted M1 ACF are omitted as they are indistinguishable from that of M2) and overestimate the parameter value of ρ (last column, in Figure 1 and Table 1) leading to underestimation of the serial dependence given by $t \mapsto \exp(-\rho t)$. In addition, bootstrapped estimates from M1 and M2 are either right-skewed or left-skewed: the bootstrapped average is placed at the first (O3) or third quartile (NO, NO2, SO2), see the last column in Figure 1.

The better behavior of M3 and M4 is unexpected as we only perform the AM on the first four lags ($G = 4$) and we expect a more erratic behavior compared to the PL-based methods. The ability to capture the serial dependence in the data from a few lags complements the satisfactory GPD fit of M4 mentioned above.

6.3 | Closing remark

In this empirical study, we have compared the performance of the four methods through a block bootstrap given we cannot simulate from the LTME model. Regarding the GPD parameters (ξ, σ), all three methods M1, M2, and M4 showcase similar quality of estimation while M3 provides a much more volatile inference of σ . The parameter controlling the probability of positive exceedances, namely κ , is best estimated by M3 while M1 is the worst-performing scheme. Finally both AM-based approaches well captured the serial dependence of the data while M1 and M2 underestimate it with a heavy skew in the block-bootstrapped distribution. In this mixed conclusion, we find that M4 to be the more balanced approach: it provides a good marginal fit, it captures serial dependence through the acf and comprises two simple steps.

7 | CONCLUSION

In this article, we compare the inference strategies of a LTME values. We find that the established PL approaches fail to capture the second-order dependence structure present in the empirical data. Although a single-stage PL scheme was preferred historically (Noven et al., 2018), a two-stage approach was recently proposed (Dupuis & Trapin, 2019) where we first fit the parameters relating to the marginal distribution before maximizing the PL in a second stage to fit the dependence structure. We tweak these approaches by capturing the serial dependence through an AM procedure, either as a single- or two-stage method. We prove the asymptotic convergence of all four strategies. Those convergences are provided in the sense of weakly approaching sequences of distributions, a stronger result compared to the convergence in distribution. This convergence mode is particularly suited for strongly mixing sequences and gives the explicit relationship between the parameter estimators and a potential sequence approximating the limiting distribution.

Next, we compare the performance of these approaches on hourly measurements of six air pollutant concentrations in London, UK. By comparing their performance on bootstrapped data, the two-stage AM procedure successfully fits both the marginal distributions and the serial dependence structure. These properties advocate for the use of interpretable inference steps without compromise on the estimation accuracy—a key property for extreme value models (Davison & Smith, 1990).

Hierarchical modeling of extremes offers flexible dependence structures suitable for many applications. Recently a family of hierarchal models for extremes arose, all sharing a similar

mixture structure based on a Gamma-distributed parameter. For example, spatial extensions of the temporal trawl process proposed by Opitz (2017) are leveraged to model exceedances in both time and space in Bacro et al. (2020). Alternatively, the Exponential-Gamma mixture has been transformed into a Gamma-Gamma mixture in Yadav et al. (2021) providing different modeling flexibility of the marginal distribution. Although those models are supported by strong empirical evidence, their asymptotic inference behavior remains an open problem and could similarly benefit from multistage schemes.

ACKNOWLEDGMENTS

We thank the Editors and both reviewers for their insightful comments that greatly improved the presentation of this article. Valentin Courgeau would like to thank the Centre for Doctoral Training in Financial Computing and Analytics for providing funding for this research (under the grant EP/L015129/1).

ENDNOTE

¹We write α^* to avoid any confusion with the Gamma distribution shape parameter α used throughout the article, see Section 2.3.2.

ORCID

Valentin Courgeau  <https://orcid.org/0000-0002-7862-1363>

Almut E.D. Veraart  <https://orcid.org/0000-0001-8582-3652>

REFERENCES

- Bacro, J.-N., Gaetan, C., Opitz, T., & Toulemonde, G. (2020). Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. *Journal of the American Statistical Association*, 115(530), 555–569. <https://doi.org/10.1080/01621459.2019.1617152>
- Bader, B., Yan, J., & Zhang, X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *The Annals of Applied Statistics*, 12(1), 310–329.
- Balkema, A. A., & de Haan, L. (1974). Residual life time at great age. *Annals of Probability*, 2(5), 792–804. <https://doi.org/10.1214/aop/1176996548>
- Barndorff-Nielsen, O. E. (2011). Stationary infinitely divisible processes. *Brazilian Journal of Probability and Statistics*, 25(3), 294–322.
- Barndorff-Nielsen, O. E., Lunde, A., Shephard, N., & Veraart, A. E. (2014). Integer-valued trawl processes: A class of stationary infinitely divisible processes. *Scandinavian Journal of Statistics*, 41(3), 693–724.
- Belyaev, Y., & Luna, S. S.-D. (2000). Weakly approaching sequences of random distributions. *Journal of Applied Probability*, 37(3), 807–822. <http://www.jstor.org/stable/3215615>
- Bortot, P., & Gaetan, C. (2014). A latent process model for temporal extremes. *Scandinavian Journal of Statistics*, 41(3), 606–621.
- Bortot, P., & Gaetan, C. (2016). Latent process modelling of threshold exceedances in hourly rainfall series. *Journal of Agricultural, Biological and Environmental Statistics*, 21(3), 531–547. <https://doi.org/10.1007/s13253-016-0254-5>
- Carter, A. V., Schnepel, K. T., & Steigerwald, D. G. (2017). Asymptotic behavior of a t-test robust to cluster heterogeneity. *The Review of Economics and Statistics*, 99(4), 698–709. https://doi.org/10.1162/REST_a_00639
- Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 52(3), 393–442.
- Davydov, Y. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory of Probability and Its Applications*, 13(4), 691–696.
- Dupuis, D. J., & Trapin, L. (2019). Ground-level ozone: Evidence of increasing serial dependence in the extremes. *The Annals of Applied Statistics*, 13(1), 34–59. <https://doi.org/10.1214/18-AOAS1183>
- Ekström, M. (2014). A general central limit theorem for strong mixing sequences. *Statistics & Probability Letters*, 94, 236–238. <http://www.sciencedirect.com/science/article/pii/S01677152%14002624>

- Francq, C., & Zakoïan, J.-M. (2013). Estimating the marginal law of a time series with applications to heavy-tailed distributions. *The Journal of Business and Economic Statistics*, 31(4), 412–425. <https://doi.org/10.1080/07350015.2013.801776>
- Fuchs, F., & Stelzer, R. (2013). ‘Mixing conditions for multivariate infinitely divisible processes with an application to mixed moving averages and the supOU stochastic volatility model. *ESAIM: Probability and Statistics*, 17, 455–471.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics*, 44(3), 423–453.
- Gumbel, E. J. (2004). *Statistics of extremes* (Vol. 4, pp. 113–155). Courier Corporation.
- Hansen, B. E., & Lee, S. (2019). Asymptotic theory for clustered samples. *Journal of Econometrics*, 210(2), 268–290. <http://www.sciencedirect.com/science/article/pii/S0304407619300247>
- Hansen, B. E., & Lee, S. (2021). Inference for iterated GMM under misspecification. *Econometrica*, 89(3), 1419–1447. <https://doi.org/10.3982/ECTA16274>
- Hogg, R. V., & Klugman, S. A. (1983). On the estimation of long tailed skewed distributions with actuarial applications. *Journal of Econometrics*, 23(1), 91–102.
- Hwang, J. (2020). *Supplement to: Simple and trustworthy cluster-robust GMM inference* [Working papers 2017–19]. University of Connecticut, Department of Economics. <https://ideas.repec.org/p/uct/uconnp/2017-19.html>
- Hwang, J. (2021). Simple and trustworthy cluster-robust GMM inference. *Journal of Econometrics*, 222(2), 993–1023. <https://www.sciencedirect.com/science/article/pii/S0304407620303043>
- Ibragimov, I., Linnik, Y., & Kingman, J. (1971). *Independent and stationary sequences of random variables*. Monographs and Textbooks on Pure and Applied Mathematics, Wolters-Noordhoff.
- Krzyzanowski, M., & Cohen, A. (2008). Update of who air quality guidelines. *Air Quality, Atmosphere & Health*, 1(1), 7–13.
- Newey, W. K., & McFadden, D. (1994). *Ch. 36 Large sample estimation and hypothesis testing*. In *Handbook of econometrics* (Vol. 4, pp. 2111–2245). Elsevier. <https://www.sciencedirect.com/science/article/pii/S1573441205800054>
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708. <http://www.jstor.org/stable/1913610>
- Noven, R., Veraart, A., & Gandy, A. (2018). A latent trawl process model for extreme values. *Journal of Energy Markets*, 11, 1–24.
- Opitz, T. (2017). Spatial random field models based on Lévy indicator convolutions. arXiv preprint arXiv:1710.06826.
- Pickands, J., III. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1), 119–131. <https://doi.org/10.1214/aos/1176343003>
- Rajput, B. S., & Rosinski, J. (1989). Spectral representations of infinitely divisible processes. *Probability Theory and Related Fields*, 82(3), 451–487.
- Smith, R., & Miller, J. (1986). A non-Gaussian state space model and application to prediction of records. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 48(1), 79–88.
- Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1), 67–90 <http://www.jstor.org/stable/2336336>
- Villaseñor-Alva, J. A., & González-Estrada, E. (2009). A bootstrap goodness of fit test for the generalized Pareto distribution. *Computational Statistics & Data Analysis*, 53(11), 3835–3841.
- Yadav, R., Huser, R., & Opitz, T. (2021). Spatial hierarchical modeling of threshold exceedances using rate mixtures. *Environmetrics*, 32(3), e2662. <https://doi.org/10.1002/env.2662>

How to cite this article: Courgeau, V., & Veraart, A. E. (2022). Asymptotic theory for the inference of the latent trawl model for extreme values. *Scandinavian Journal of Statistics*, 49(4), 1448–1495. <https://doi.org/10.1111/sjos.12563>

APPENDIX A. LÉVY MEASURES AND INFINITE DIVISIBILITY

Definition 6. A **random measure** M on $(S, \mathcal{B}(S))$ is a collection of $\mathbb{R}$ -valued random variables in $\{M(A) : A \in \mathcal{B}_b(S)\}$ such that for any sequence $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{B}_b(S)$ of disjoint elements satisfying $\cup_{n \in \mathbb{N}} A_n \in \mathcal{B}_b(S)$, we have $M(\cup_{n \in \mathbb{N}} A_n) = \cup_{n \in \mathbb{N}} M(A_n)$ a.s.

Additional properties are defined as follows:

- M is **stationary** if it is stable by translation, that is, if for any $\mathbf{s} \in S$, $n \in \mathbb{N}$ and any $A_1, \dots, A_n \in \mathcal{B}_b(S)$ such that $A_i + \mathbf{s} \subseteq S$ for $i = 1, \dots, n$, we have $(M(A_1 + \mathbf{s}), \dots, M(A_n + \mathbf{s})) \stackrel{d}{=} (M(A_1), \dots, M(A_n))$, where $\stackrel{d}{=}$ is the equality in law.
- M is **independently scattered** if for any sequence $(A_n, n \in \mathbb{N}) \subseteq \mathcal{B}_b(S)$ of disjoint elements, $(M(A_n), n \in \mathbb{N})$ is a collection of mutually independent random variables.
- M is **infinitely divisible**, or ID, if and only if for any $M(A) \in M$ with law μ and any $n \in \mathbb{N}$, there exists a law μ_n on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mu = \mu_n^{*n}$ where μ_n^{*n} denotes the n -fold convolution of μ_n on itself.
- A **Lévy basis** L on $(S, \mathcal{B}(S))$ is an independently scattered, infinitely divisible random measure. Additionally, it is **homogeneous** if it is stationary.
- A **Lévy measure** is any Borel measure ν on $\mathbb{R}$ such that $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}} \min(1, |x|^2) \nu(dx) < \infty$.

APPENDIX B. PL FUNCTION

We recall the PL function derived in appendix A.1 of Noven et al. (2018). For $s < t \in \mathbb{R}$, recall the notations of $b_{s \setminus t}$ and $b_{s,t}$ as in (5). If $X_s = 0$ and $X_t = 0$ we have

$$f_{\text{PL}}(0, 0) = 1 - 2 \left(1 + \frac{\kappa}{\beta}\right)^{-\alpha} + \left(1 + \frac{\kappa}{\beta}\right)^{2b_{s \setminus t} - 1} \left(1 + \frac{2\kappa}{\beta}\right)^{b_{s,t}}.$$

Alternatively, for $X_s = x_s > 0$ and $X_t = 0$:

$$f_{\text{PL}}(x_s, 0) = \frac{\alpha}{\beta} \left(1 + \frac{\kappa + x_s}{\beta}\right)^{-\alpha-1} + \left(1 + \frac{\kappa + x_s}{\beta}\right)^{b_{s \setminus t} - 1} \left(1 + \frac{2\kappa + x_s}{\beta}\right)^{b_{s,t} - 1} \left(1 + \frac{\kappa}{\beta}\right)^{b_{s \setminus t} - 1} \\ \times \left[\frac{b_{s \setminus t}}{\beta} - \alpha \left(1 + \frac{\kappa + x_s}{\beta}\right) \right].$$

Note that the case $X_s = 0$ and $X_t = x_t > 0$ is similar. Finally, we have for $X_s = x_s > 0$ and $X_t = x_t > 0$:

$$f_{\text{PL}}(x_s, x_t) = \beta^{-2} \left(1 + \frac{2\kappa + x_s + x_t}{\beta}\right)^{b_{s,t} - 2} \left(1 + \frac{\kappa + x_s}{\beta}\right)^{b_{s \setminus t} - 1} \left(1 + \frac{2\kappa + x_t}{\beta}\right)^{b_{s \setminus t} - 1} \\ \times \left[b_{s,t}(b_{s \setminus t} - 1) \left(1 + \frac{\kappa + x_s}{\beta}\right) \left(1 + \frac{\kappa + x_t}{\beta}\right) + b_{s \setminus t}^2 \left(1 + \frac{2\kappa + x_s + x_t}{\beta}\right)^2 \right. \\ \left. b_{s,t} b_{s \setminus t} \left(1 + \frac{2\kappa + x_s + x_t}{\beta}\right) \left\{ 1 + \frac{\kappa + x_s}{\beta} + 1 + \frac{\kappa + x_t}{\beta} \right\} \right].$$

APPENDIX C. STRONG MIXING PROPERTY FOR TRAWLS AS LATENT PROCESS

The proof of Proposition 3 is presented below

Proof of Proposition 3. In this proof, we use repeatedly the conditional independence assumption from Equation (2). One derives the marginal density function for $k > 0$ as follows (Noven et al. (2018), equation (11)):

$$f^{(1)}(x_i) = \begin{cases} \int_{(0,\infty)} (1 - e^{-\kappa \lambda_i}) dF(\lambda_i), & \text{if } x_i = 0, \\ \int_{(0,\infty)} J_h(x_i) \lambda_i e^{-(\kappa+h^{-1}(x_i))\lambda_i} dF(\lambda_i), & \text{if } x_i > 0, \end{cases}$$

and the bivariate density function:

$$f^{(2)}(x_i, x_{i+k}) = \int_{(0,\infty)^2} r(x_i, x_{i+k}, \lambda_i, \lambda_{i+k}; \kappa) dF(\lambda_i, \lambda_{i+k}),$$

where

$$r(x_i, x_{i+k}, \lambda_i, \lambda_{i+k}; \kappa) := \prod_{\substack{j \in \{i, i+k\} \\ x_j = 0}} (1 - e^{-\kappa \lambda_j}) \prod_{\substack{j \in \{i, i+k\} \\ x_j > 0}} (J_h(x_j) \lambda_j e^{-(\kappa+h^{-1}(x_j))\lambda_j}), \quad (C1)$$

as a function of $(x_i, x_{i+k}, \lambda_i, \lambda_{i+k})$.

Let $S'_i \in \mathcal{F}_{-\infty}^i(X)$ and $S'_{i+k} \in \mathcal{F}_{i+k}^{+\infty}(X)$ be nonempty sets. To simplify the notation, we set $S_i := X_i^{-1}(S'_i)$ and $S_{i+k} := X_{i+k}^{-1}(S'_{i+k})$, the preimages of S'_i and S'_{i+k} with respect to X_i and X_{i+k} , respectively. (In general, one could extend to preimages of any $Y \in \mathcal{F}_{-\infty}^i(X)$ and any $Z \in \mathcal{F}_{i+k}^{+\infty}(X)$, extending the calculations presented in Proposition 10 and its proof in Appendix F.0.4 where we present the result with two pairs of indices (i_1, i_2) and (i_3, i_4) such that $i_1 < i_2 \leq i$ and $i + k \leq i_3 < i_4$.) Remark that

$$\int_{S_i \times S_{i+k}} r(x_i, x_{i+k}, \lambda_i, \lambda_{i+k}; \kappa) dx_i dx_{i+k} \leq \int_{(0, x_F)^2} \prod_{j \in \{i, i+k\}} [J_h(x_j) \lambda_j e^{-(\kappa+h^{-1}(x_j))\lambda_j}] dx_i dx_{i+k}.$$

Let $j \in \{i, i+k\}$ and observe that

$$J_h(x) \lambda_j e^{-(\kappa+h^{-1}(x))\lambda_j} = \frac{1+\kappa}{\sigma} \lambda_j \left(1 + \frac{\xi x}{\sigma}\right)^{\frac{1}{\xi}-1} \exp \left\{ - \left[(1+\kappa) \left(1 + \frac{\xi x}{\sigma}\right)^{\frac{1}{\xi}} - 1 \right] \lambda_j \right\}.$$

Using the change of variable $y := (1 + \xi x/\sigma)^{1/\xi}$, we obtain

$$\begin{aligned} \int_{(0, x_F)} J_h(x) \lambda_j e^{-(\kappa+h^{-1}(x))\lambda_j} dx &= \int_{(1,\infty)} (1+\kappa) \lambda_j e^{\lambda_j} e^{-(1+\kappa)\lambda_j y} dy \\ &= [-e^{\lambda_j} e^{-(1+\kappa)\lambda_j y}]_1^{(1+\xi x_F/\sigma)^{1/\xi}} = e^{-\kappa \lambda_j}. \end{aligned}$$

Therefore, we have the following inequality

$$\int_{S_i \times S_{i+k}} r(x_i, x_{i+k}, \lambda_i, \lambda_{i+k}; \kappa) dx_i dx_{i+k} \leq \prod_{j \in \{i, i+k\}} e^{-\kappa \lambda_j}. \quad (C2)$$

Also, by Fubini's theorem, we have

$$f^{(1)}(x_i)f^{(1)}(x_{i+k}) = \int_{(0,\infty)^2} r(x_i, x_{i+k}, \lambda_i, \lambda_{i+k}; \kappa) dF(\lambda_i) dF(\lambda_{i+k}).$$

The mixing properties X are solely determined by those of the latent process Λ . Indeed, one has

$$\begin{aligned} & P(S'_i \cap S'_{i+k}) - P(S'_i)P(S'_{i+k}) \\ &= \int_{S_i \times S_{i+k}} f^{(2)}(x_i, x_{i+k}) dx_i dx_{i+k} - \int_{S_i} f^{(1)}(x_i) dx_i \int_{S_{i+k}} f^{(1)}(x_{i+k}) dx_{i+k} \\ &= \int_{S_i \times S_{i+k}} [f^{(2)}(x_i, x_{i+k}) - f^{(1)}(x_i)f^{(1)}(x_{i+k})] dx_i dx_{i+k} \\ &= \int_{S_i \times S_{i+k}} \int_{(0,\infty)^2} r(x_i, x_{i+k}, \lambda_i, \lambda_{i+k}; \kappa) [dF(\lambda_i, \lambda_{i+k}) - dF(\lambda_i)dF(\lambda_{i+k})] dx_i dx_{i+k}. \end{aligned}$$

Then, by taking the triangle inequality and Equation (C2):

$$\begin{aligned} & |P(S'_i \cap S'_{i+k}) - P(S'_i)P(S'_{i+k})| \\ & \leq \left| \int_{(0,\infty)^2} \left(\prod_{j \in \{i, i+k\}} e^{-\kappa \lambda_j} \right) \{dF(\lambda_i, \lambda_{i+k}) - dF(\lambda_i)dF(\lambda_{i+k})\} \right| \\ & \leq \alpha^*(\Lambda, k), \quad \text{since } \Lambda_t \geq 0 \text{ a.s. for all } t \geq 0. \end{aligned}$$

This yields that the strong mixing property for X as one has $\alpha(X, k) \leq \alpha(\Lambda, k) \xrightarrow{k \rightarrow \infty} 0$. Additionally, one can refine the computation as follows

$$\begin{aligned} & |P(S'_i \cap S'_{i+k}) - P(S'_i)P(S'_{i+k})| \\ & \leq \left| \int_{(0,\infty)^2} \prod_{j \in \{i, i+k\}} e^{-\kappa \lambda_j} dF(\lambda_i, \lambda_{i+k}) - \int_{(0,\infty)^2} \prod_{j \in \{i, i+k\}} e^{-\kappa \lambda_j} dF(\lambda_i) dF(\lambda_{i+k}) \right| \\ & \leq \left| \mathbb{E}(e^{-\kappa(\Lambda_i + \Lambda_{i+k})}) - \mathbb{E}(e^{-\kappa \Lambda_i}) \mathbb{E}(e^{-\kappa \Lambda_{i+k}}) \right|. \end{aligned}$$

The independent scattering property of trawls yields that for any $s, t \geq 0$, $\Lambda_t = L(A_t) \stackrel{d}{=} L(A_t \setminus A_s) + L(A_s \cap A_t) =: \Lambda_{t \setminus s} + \Lambda_{s,t}$ (respectively) where $\Lambda_{t \setminus s}$ and $\Lambda_{s,t}$ are independent. Thus, using these and the notations from Equation (5) along with the strict stationarity and Gamma distribution assumptions, we find that

$$\begin{aligned} & |P(S'_i \cap S'_{i+k}) - P(S'_i)P(S'_{i+k})| \\ & \leq \left| \mathbb{E}(e^{-\kappa \Lambda_{i+k \setminus i}}) \mathbb{E}(e^{-2\kappa \Lambda_{i,i+k}}) \mathbb{E}(e^{-\kappa \Lambda_{i \setminus i+k}}) - \mathbb{E}(e^{-\kappa \Lambda_{i \setminus i+k}}) \mathbb{E}(e^{-\kappa \Lambda_{i+k \setminus i}}) \mathbb{E}(e^{-\kappa \Lambda_{i,i+k}}) \right| \\ & \leq \left| \left(1 + \frac{\kappa}{\beta}\right)^{b_{k \setminus 0}} \left(1 + \frac{2\kappa}{\beta}\right)^{b_{0,k}} \left(1 + \frac{\kappa}{\beta}\right)^{b_{0 \setminus k}} - \left(1 + \frac{\kappa}{\beta}\right)^{b_{k \setminus 0}} \left(1 + \frac{\kappa}{\beta}\right)^{2b_{0,k}} \left(1 + \frac{\kappa}{\beta}\right)^{b_{0 \setminus k}} \right| \\ & \leq \left(1 + \frac{\kappa}{\beta}\right)^{b_{k \setminus 0}} \left(1 + \frac{\kappa}{\beta}\right)^{b_{0 \setminus k}} \left(1 + \frac{2\kappa}{\beta}\right)^{b_{0,k}} \left\{ 1 - \left(1 + \frac{2\kappa}{\beta}\right)^{b_{0,k}} \right\} \end{aligned}$$

$$\leq \left(1 + \frac{\kappa}{\beta}\right)^{b_{k \setminus 0}} \left(1 + \frac{\kappa}{\beta}\right)^{b_{0 \setminus k}} \left(1 + \frac{2\kappa}{\beta}\right)^{b_{0,k}} \text{Cor}(\Lambda_0, \Lambda_k).$$

Hence, we obtain that $\alpha^*(X, k) = O(\text{Cor}(\Lambda_0, \Lambda_k))$ which concludes the proof. $\blacksquare$

APPENDIX D. PROOFS FOR THE UL

D.1 Proof of Proposition 4

Proof of Proposition 4. Throughout the proof, we use the change of variable $y := 1 + \xi x / \sigma$ for $x > 0$ if $\xi \neq 0$ and $y = x / \sigma$ if $\xi = 0$. If $\xi < 0$, the GPD support is $(0, -\sigma / \xi)$ and the transformed variable y sits in $(0, 1)$ P-a.s.. For $\xi > 0$, the transformed variable y belongs to $(1, \infty)$ P-a.s. and in $(0, \infty)$ for $\xi = 0$. The condition $\mathbb{E}[|\ln f_{\text{UL}}(X_i; \xi, \sigma, \kappa)|] < \infty$ holds for any $\xi > -1/2$: for $\xi < 0$, the upper-bound $\int_0^1 \ln(y) y^{-\xi-1} dy$ is finite if $\xi > -1/2$. For $\xi = 0$, it holds since $\int_1^\infty \ln(y) e^{-y} dy < \infty$. For $\xi > 0$, an upper-bound is $\int_1^\infty \ln(y) y^{-\xi-1} dy$ which is also finite for any $\xi > 0$. Next, we prove the integrability of the score function and have

$$\nabla_{(\xi, \sigma, \kappa)} \ln f_{\text{UL}}(x; \xi, \sigma, \kappa) = \begin{pmatrix} \left[\xi^{-2} \ln(1 + \xi x / \sigma) + \frac{1+\xi}{\xi \sigma} x (1 + \xi x / \sigma)^{-1} \right] \mathbb{I}_{(0, x_F)}(x) \\ \left[-\sigma^{-1} + \frac{1+\xi}{\xi \sigma^2} x (1 + \xi x / \sigma)^{-1} \right] \mathbb{I}_{(0, x_F)}(x) \\ -(1 + \kappa)^{-1} \mathbb{I}_{(0, x_F)}(x) + (1 + \kappa)^{-1} \kappa^{-1} \mathbb{I}_{\{0\}}(x) \end{pmatrix},$$

where $x_F = -\sigma / \xi$ if $\xi < 0$ and $x_F = \infty$ if $\xi \geq 0$. By the triangle inequality, we obtain

$$\begin{aligned} \mathbb{E}(\|s_{\text{UL}}(\xi_0, \sigma_0, \kappa_0)\|^2) &= \mathbb{E}(\|\nabla_{(\xi, \sigma, \kappa)} \ln f_{\text{UL}}(X_i; \xi, \sigma, \kappa)\|^2) \\ &\leq C_1 \int_0^{x_F} \ln\left(1 + \frac{\xi_0 x}{\sigma_0}\right)^2 \left(1 + \frac{\xi_0 x}{\sigma_0}\right)^{-\xi_0-1} dx \\ &\quad + C_2 \int_0^{x_F} \left(1 + \frac{\xi_0 x}{\sigma_0}\right)^{-\xi_0-3} dx + C_3, \end{aligned}$$

for some finite constants $C_i := C_i(\xi_0, \sigma_0, \kappa_0) \in \mathbb{R}$, $i \in \{1, 2, 3\}$. For negative values of ξ_0 , the condition $\mathbb{E}(\|s_{\text{UL}}(\xi_0, \sigma_0, \kappa_0)\|^2) < \infty$ holds if both $\int_0^1 \ln(y)^2 y^{-\xi_0-1} dy < \infty$ and $\int_0^1 (1-y)^2 y^{-\xi_0-3} dy < \infty$ are satisfied. The latter being more restrictive as it holds for any $\xi_0 > -1/2$ by definition of the Beta distribution while the former is satisfied for any $\xi_0 > -1$. On the other hand, if ξ_0 is positive then the same condition holds if $\int_1^\infty \ln(y)^2 y^{-\xi_0-1} dy < \infty$ and $\int_1^\infty (y-1)^2 y^{-\xi_0-3} dy < \infty$ which are both satisfied for any positive ξ_0 . Next, the last condition

$$\mathbb{E}\left(\sup_{(\xi, \sigma, \kappa) \in \mathcal{N}_0} \|\nabla_{(\xi, \sigma, \kappa)} s_{\text{UL}}(\xi, \sigma, \kappa)\|\right) < \infty,$$

holds by noticing that ξ is the only parameter impacting the integrability of the k th-order partial derivatives of UL for $k \geq 1$. Conditionally on X_i being positive, repeating a similar argument as given above yields that the second-order derivatives of the log UL likelihood evaluated at $(\xi_0, \sigma_0, \kappa_0)$ are finite if

$$\int_0^{x_F} |\ln(1 + \xi x/\sigma)| (1 + \xi x/\sigma)^{-\xi^{-1}-1} dx < \infty \quad \text{and} \quad \int_0^{x_F} |x|^l (1 + \xi x/\sigma)^{-\xi^{-1}-1-l} dx < \infty,$$

for $l \in \{1, 2\}$, that is, if $\xi_0 > -1/2$; this coincides with the classical result that the GPD Fisher information matrix is finite if $\xi_0 > -1/2$ (Smith, 1985, p. 89). If X_i is zero, the probability of the unique UL likelihood singleton at $x = 0$ is independent from the GPD marginal parameters ξ_0 and σ_0 and is infinitely differentiable w.r.t. κ hence finite. Therefore, by infinite differentiability of the UL likelihood w.r.t. its parameters, we conclude that $\mathbb{E} \left(\sup_{(\xi, \sigma, \kappa) \in \mathcal{N}_0} \|\nabla_{(\xi, \sigma, \kappa)} s_{\text{UL}}(\xi, \sigma, \kappa)\| \right) < \infty$ if $\xi_0 > -1/2$. ■

D.2 Proof of Proposition 6

Proof. We verify that Conditions A*1, A*3, and A*4, and A2 and A5 from Francq and Zakořan (2013) hold: Condition A*1 holds since UL is continuous in (ξ, σ, κ) . Condition A2 requires that, for $X \sim \text{UL}(\xi_0, \sigma_0, \kappa_0)$, we have $f_{\text{UL}}(X; \xi, \sigma, \kappa) = f_{\text{UL}}(X; \xi_0, \sigma_0, \kappa_0)$ a.s. only if $(\xi, \sigma, \kappa) = (\xi_0, \sigma_0, \kappa_0)$. The singleton at $\{0\}$ yields that $\kappa = \kappa_0$ whilst, by taking $x \downarrow 0$, we obtain that $\sigma = \sigma_0$. Then, the GPD support (i.e., $(0, x_F)$ with $x_F = -\sigma_0/\xi$ if $\xi < 0$ and $x_F = \infty$ otherwise) implies that ξ and ξ_0 are both negative or both nonnegative for the almost-sure equality of the GPD densities to hold. We write $x_F(\xi) = -\sigma/\xi$ and suppose that $-1/3 < \xi < \xi_0 < 0$ without loss of generality. The set $(x_F(\xi), x_F(\xi_0))$ has a positive Lebesgue measure and is included in the support of $f(\cdot; \xi_0, \sigma_0, \kappa_0)$ such that $1 = \int_0^{x_F(\xi)} f(x; \xi, \sigma_0, \kappa_0) dx > \int_0^{x_F(\xi)} f(x; \xi_0, \sigma_0, \kappa_0) dx$ which contradicts the almost-sure equality that imposes that $\int_0^{x_F(\xi)} f(x; \xi, \sigma_0, \kappa_0) dx = \int_0^{x_F(\xi)} f(x; \xi_0, \sigma_0, \kappa_0) dx$. Thus, $\xi = \xi_0 < 0$. The case $\xi_0 = 0$ implies that $1 = \lim_{x \rightarrow \infty} \sigma_0(\xi^{-1} + 1) \ln(1 + \xi x/\sigma_0)/x$. This limit is zero if $\xi \neq 0$, hence $\xi = \xi_0 = 0$. Finally, if $\xi_0 > 0$ then it holds that

$$1 = \lim_{x \rightarrow \infty} \frac{(\xi^{-1} + 1) \ln(1 + \xi x/\sigma_0)}{(\xi_0^{-1} + 1) \ln(1 + \xi_0 x/\sigma_0)} = \frac{1 + \xi}{1 + \xi_0} \Leftrightarrow \xi = \xi_0,$$

and Condition A2 is satisfied. Condition A*3 and all but two (sub)assumptions of Condition A*4 are given by Proposition 4. The first remaining condition is that

$$\mathbb{E} \left[\sup_{(\xi, \sigma, \kappa) \in \mathcal{N}_0} \left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln f_{\text{UL}}(X_i; \xi, \sigma, \kappa) \right| \right] < \infty,$$

for any $\theta_i, \theta_j, \theta_k \in \{\xi, \sigma, \kappa\}$. It is proved to hold for $\xi_0 > -1/3$ by proceeding similarly to the proof of Prop. 4. Indeed, UL is infinitely differentiable and the third-order derivatives of the log UL likelihood evaluated at $(\xi_0, \sigma_0, \kappa_0)$ have finite moments if

$$\int_0^{x_F} \left| \ln \left(1 + \frac{\xi_0 x}{\sigma_0} \right) \right| \left(1 + \frac{\xi_0 x}{\sigma_0} \right)^{-\xi_0^{-1}-1} dx < \infty \quad \text{and} \quad \int_0^{x_F} \left(1 + \frac{\xi_0 x}{\sigma_0} \right)^{-\xi_0^{-1}-3} dx < \infty,$$

that is, if $\xi_0 > -1/3$. Since $\mathbb{E}(s_{\text{UL}}(X_0; \xi_0, \sigma_0, \kappa_0)) = 0$, the last assumption of Condition A4, Francq and Zakořan (2013), is equivalent to

$$\sum_{k \in \mathbb{Z}} |\text{Cov}(s_{\text{UL}}(X_0; \xi_0, \sigma_0, \kappa_0), s_{\text{UL}}(X_k; \xi_0, \sigma_0, \kappa_0))| < \infty. \quad (\text{D1})$$

By stationary and since $x \mapsto s_{UL}(x; \xi_0, \sigma_0, \kappa_0)$ is measurable, each member is the upper-bounded (Davydov, 1968) by

$$12 \times \alpha^*(X, k)^{\delta/(2+\delta)} \times \mathbb{E}(|s_{UL}(\xi_0, \sigma_0, \kappa_0)|^{2+\delta})^{2/(2+\delta)}.$$

Then, Proposition 5 yields that both this upper-bound and the sum presented in (D1) are finite; and that Condition A5 is also satisfied. Therefore, we apply theorem 2.2 of Francq and Zakoïan (2013) and obtain the strong consistency and asymptotic normality of the estimators. ■

APPENDIX E. PROOF OF THE SLUTSKY-TYPE RESULT

We devote this section to the proof of Proposition 7. First, we prove that weakly approaching sequences of distributions are closed under the multiplication of same sequence on both sides under some technical assumption detailed below.

First, we prove the transitivity of this mode of convergence

Lemma 6. Suppose that three sequence of random variables on $\mathbb{R}^d$ are such that $\mathcal{L}(K_n) \xrightarrow{\text{wa}} \mathcal{L}(L_n)$ and $\mathcal{L}(L_n) \xrightarrow{\text{wa}} \mathcal{L}(M_n)$. Then, $\mathcal{L}(K_n) \xrightarrow{\text{wa}} \mathcal{L}(M_n)$.

Proof. Let $f(\cdot)$ be a bounded continuous function. Hence, one has

$$\mathbb{E}\{f(K_n) - f(M_n)\} = \mathbb{E}\{f(K_n) - f(L_n)\} + \mathbb{E}\{f(L_n) - f(M_n)\}.$$

By boundedness of f , those elements are all finite, and by definition of weakly approaching distributions, one obtains the result. ■

The final requirement is a Slutsky-type convergence in the sense of the following lemma. First, recall that one says that a sequence $(K_n, n \in \mathbb{N})$ is *tight* if and only if for any $\epsilon > 0$ there is a compact set $C_\epsilon \subset \mathbb{R}^d$ such that $P\{K_n \notin C_\epsilon\} < \epsilon$ for all $n \in \mathbb{N}$. We are interested in proving the following lemma

Lemma 7. Let $(M_n, n \in \mathbb{N}) \subseteq \mathbb{R}^{m \times d}$ for some $m \in \mathbb{N}$ be such that $M_n \xrightarrow{P} M$. Then, for any tight sequence $(K_n, n \in \mathbb{N}) \subseteq \mathbb{R}^d$, one has $\mathcal{L}(M_n K_n) \xrightarrow{\text{wa}} \mathcal{L}(M K_n)$ as $n \rightarrow \infty$.

Note that lemma 7 of Belyaev and Luna (2000) proved a similar result for the sum instead of the product. Its proof is instrumental in the following:

Proof. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a bounded continuous function and $\epsilon > 0$. Since (K_n) is tight, let C_ϵ be a compact set such that $P(K_n \notin C_\epsilon) < \epsilon/2$ for any $n \in \mathbb{N}$. Consider the Euclidean norm on $\mathbb{R}^d$ denoted $\|\cdot\|$. Then, $C_\epsilon \times \{z : \|z - M\| \leq \delta\}$ is compact (with respect to the product topology) hence one has the uniform continuity of $w \mapsto f(w)$ on this particular set. This means that, since $M_n \xrightarrow{P} M$, there exist $\delta := \delta(\epsilon)$ sufficiently small and $n_0 := n_0(\epsilon, \delta)$ sufficiently large such that, for $n \geq n_0(\epsilon, \delta)$, one has $P\{\|M_n - M\| > \delta\} < \epsilon/2$ and

$$\sup \{|f(yx) - f(Mx)| : (x, y) \in C_\epsilon \times \{z : \|z - M\| \leq \delta\}\} < \epsilon.$$

One finds that

$$P(|f(M_n K_n) - f(M K_n)| < \epsilon) \geq P(\{K_n \in C_\epsilon\} \cap \{\|M_n - M\| \leq \delta\}) \geq 1 - \epsilon,$$

independently from $\epsilon > 0$ therefore $f(M_n K_n) - f(M K_n) \xrightarrow{P} 0$ as $n \rightarrow \infty$. Recall that f is bounded, hence $|f(yx) - f(Mx)| \leq 2 \sup_{z \in \mathbb{R}^d} |f(z)| < \infty$ for $(x, y) \in C_\epsilon \times \{z : \|z - M\| \leq \delta\}$. Hence, the difference $(x, y) \mapsto |f(yx) - f(Mx)|$ is either bounded on a compact set, that is with finite Lebesgue measure, or potentially unbounded on a set with an arbitrarily small Lebesgue measure. By application of the dominated convergence theorem, one concludes with the result. ■

With those lemmas at hand, we can prove Proposition 7 as follows:

Proof. Given Lemma 7, we have that $\mathcal{L}(M_n L_n) \xleftrightarrow{\text{wa}} \mathcal{L}(M L_n)$ and $\mathcal{L}(M_n K_n) \xleftrightarrow{\text{wa}} \mathcal{L}(M K_n)$ as $n \rightarrow \infty$. Now, say $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is a bounded continuous function, then in particular $x \mapsto f(Mx)$ is also bounded and continuous as the matrix product is continuous hence, since $(K_n, n \in \mathbb{N})$ and $(L_n, n \in \mathbb{N})$ have weakly approaching distributions, we have $\mathcal{L}(M L_n) \xleftrightarrow{\text{wa}} \mathcal{L}(M K_n)$ as $n \rightarrow \infty$. Therefore, by transitivity (Lemma 6), one obtains that $\mathcal{L}(M_n L_n) \xleftrightarrow{\text{wa}} \mathcal{L}(M_n K_n)$ as $n \rightarrow \infty$. ■

APPENDIX F. FUNDAMENTAL CONVERGENCE RESULTS

In this section, we prove the asymptotic normality in the *weakly approaching* sense (Belyaev & Luna, 2000) as in Lemma 1. We commence this section with the almost sure convergence and tightness of $\nabla_{\theta} \zeta_i^g(\theta)$. Then, we connect the assumptions made in Lemma 1 and the set assumptions in Hwang (2021) as per Lemma 3. Finally, we prove the *pairwise strong mixing property* of $(\zeta_i^g(\theta), i \in \mathbb{N})$ from Condition (A1).

F.1 Proof of Proposition 8

Proof. By the strong mixing assumption, we have that $(\zeta_i^g(\theta_0), i \in \mathbb{N})$ is ergodic which, in turn, implies that the sequence converge almost surely to its limit $\Gamma(\theta_0)$. Hence, it also converges in probability. For simplicity, denote by Z_N this said sequence and Z its limit. Let $\epsilon > 0$ and let $\|\cdot\|$ be the Euclidean norm on $\mathbb{R}^d$. We find a compact set C_ϵ with respect to $\|\cdot\|$ such that $P\{Z_N \notin C_\epsilon\} < \epsilon$, for all $N \in \mathbb{N}$. Let $\delta > 0$, by convergence in probability, there exists a rank $N_0 = N_0(\epsilon)$ such that for all $N \geq N_0$, one has $P\{\|Z_N - Z\| \geq \delta\} \leq \epsilon/2$. Define $\bar{\delta} := \max\{\delta\} \cup \{\|Z_N - Z\| \times (N+1)/N : N \in \{0, \dots, N_0\}\}$. Then we have for any $N \geq N_0$ that

$$P\left(\|Z_N - Z\| \geq \bar{\delta}\right) \leq P(\|Z_N - Z\| \geq \delta) < \epsilon.$$

In addition, we obtain for any $N \in \{1, \dots, N_0\}$, if $\|Z_N - Z\| = 0$, we conclude, otherwise observe that

$$P\left(\|Z_N - Z\| \geq \bar{\delta}\right) \leq P(\|Z_N - Z\| \geq \|Z_N - Z\| \times (N+1)/N) = 0 < \epsilon.$$

Hence we take C_ϵ to be the closed ball with center Z and radius $\bar{\delta}$ with respect to $\|\cdot\|$. This implies that the sequence is tight. ■

F.2 Proof of Lemma 1

Proof. From Conditions (A1), we deduce that ζ^g is strongly mixing. We define the cluster-wise sample average vector as follows

$$\bar{\mathbf{Y}}_N^G(\theta_0) := \left(N^{-1} \sum_{i=1}^{N_1} \zeta_i^1(\theta_0)^\top, \dots, N^{-1} \sum_{i=1}^{N_G} \zeta_i^G(\theta_0)^\top \right)^\top \in \mathbb{R}^{G(3+p)},$$

where we recall that $N_g/N \rightarrow 1$ and $N_g/\bar{N} \rightarrow 1$ as $N \rightarrow \infty$ for any $g \in \{1, \dots, G\}$. If Conditions (A1) and (A2) hold, the central limit theorem for strongly mixing sequences proposed in Ekström (2014) states that

$$\mathcal{L} \left(N^{1/2} \bar{\mathbf{Y}}_N^G(\theta_0) \right) \stackrel{\text{wa}}{\leftrightarrow} \mathcal{N} \left(\mathbf{0}, \text{Var}(N^{1/2} \bar{\mathbf{Y}}_N^G(\theta_0)) \right), \quad \text{as } N \rightarrow \infty.$$

Note that $\text{Var}(N^{1/2} \bar{\mathbf{Y}}_N^G(\theta_0))$ is a $G(3+p)$ -square matrix. We show that the $(3+p) \times (3+p)$ off-diagonal blocks are asymptotically vanishing as $N \rightarrow \infty$ similarly to appendix B.1 of Hwang (2020). By Assumption 1, recall that $\mathbb{E}(\zeta_i^g(\theta_0)) = 0$ and we define the (variance) sub-matrices by

$$V_N^{g,h} := \mathbb{E} \left[\left(\sum_{i=1}^{N_g} \zeta_i^g(\theta_0) \right) \cdot \left(\sum_{j=1}^{N_h} \zeta_j^h(\theta_0) \right)^\top \right], \quad \text{for } g, h \in \{1, \dots, G\}.$$

Without loss of generality, suppose that $g < h$. By the triangle inequality, one has

$$N^{-1} |V_N^{g,h}| \leq N^{-1} \sum_{i=1}^{N_g} \sum_{j=1}^{N_h} \left| \mathbb{E} \left(\zeta_i^g(\theta_0) \zeta_j^h(\theta_0)^\top \right) \right| \leq N^{-1} \sum_{k=0}^{\infty} \sum_{\substack{1 \leq i \leq N_g, \\ i+k \leq N_h}} \left| \mathbb{E} \left(\zeta_i^g(\theta_0) \zeta_{i+k}^h(\theta_0)^\top \right) \right|.$$

For some $\delta > 0$, an established upper-bound on the covariance (Ibragimov et al., 1971, theorems 7.2.2 & 7.2.3) yields

$$\left| \mathbb{E} \left(\zeta_i^g(\theta_0) \zeta_{i+k}^h(\theta_0)^\top \right) \right| \leq C_\delta^{g,h}(\theta_0) \cdot \alpha^*(\zeta, k)^{\delta/(2+\delta)},$$

where $C_\delta^{g,h}(\theta_0) := c_\delta^{g,h}(\theta_0) \mathbb{E}(\|\zeta_i^g(\theta_0)\|^{2+\delta})^{(2+\delta)^{-1}} \mathbb{E}(\|\zeta_{i+k}^h(\theta_0)\|^{2+\delta})^{(2+\delta)^{-1}} < \infty$ by Condition (A2) for some $c_\delta^{g,h}(\theta_0) > 0$. Then, observe that for a given $k \in \mathbb{N}$ there exists a positive constant η such that $|\{i : 1 \leq i \leq N_g, i+k \leq N_h\}| = O(\bar{N}^{1-\eta})$ as $N \rightarrow \infty$. Hence, by Condition (A1), we obtain that

$$N^{-1} |V_N^{g,h}| \leq O(\bar{N}^{-\eta}) \sum_{k=0}^{\infty} k \alpha^*(\zeta, k)^{\delta/(2+\delta)} = o(1), \quad \text{as } N \rightarrow \infty.$$

Therefore, the off-diagonal variance submatrices vanish asymptotically as well as their sample equivalent by ergodicity. On the other hand, for diagonal submatrices, that is, for $g = h$, we obtain by ergodicity that $N^{-1} V_N^{g,h} \rightarrow \Omega_g(\theta)$. We deduce that, as $N \rightarrow \infty$, we have

$$\mathcal{L} \left(\left(N_1^{-1/2} \sum_{i=1}^{N_1} \zeta_i^1(\theta_0)^\top, \dots, N_G^{-1/2} \sum_{i=1}^{N_G} \zeta_i^G(\theta_0)^\top \right)^\top \right) \stackrel{\text{wa}}{\leftrightarrow} \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \Omega_1(\theta_0) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Omega_G(\theta_0) \end{bmatrix} \right),$$

similar to assumption 1 (ii) in Hwang (2021). In particular, we conclude that

$$\mathcal{L} \left(G^{1/2} \bar{N}^{1/2} \left[\bar{\zeta}_N(\theta_0) - \mathbb{E}\{\bar{\zeta}_N(\theta_0)\} \right] \right) \stackrel{\text{wa}}{\leftrightarrow} \mathcal{N}(\mathbf{0}, \Omega(\theta_0)).$$

with the cluster-wise decomposition $\Omega(\theta_0) = G^{-1} \sum_{g=1}^G \Omega_g(\theta_0)$. ■

F.3 Proof of Proposition 9

We prove a core result that links the asymptotic property of ζ to the ones of $\hat{\theta}$:

Proof. Given the continuous differentiability of $(\zeta_i^g(\theta), i \in \mathcal{G}_g)$ in θ for a fixed $g \in \{1, \dots, G\}$, a Taylor's expansion around the true value θ_0 yields that $G^{1/2} \bar{N}^{1/2} (\hat{\theta} - \theta_0)$ is asymptotically equivalent in the weakly approaching sequences sense to

$$- \left(G^{-1} \sum_{g=1}^G N_g^{-1} \sum_{i=1}^{N_g} \nabla_{\theta}(\zeta_i^g(\theta_0)) \right)^{-1} \cdot G^{1/2} \bar{N}^{1/2} \bar{\zeta}_N(\theta_0), \text{ as } N_g \rightarrow \infty, \quad (\text{F1})$$

where $\bar{\zeta}_N(\theta_0) = G^{-1} \sum_{g=1}^G N_g^{-1} \sum_{i=1}^{N_g} \zeta_i^g(\theta_0)$; under the assumption that $\Gamma(\theta)$ is invertible around θ_0 . From Lemma 1, one obtains that the right-hand side term converges in the weakly approaching sequences sense, that is, we have

$$\mathcal{L} \left(G^{1/2} \bar{N}^{1/2} \left[\bar{\zeta}_N(\theta_0) - \mathbb{E}\{\bar{\zeta}_N(\theta_0)\} \right] \right) \stackrel{\text{wa}}{\leftrightarrow} \mathcal{N}(\mathbf{0}, \Omega(\theta_0)).$$

The left-hand side sequence is tight (Proposition 8) and the continuous mapping theorem yields that it converges in probability to $-\Gamma(\theta)^{-1}$. Finally, the Slutsky-type result for weakly approaching sequences (Proposition 7) gives that

$$\mathcal{L} \left(G^{1/2} \bar{N}^{1/2} (\hat{\theta} - \theta_0) \right) \stackrel{\text{wa}}{\leftrightarrow} \mathcal{N}(\mathbf{0}, \Gamma(\theta_0)^{-1} \cdot \Omega(\theta_0) \cdot (\Gamma(\theta_0)^{-1})^\top), \text{ as } N \rightarrow \infty. \quad \blacksquare$$

F.4 Proof of Proposition 10

We hereby prove Proposition 10 which states that—for the latent trawl model—we have the strong mixing property for any pairwise measurable mapping ζ ; more precisely, the strong mixing coefficient is still dominated by the latent trawl correlation as in Proposition 3. Also, Condition (A1) holds for asymptotically strictly increasing trawl functions, that is, such that $\lim_{t \rightarrow \infty} \mu^{\text{leb}}(A \cap A_{t+1}) / \mu^{\text{leb}}(A \cap A_t) < 1$.

Proof of Proposition 10. Let $g, h \in \{1, \dots, G\}$. Suppose that $g = h$ to compute the strong mixing coefficient of $\zeta^g(\theta)$ seen as a measurable mapping of elements from $\mathcal{G}_g$. We retrieve similar results to the ones of Proposition 3 and use the notation from its proof presented in Appendix C,

that is,

$$f^{(2)}(x_i, x_{i+k}) = \int_{(0, \infty)^2} r(x_i, x_{i+k}, \lambda_i, \lambda_{i+k}; \kappa) dF(\lambda_i, \lambda_{i+k}),$$

and we extend it to the four-dimensional case.

Let $i \in \mathbb{Z}$, $k \in \mathbb{N}$ and take $\mathbb{S}_i^{g'} \in \mathcal{F}_{-\infty}^i(\zeta^g(\theta)) \subseteq \mathcal{B}(\mathbb{R}^2)$ and $\mathbb{S}_{i+k}^{g'} \in \mathcal{F}_{i+k}^{+\infty}(\zeta^g(\theta)) \subseteq \mathcal{B}(\mathbb{R}^2)$. To simplify the notation, we consider the preimages $\mathbb{S}_i^g := (X_i, X_{i+g})^{-1}(\mathbb{S}_i^{g'})$ and $\mathbb{S}_{i+k}^g := (X_{i+k}, X_{i+g+k})^{-1}(\mathbb{S}_{i+k}^{g'})$. (The arguments presented here can be extended to the general case of preimages with respect to any pair of sets in $\mathcal{F}_{-\infty}^i(\zeta^g(\theta))$ and $\mathcal{F}_{i+k}^{+\infty}(\zeta^g(\theta))$, and this proof generalizes the proof of Proposition 3 in Appendix C). By stationarity, we consider the set of indices $J_k^g := \{0, g, k, g+k\}$ and simplify the notations by omitting the index i . Then, using Fubini's theorem and the conditional independence and with loss of generality for $k > g$:

$$\begin{aligned} P(\mathbb{S}_0^{g'} \cap \mathbb{S}_k^{g'}) &= \int_{\mathbb{S}_0^g \times \mathbb{S}_k^g} f(x_0, x_g, x_k, x_{g+k}) dx_0 dx_g dx_k dx_{g+k} \\ &= \int_{\mathbb{S}_0^g \times \mathbb{S}_k^g} \int_{(0, \infty)^4} f(\lambda_0, \lambda_g, \lambda_k, \lambda_{g+k}) \prod_{j \in J_k^g} f(x_j | \lambda_j) \prod_{j \in J_k^g} d\lambda_j \prod_{j \in J_k^g} dx_j \\ &= \int_{(0, \infty)^4} \left[\int_{\mathbb{S}_0^g \times \mathbb{S}_k^g} \prod_{j \in J_k^g} f(x_j | \lambda_j) dx_j \right] f(\lambda_0, \lambda_g, \lambda_k, \lambda_{g+k}) \prod_{j \in J_k^g} d\lambda_j. \end{aligned}$$

On the other hand, one has

$$\begin{aligned} P(\mathbb{S}_0^{g'}) P(\mathbb{S}_k^{g'}) &= \int_{\mathbb{S}_0^g \times \mathbb{S}_k^g} f^{(2)}(x_0, x_g) f^{(2)}(x_k, x_{g+k}) dx_0 dx_g dx_k dx_{g+k} \\ &= \int_{\mathbb{S}_0^g \times \mathbb{S}_k^g} \int_{(0, \infty)^4} f(\lambda_0, \lambda_g) f(\lambda_k, \lambda_{g+k}) \prod_{j \in J_k^g} f(x_j | \lambda_j) \prod_{j \in J_k^g} d\lambda_j \prod_{j \in J_k^g} dx_j \\ &= \int_{(0, \infty)^4} \left[\int_{\mathbb{S}_0^g \times \mathbb{S}_k^g} \prod_{j \in J_k^g} f(x_j | \lambda_j) dx_j \right] f(\lambda_0, \lambda_g) f(\lambda_k, \lambda_{g+k}) \prod_{j \in J_k^g} d\lambda_j. \end{aligned}$$

Remark that, similarly to using the function defined in (C1) and the upper-bound given in (C2) adapted to the four indices used, the following upper-bound can be found

$$\int_{\mathbb{S}_0^g \times \mathbb{S}_k^g} \prod_{j \in J_k^g} f(x_j | \lambda_j) dx_j \leq \prod_{j \in J_k^g} e^{-\kappa \lambda_j}.$$

Therefore, by the triangle inequality

$$\begin{aligned} &|P(\mathbb{S}_0^{g'} \cap \mathbb{S}_k^{g'}) - P(\mathbb{S}_0^{g'}) P(\mathbb{S}_k^{g'})| \\ &\leq \left| \mathbb{E} \left(\prod_{j \in J_k^g} e^{-\kappa \Lambda_j} \right) - \mathbb{E} \left(\prod_{j \in \{0, g\}} e^{-\kappa \Lambda_j} \right) \mathbb{E} \left(\prod_{j \in \{k, g+k\}} e^{-\kappa \Lambda_j} \right) \right| \end{aligned}$$

$$\leq \left| \mathbb{E} \left(\prod_{j \in \mathcal{J}_k^g} e^{-\kappa \Lambda_j} \right) - \mathbb{E} \left(\prod_{j \in \{0, g\}} e^{-\kappa \Lambda_j} \right) \mathbb{E} \left(\prod_{j \in \{k, g+k\}} e^{-\kappa \Lambda_j} \right) \right|.$$

Given we consider a monotonic trawl function, the *slice* partition (Noven et al., 2018, p. 5) yields a partition $\{S(i, j) : i \leq j \in \mathcal{J}_k^g\}$ such that

$$\cup_{j \in \mathcal{J}_k^g} A_j = \cup_{i \leq j \in \mathcal{J}_k^g} S(i, j),$$

and $\{L(S(i, j)) : i \leq j \in \mathcal{J}_k^g\}$ are independent. Therefore, the first term is given by

$$\mathbb{E} \left(\prod_{j \in \mathcal{J}_k^g} e^{-\kappa \Lambda_j} \right) = \prod_{i \in \mathcal{J}_k^g} \prod_{j \in \mathcal{J}_k^g, j \geq i} \left(1 + \frac{\kappa |\{l : i \leq l \leq j, l \in \mathcal{J}_k^g\}|}{\beta} \right)^{-\alpha s_{ij}},$$

where $s_{ij} = \mu^{\text{leb}}(S(i, j)) / \mu^{\text{leb}}(A)$ and $|\cdot|$ is the cardinality of the set. Similarly, using the same partition, we write

$$\mathbb{E} \left(\prod_{j \in \{0, g\}} e^{-\kappa \Lambda_j} \right) = \prod_{i \in \{0, g\}} \prod_{j \in \mathcal{J}_k^g, j \geq i} \left(1 + \frac{\kappa (\mathbb{I}\{i = j\} + 2\mathbb{I}\{i = 0, j > 0\})}{\beta} \right)^{-\alpha s_{ij}},$$

and

$$\begin{aligned} & \mathbb{E} \left(\prod_{j \in \{k, k+g\}} e^{-\kappa \Lambda_j} \right) \\ &= \prod_{i \in \mathcal{J}_k^g} \prod_{j \in \mathcal{J}_k^g, j \geq i} \left(1 + \frac{\kappa (\mathbb{I}\{j = k\} \cup \{i = k + g, j = k + g\} + 2\mathbb{I}\{i < k + g, j = k + g\})}{\beta} \right)^{-\alpha s_{ij}}. \end{aligned}$$

Going back to the difference of probabilities, we have

$$\begin{aligned} & |P(\mathbb{S}_0^{g'} \cap \mathbb{S}_k^{g'}) - P(\mathbb{S}_0^{g'})P(\mathbb{S}_k^{g'})| \\ & \leq \left(1 + \frac{\kappa}{\beta} \right)^{-\alpha |\mathcal{J}_k^g|} \left(1 + \frac{\kappa}{\beta} \right)^{-\alpha \text{Cor}(\Lambda_0, \Lambda_g)} \left\{ 1 - \prod_{i \in \{0, g\}} \prod_{j \in \mathcal{J}_k^g, j \geq i} \left(1 + \frac{\kappa}{\beta} \right)^{-\alpha s_{ij}} \right\} \\ & \leq \left(1 + \frac{\kappa}{\beta} \right)^{-\alpha [|\mathcal{J}_k^g| + \text{Cor}(\Lambda_0, \Lambda_g)]} \left\{ 1 - \left(1 + \frac{\kappa}{\beta} \right)^{-\alpha \text{Cor}(\Lambda_g, \Lambda_k)} \right\}. \end{aligned}$$

Then, we obtain that

$$|P(\mathbb{S}_0^{g'} \cap \mathbb{S}_k^{g'}) - P(\mathbb{S}_0^{g'})P(\mathbb{S}_k^{g'})| = O(\text{Cor}(\Lambda_0, \Lambda_{k-g})), \quad \text{as } k \rightarrow \infty.$$

This is true for any such sets $\mathbb{S}_0^{g'}$ and $\mathbb{S}_k^{g'}$, we conclude

$$\alpha^*(\zeta^g(\theta), k) = O(\text{Cor}(\Lambda_0, \Lambda_k)), \quad \text{as } k \rightarrow \infty,$$

where the correlation can be computed using $\mu^{\text{leb}}(A \cap A_k)$ and $\mu^{\text{leb}}(A)$, similarly to Proposition 3. The strongly mixing coefficient of $\zeta(\theta_0)$ is computed in a similar fashion for the indices $J_k^{g,h} = \{0, g, k, k+h\}$ for $g, h \in \{1, \dots, G\}$. Finally, for strictly monotonic trawls, if $\lim_{t \rightarrow \infty} \mu^{\text{leb}}(A \cap A_{t+1}) / \mu^{\text{leb}}(A \cap A_t) < 1$, Condition (A1) holds by D'Alembert's criterion. ■

F.5 $(2 + \delta)$ -Integrability in the PL case (Lemma 2)

Proof of Lemma 2. We consider the case where the scale parameter β of the transformed marginals is set to 1. Similarly to Proposition 6, we only prove integrability conditions of derivatives with respect to ξ and ρ since σ and κ have no impact.

Suppose $\xi \geq 0$. From appendix A.1 of Noven et al. (2018) or Appendix B herein presented, without loss of generality, for $x_0 = 0$ and $x_g > 0$, we have $f(0, x_g) = J_h(x_g)f(0, h^{-1}(x_g))$ where

$$\begin{aligned} f_{\text{PL}}(0, y_g) &= \alpha(1 + \kappa + y_g)^{-\alpha-1} \\ &\quad - \alpha(1 + \kappa)^{b_{0 \setminus g}}(1 + \kappa + y_g)^{b_{0 \setminus g}}(1 + 2\kappa + y_g)^{(b_{0 \setminus g}-1)} \\ &\quad + b_{0 \setminus g}(1 + \kappa)^{b_{0 \setminus g}}(1 + \kappa + y_g)^{b_{0 \setminus g}-1}(1 + 2\kappa + y_g)^{b_{0 \setminus g}-1} \\ &= O\left((1 + \kappa + y_g)^{-\alpha-1}\right). \end{aligned}$$

Recall that

$$J_h(h(y)) = \frac{1 + \kappa}{\sigma} \left(1 + \frac{y}{1 + \kappa}\right)^{1-\xi}.$$

By the change of variable $y_g = h^{-1}(x_g)$, we have

$$\begin{aligned} \mathbb{E}[\|\nabla_{\theta} \ln f(0, X_g)\|^{2+\delta} \mathbb{I}_{(0, x_F)}(X_g)] \\ &= \int_0^\infty \|\nabla_{\theta} \ln f(0, x_g)\|^{2+\delta} f(0, x_g) dx_g \\ &= \int_0^\infty \|\nabla_{\theta} \ln J_h(x_g) + \nabla_{\theta} \ln f_{\text{PL}}(0, h^{-1}(x_g))\|^{2+\delta} J_h(x_g) f_{\text{PL}}(0, h^{-1}(x_g)) dx_g \\ &= \int_0^\infty \|\nabla_{\theta} \ln J_h(h(y_g)) + \nabla_{\theta} \ln f_{\text{PL}}(0, y_g)\|^{2+\delta} f_{\text{PL}}(0, y_g) dy_g. \end{aligned}$$

The derivative with respect to ξ yields

$$\int_0^\infty (\nabla_{\xi} \ln J_h(h(y_g)))^{2+\delta} f_{\text{PL}}(0, y_g) dy_g = \int_0^\infty (\ln(1 + y_g/(1 + \kappa)))^{2+\delta} f_{\text{PL}}(0, y_g) dy_g,$$

which finite if $\int_0^\infty (\ln(1 + y/(1 + \kappa)))^{2+\delta} (1 + \kappa + y_g)^{-\alpha-1} dy_g < \infty$. This is the case since $\alpha > 0$.

The derivative of f_{PL} w.r.t ρ is given by

$$\begin{aligned} \nabla_{\rho} f_{\text{PL}}(0, y_g) \\ &= \{(\nabla_{\rho} b_{0 \setminus g})[\ln(1 + \kappa + y_g) + \ln(1 + \kappa) + \kappa] + (\nabla_{\rho} b_{0 \setminus g}) \ln(1 + 2\kappa + y_g) - \alpha(1 + \kappa + y_g) \\ &\quad + b_{0 \setminus g} \kappa\} \times (1 + \kappa + y_g)^{b_{0 \setminus g}-1} (1 + 2\kappa + y_g)^{b_{0 \setminus g}-1} (1 + \kappa)^{b_{0 \setminus g}-1}, \end{aligned}$$

and, additionally, the likelihood is lower-bounded as follows

$$f_{\text{PL}}(0, y_g) \geq \alpha(1 + \kappa + y_g)^{-\alpha-1} [1 - (1 + \kappa)^{b_{0\setminus g}}] \\ + b_{0\setminus g} \kappa (1 + \kappa + y_g)^{b_{0\setminus g}-1} (1 + 2\kappa + y_g)^{b_{0\setminus g}-1} (1 + \kappa)^{b_{0\setminus g}-1}.$$

Hence, we obtain as $y_g \rightarrow \infty$

$$|\nabla_\rho \ln f_{\text{PL}}(0, y_g)| = |\nabla_\rho f_{\text{PL}}(0, y_g) / f_{\text{PL}}(0, y_g)| \leq C(1 + \kappa + y_g)^{-\alpha-1} (1 + 2\kappa + y_g)^{\alpha+1} = O(1),$$

for some constant $C > 0$. Hence, the derivative to the $(2 + \delta)$ th power is constant if $\int_0^\infty f_{\text{PL}}(0, y_g) dy_g < \infty$ which is true since $\int_0^\infty f_{\text{PL}}(0, y_g) dy_g < 1$. On the other hand, if $x_s, x_t > 0$, one has:

$$f(x_s, x_t) = J_h(x_s) J_h(x_t) f_{\text{PL}}(h^{-1}(x_s), h^{-1}(x_t)).$$

By the change of variables $y_0 = h^{-1}(x_0)$ and $y_g = h^{-1}(x_g)$, we have

$$\mathbb{E}[\|\nabla_\theta \ln f(X_0, X_g)\|^{2+\delta} \mathbb{I}_{(0, x_F)}(X_0) \mathbb{I}_{(0, x_F)}(X_g)] \\ = \int_0^\infty \int_0^\infty \|\nabla_\theta \ln J_h(h(y_0)) + \nabla_\theta \ln J_h(h(y_g)) + \nabla_\theta \ln f_{\text{PL}}(y_0, y_g)\|^{2+\delta} f_{\text{PL}}(y_0, y_g) dy_0 dy_g.$$

Again, by the change of variables $z_0 = y_0$ and $z_g = y_0 + y_g$, we reformulate the integral as follows

$$\int_0^\infty \int_0^{z_g} \|\nabla_\theta \ln J_h(h(z_0)) + \nabla_\theta \ln J_h(h(z_g - z_0)) + \nabla_\theta \ln f_{\text{PL}}(z_0, z_g)\|^{2+\delta} f_{\text{PL}}(z_0, z_g - z_0) dz_0 dz_g.$$

For the derivative with respect to ξ , we obtain the following upper-bound:

$$\int_0^\infty \int_0^{z_g} \left[\ln \left(1 + \frac{z_0}{1 + \kappa} \right) + \ln \left(1 + \frac{z_g - z_0}{1 + \kappa} \right) \right]^{2+\delta} f_{\text{PL}}(z_0, z_g - z_0) dz_0 dz_g.$$

From appendix A.1 of Noven et al. (2018) (see Appendix B), we bound the PL as follows

$$f_{\text{PL}}(z_0, z_g - z_0) \leq C \left(\sum_{i=1}^d f_{\text{PL}}^{(i)}(z_0, z_g - z_0) \right),$$

for some constant $C > 0$ where

$$f_{\text{PL}}^{(1)}(z_0, z_g - z_0) = (1 + 2\kappa + z_g)^{b_{0,h}-2} (1 + \kappa + z_0)^{b_{0\setminus h}} (1 + \kappa + z_g - z_0)^{b_{0\setminus h}}, \\ f_{\text{PL}}^{(2)}(z_0, z_g - z_0) = (1 + 2\kappa + z_g)^{b_{0,h}} (1 + \kappa + z_0)^{b_{0\setminus h}-1} (1 + \kappa + z_g - z_0)^{b_{0\setminus h}-1}, \\ f_{\text{PL}}^{(3)}(z_0, z_g - z_0) = (1 + 2\kappa + z_g)^{b_{0,h}-1} (1 + \kappa + z_0)^{b_{0\setminus h}-1} (1 + \kappa + z_g - z_0)^{b_{0\setminus h}}, \\ f_{\text{PL}}^{(4)}(z_0, z_g - z_0) = (1 + 2\kappa + z_g)^{b_{0,h}-1} (1 + \kappa + z_0)^{b_{0\setminus h}} (1 + \kappa + z_g - z_0)^{b_{0\setminus h}-1}.$$

Taking the upper-bound of the inner integral with $f_{\text{PL}}^{(1)}$ yields

$$z_g \cdot 2^{2+\delta} \ln \left(1 + \frac{z_g}{2(1+\kappa)} \right)^{2+\delta} (1+2\kappa+z_g)^{b_{0,h}-2} (1+\kappa+z_g/2)^{2b_{0,h}},$$

such that the outer integral is bounded by

$$2^{2+\delta} \int_0^\infty \ln \left(1 + \frac{z_g}{2(1+\kappa)} \right)^{2+\delta} (1+\kappa+z_g/2)^{1-\alpha+b_{0,h}} dz_g,$$

which is finite if $1-\alpha-2+b_{0,h} < -1 \Leftrightarrow \alpha(1+\mu^{\text{leb}}(A \setminus A_g)/\mu^{\text{leb}}(A)) > 1$. This condition holds since $\alpha \geq 1$ and $\mu^{\text{leb}}(A_0 \setminus A_g) > 0$ for $g > 0$ independently of δ . We can apply a similar proof for $f_{\text{PL}}^{(i)}$ for $i = 2, 3, 4$ and reach the same conclusion: we obtain the $(2+\delta)$ -integrability for any $\xi > 0$. For $\xi < 0$, we proceed similarly with the change of variables $y = h^{-1}(x)$. ■

F.6 Proof of Lemma 3

In this section, we assume that Assumptions 1 and 3 and Conditions (A1) and (A2) hold.

Lemma 8. *Assumption (P2) is satisfied.*

Proof. By Proposition 9, we have that

$$G^{1/2}(N-G)^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Gamma(\theta_0)^{-1} \cdot \Omega(\theta_0) \cdot (\Gamma(\theta_0)^{-1})^\top), \quad \text{as } N \rightarrow \infty.$$

Hence, for some $\epsilon > 0$, we write that

$$\mathbb{P}(\|\hat{\theta} - \theta_0\| > \epsilon) = \mathbb{P}(G^{-1/2}(N-G)^{-1/2}\|\mathbf{Z}\| > \epsilon/2) + o_p(1),$$

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \Gamma(\theta_0)^{-1} \cdot \Omega(\theta_0) \cdot (\Gamma(\theta_0)^{-1})^\top)$. Since $\Gamma(\theta_0)^{-1} \cdot \Omega(\theta_0) \cdot (\Gamma(\theta_0)^{-1})^\top < \infty$, the asymptotic variance of $\hat{\theta} - \theta_0$ vanishes as $N \rightarrow \infty$. We have the convergence in distribution of the estimator to the constant θ_0 , hence we obtain the (stronger) convergence in probability, that is, $\hat{\theta} \xrightarrow{\text{P}} \theta_0$. ■

Lemma 9. *Assumption (P3) is satisfied except the full rank assumption of $\Gamma(\theta_0)$.*

Proof. By Proposition 8, the sequence $\sum_{i=1}^{N_g} \nabla_{\theta} \zeta_i^g(\theta)$ converges almost surely to $\Gamma(\theta)$ as $N_g \rightarrow \infty$ for any θ in a neighborhood of θ_0 . Hence, it converges in probability with respect to the Euclidean norm and the first subassumption is satisfied. By assumption $\theta \mapsto \nabla_{\theta} \zeta_i^g(\theta)$ is continuous around θ_0 and implying that $\theta \mapsto \Gamma(\theta)$ is continuous too. Indeed, let θ be in a neighborhood of θ_0 with respect to the Euclidean norm $\|\cdot\|$ and let $\epsilon > 0$. There exists $\delta = \delta(\epsilon) > 0$ such that if $\|\theta - \theta'\| \leq \delta$ then $\|\nabla_{\theta} \zeta_i^g(\theta) - \nabla_{\theta} \zeta_i^g(\theta')\| \leq \epsilon^2/4$. Hence,

$$\begin{aligned} \|\Gamma(\theta) - \Gamma(\theta')\| &\leq G^{-1} \sum_{g=1}^G \|\Gamma_g(\theta) - \Gamma_g(\theta')\| \\ &\leq G^{-1} \sum_{g=1}^G \left\| \mathbb{E}(\nabla_{\theta} \zeta_i^g(\theta)) - \mathbb{E}(\nabla_{\theta} \zeta_i^g(\theta')) \right\| \end{aligned}$$

$$\leq G^{-1} \sum_{g=1}^G \mathbb{E} \left(\left\| \nabla_{\theta} \zeta_i^g(\theta) - \nabla_{\theta} \zeta_i^g(\theta') \right\| \right)^{1/2} \leq G^{-1} \sum_{g=1}^G \epsilon/2 < \epsilon.$$

This proves the continuity of $\theta \mapsto \Gamma(\theta)$. ■

Lemma 10. *Assumption (P4) is satisfied.*

Proof. Consider the result of Lemma 1 which could be generalized to $N_G^{-1/2} \sum_{i=1}^{N_g} \zeta_i^g(\theta_0)$ for any $g \in \{1, \dots, G\}$ and resulting in

$$\mathcal{L} \left(G^{-1/2} \sum_{i=1}^{N_g} \zeta_i^g(\theta_0) \right) \xrightarrow{\text{wa}} \mathcal{N}(\mathbf{0}, \Omega_g(\theta_0)).$$

Then, given Condition (A2), the second moments are finite which means that one can interpret the weakly approaching convergence as a convergence in distribution similarly to Proposition 9. Finally, given Assumption 3, we have independence of those limiting normal distributions which gives the desired result. ■

F.7 Proof of Lemma 5

We hereby prove that Condition (A2) holds for any $\xi_0 < 1/4$ as presented in Lemma 5.

Proof of Lemma 5. Let $\delta \in (0, 1)$. Suppose that the Gamma scale parameter β is set to 1. Similarly to the proof of Proposition 4, we obtain the $(2 + \delta)$ -integrability of the first three components if $\xi_0 < (1 + \delta)^{-1}$ for any $\delta > 0$. The last condition is given by

$$\mathbb{E} \left(|X_0 X_g - c_g|^{2+\delta} \right) = \int_0^\infty \int_0^\infty (x_0 x_g - c_h)^{2+\delta} J_h(x_0) J_h(x_g) f_{\text{PL}}(h^{-1}(x_0), h^{-1}(x_g)) dx_0 dx_g < \infty,$$

where $c_g := \mathbb{E}(X_0 X_g)$. Using $x = h(y)$ for each variable, we obtain

$$\mathbb{E} \left(|X_0 X_g - c_g|^{2+\delta} \right) = \int_0^\infty \int_0^\infty (h(y_0) h(y_g) - c_h)^{2+\delta} f_{\text{PL}}(y_0, y_g) dy_0 dy_g,$$

where we recall that $h(y) = \frac{\sigma}{\xi} \left[\left(1 + \frac{y}{1+\kappa} \right)^\xi - 1 \right]$. Then, by change of variables $z_0 = y_0$ and $z_g = y_0 + y_g$, we reformulate as follows

$$\int_0^\infty \int_0^{z_g} (h(z_0) h(z_g - z_0) - c_h)^{2+\delta} f_{\text{PL}}(z_0, z_g - z_0) dz_0 dz_g.$$

By taking upper bounds, we prove the integrability of

$$\int_0^\infty \int_0^{z_g} \left(1 + \frac{z_0}{1+\kappa} \right)^{\xi(2+\delta)} \left(1 + \frac{z_g - z_0}{1+\kappa} \right)^{\xi(2+\delta)} f_{\text{PL}}(z_0, z_g - z_0) dz_0 dz_g.$$

Taking the notation from the proof of Lemma 2 as presented in Appendix F.0.5. Recall that $b_k < 0$ for $k \in \{\{0 \setminus g\}, \{0, g\}\}$, hence we prove the finiteness of

$$\int_0^\infty (1+z_g)^{b_{0,g}-2} \int_0^{z_g} (1+z_0)^{\xi(2+\delta)+b_{0,g}-1} (1+z_g-z_0)^{\xi(2+\delta)+b_{0,g}} dz_0 dz_g.$$

For $z_g > 0$, the inner integral is upper-bounded by

$$\begin{aligned} & \int_0^{z_g} (1+z_0)^{\xi(2+\delta)+b_{0,g}-1} (1+z_g-z_0)^{\xi(2+\delta)+b_{0,g}-1} dz_0 \\ &= z_g (1+z_g/2)^{2(\xi(2+\delta)+b_{0,g})} \leq (1+z_g/2)^{2(\xi(2+\delta)+b_{0,g})+1}. \end{aligned}$$

Therefore, for the double integral to be finite, we should have

$$(1+z_g)^{b_{0,g}-2} (1+z_g/2)^{2(\xi(2+\delta)+b_{0,g}-1)+1} = o\left((1+z_g)^{-1}\right), \quad \text{as } z_g \rightarrow \infty,$$

that is, $b_{0,g} - 2 + 2(\xi(2+\delta) + b_{0,g}) + 1 < -1$ such that

$$\xi < \frac{\alpha - b_{0,g}}{2(2+\delta)} = \frac{\alpha}{2(2+\delta)} \left(1 + \frac{\mu^{leb}(A \setminus A_g)}{\mu^{leb}(A)} \right).$$

For this to hold for any $g > 0$, we obtain that

$$\xi_0 < \frac{\alpha}{2(2+\delta)}, \quad \text{for any } \delta \in (0, 1).$$

Therefore, since $\alpha \geq 1$, having $\xi_0 < 1/4$ is a sufficient condition for the moment condition to hold. ■