PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

**Cite this article:** Truquet L, Cohen JE, Doukhan P. 2025 Inferring the parameters of Taylor's power law of fluctuation scaling. *Proc. R. Soc. A* **481**: 20250248. https://doi.org/10.1098/rspa.2025.0248

Received: 22 March 2025
Accepted: 6 October 2025

**Subject Areas:**
statistics

**Keywords:**
asymptotic statistics for space-time data, biased confidence interval, non-stationary time series

**Author for correspondence:**
L. Truquet
e-mail: Lionel.TRUQUET@ensai.fr

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.8115997.

THE ROYAL SOCIETY PUBLISHING



# Inferring the parameters of Taylor's power law of fluctuation scaling

L. Truquet[1], J. E. Cohen[2] and P. Doukhan[3]

[1]Ensai, CNRS, CREST—UMR 9194, University of Rennes, F-35000 Rennes, France
[2]Laboratory of Populations, The Rockefeller University, New York, NY 10065, USA
[3]AGM Mathematics, CY Cergy Paris University, 2 Bd. Adolphe Chauvin, AGM-UMR 8080, 95000, Cergy-Pontoise, France

LT, 0000-0002-0017-1438; JEC, 0000-0002-9746-6725

Taylor's power law of fluctuation scaling has been well approximated empirically in many fields, including physics, meteorology, computer science, finance and ecology. In ecology, it describes well the abundances of many species, humans and non-humans. Taylor's law asserts that, in a collection of probability distributions, the variance of each distribution is directly proportional to a power of the mean of each distribution, exactly for population moments and, whether or not population moments exist, approximately for sample moments. Linear regression of log variance as a function of log mean is widely used to estimate the parameters of Taylor's power law. We provide large-sample asymptotics for this kind of inference under general conditions, and we derive confidence intervals for the parameters. In many ecological applications, the means and variances are estimated over time or across space from arrays of abundance data collected at different locations and time points. Our results depend on the asymptotic relation between the time-series length and the number of spatial points. When their ratio converges to a constant as both become large, the usual normalized statistics are asymptotically biased. We provide a bias correction to get correct confidence intervals. Taylor's law, widely studied in multiple sciences, is a source of challenging new statistical problems in a non-stationary spatio-temporal framework. We illustrate our results with both simulated and real data.

# 1. Introduction

Taylor's law is an important family of relations between the mean and the variance of some probability distributions. Taylor's law is well approximated empirically in many datasets in the natural and social sciences, finance and technology. Since [1] publicized this relationship, which was discovered by others at least 20 years earlier, such as [2], thousands of papers have studied Taylor's law for species abundance data in ecology and more generally to assess time-dependent or space-dependent changes of distributions of population or other non-negative variables. Taylor [3] and the references therein give a broad introduction to Taylor's law.

The aim of this paper is to study Taylor's power law, which is a simple power relation between the variance and the mean of distributions in a family of marginal distributions. Informally, Taylor's law says that the variance of each distribution in the family is proportional to some (usually positive) power of the mean of that distribution: variance of distribution = $\alpha \times (\text{ mean of distribution})^{\beta}$, where $\alpha > 0$ and $\beta$ are constant for all distributions in the family. We define Taylor's law more formally in equation (2.1). Because Taylor's law has an interpretation in terms of aggregation of 'individuals' across time or space, it became widely popular in ecology, even though many other interpretations are equally plausible. Cohen & Saitoh [4] discuss the interpretation of the exponent $\beta$ (also called the slope parameter) of Taylor's power law. In scientific use and testing of Taylor's law, equation (2.1) is the most basic relation against many plausible alternative models.

Many extensions of Taylor's law can be more appropriate than equation (2.1). For example, the quadratic Taylor's law [5], where log variance is a quadratic polynomial function of log mean, describes U.S. county populations more accurately [6]. A cubic generalization of Taylor's law, where log variance is a cubic polynomial function of log mean, is required to describe mortality rates in Japan [7]. And sometimes, a variance function expressed as a linear combination of the mean and the square of the mean [8] (a generalization of one of the negative binomial distribution's variance functions) is superior to any form of Taylor's law [9]. Fitting Taylor's law (equation (2.1)) is a step that follows exploratory data analysis of multiple alternative models, each of which poses its own problems of estimating parameters.

Whatever the form of Taylor's law, estimation of unknown parameters roughly follows the same idea. We compute sample means and sample variances along one dimension (say, space) of a rectangular array of non-negative data indexed by (say) space and time. This produces a pair (sample mean, sample variance) for each point in time. The sample mean and sample variance are assumed to be positive so their logarithms are well defined. The slope and intercept of log variance as a linear function of log mean are calculated, typically by ordinary least squares but sometimes by more robust methods to reduce vulnerability to outliers, e.g. Hampel's function and Tukey's biweight function [7] to optimize the fit of Taylor's power law to these pairs of sample moments.

For example, in ecological species abundance data, the spatial Taylor's law, on which we focus in this paper, assumes that the time-varying mean and variance of the marginal distribution are linked at each time $t$ by equation (2.1) and these population quantities are approximated by the corresponding sample moments computed from abundance data of a given species observed at different locations for each time $t$. After taking the logarithm of equation (2.1), this leads to the least-squares estimator in §2.

A simple log–log plot (log variance on the vertical axis, log mean on the horizontal axis) is the classical graphical tool used by practitioners to assess whether Taylor's law describes the data. We show in §4 that the confidence intervals of parameter estimates resulting from ordinary least squares may be seriously biased, and we show how to correct that bias. This is illustrated in §5, with real data.

A wide statistical literature links variance laws like Taylor's law with some classes of distributions [10–12]; those classes are compared in [13]. Cohen & Huillet [14] provide infinitely divisible examples of the Bar-Lev Enis type. However, formal statistical analyses of Taylor's law are rather scarce. Some methodological contributions include [15–20], among others. Beyond

the previous papers which state that Taylor's law holds for interesting classes of distributions, these contributions mainly focus on inference procedures or computational issues and compare different approaches for estimating the parameters of Taylor's law in various situations such as sparsity, small samples or data with covariates. None of these contributions discuss formally the large-sample behaviour of least-squares or likelihood estimators, which is precisely the scope of the present paper.

We believe that the statistical analysis here can guide the analysis of more complex versions of Taylor's law. In this paper, we study some detailed conditions that justify the commonly used method of log variance-log mean least-squares regression. However, many extensions of our results are possible. Our approach to studying Taylor's law requires a careful analysis of a specific minimum contrast estimator derived from some sample moments instead of the usual pair $(x, y) = (\text{covariate, response})$ used in standard regression analysis. Analyzing other methods such as likelihood estimation or more general models that use log-regression could be conducted using similar arguments.

This work can be useful for both quantitative ecologists and statisticians. Our analysis can help ecologists to measure the uncertainty in the estimated parameters with precise confidence intervals. Our analysis also shows that the asymptotic distribution of the regression estimates depends on the asymptotic relation between the time-series length and the number of spatial points. By using our results, practitioners can avoid using standard but biased methods to compute confidence intervals for the parameters of Taylor's law. By biased, we mean a non-centred asymptotic distribution for the parameters of Taylor's law. In general, in linear regression models, the difference between the ordinary least-squares estimator and the target parameter has a probability distribution that can be well approximated by a zero-mean Gaussian distribution. This will not be necessarily the case for Taylor's law. In §4, we show that in some cases, the Gaussian asymptotic distribution of least-squares estimates is not centred. For statisticians, our work reveals that a popular family of stochastic relations widely studied in several sciences is a source of interesting new theoretical problems, in particular for getting consistency properties or for constructing testing problems in a non-stationary spatio-temporal framework.

Many extensions of our results are possible, for example to the numerous variants of Taylor's law discussed above. In addition, this paper focuses only on distributions with light tails, i.e. when the probability distributions of interest have some finite moments of order $k > 2$. Though sufficient to prove rigorous consistency results for our estimates, such a restriction is not necessary for Taylor's law to occur. Taylor's law may hold for the sample mean and sample variance from stable laws with infinite variance or infinite mean [21–24]. However, even with our restriction to light-tailed distributions, ours is the first rigorous analysis of the asymptotic properties of the least-squares estimators of Taylor's power law parameters.

The paper is organized as follows. In §2, we specify our model and the least-squares estimators of the parameters of Taylor's law. In §3, we provide consistency properties. In §4, we prove the asymptotic normality of our log-regression ordinary least-squares estimators. In §5, we apply our results to real data. In §6, we summarize our conclusions and, in §7, we provide proofs of our results. Appendix A contains some technical lemmas needed in §7. Electronic supplementary material available online contains an additional proof as well as some numerical experiments that assess the quality of the estimators studied in the paper, depending on various assumptions about the process that generates the data. The dataset used in §5 and the R or Matlab codes used for this paper are available at [25].

## 2. Least-squares estimators for the parameters of Taylor's law

Let $(X_{j,t})_{1\le j\le n, 1\le t\le T}$ be an array of non-negative random variables indexed by integers $j, t$. These variables can take integer values, for instance, when $X_{j,t}$ represents the number of individuals observed at time $t$ and at location $j$, or more generally non-negative real values, for instance,

when $X_{j,t}$ represents the number of individuals observed at time $t$ divided by the area of site $j$, in which case $X_{j,t}$ represents a number of individuals per unit of area (population density).

We assume that the spatial variance $\sigma_t^2 := \mathrm{Var}(X_{j,t})$ (that is, the variance of the $n$ elements in column $t$ of the array $(X_{j,t})$) and the spatial mean $\mu_t := \mathbb{E}(X_{j,t})$ (that is, the mean of the $n$ elements in column $t$ of the array $(X_{j,t})$) at time $t$ do not depend on the location $j$, are strictly positive, and satisfy the spatial Taylor's law (to be explained further below),

$$\sigma_t^2 = \alpha \mu_t^\beta, \quad 1 \le t \le T, \tag{2.1}$$

for a pair of parameters $(\alpha, \beta) \in (0, \infty) \times \mathbb{R}$. In most ecological applications, the estimate of $\beta$ is positive. A value $\beta > 1$ is sometimes interpreted as indicating aggregation or clumping of the organisms because a family of Poisson distributions in which the mean varies in time and remains constant over space obeys $\alpha = \beta = 1$. Values below 1 are uncommon but possible, for example, in models of a population in a Markovian environment, e.g. [26,27]. Numerical simulations of some fisheries models gave $\beta < 0$ [28].

The population parameters $(\mu_t, \sigma_t^2)$ are estimated by their empirical counterparts from the sample $X_{1,t}, \ldots, X_{n,t}$. See below for details. This form of Taylor's law is known as a spatial Taylor's law because the empirical mean and the empirical variance depend on $n$ spatial locations at each time. We do not consider here other variants of Taylor's law, such as the temporal Taylor's law, in which the mean and the variance depend on $T$ time points at each spatial location, or the hierarchical spatial Taylor's law or the hierarchical temporal Taylor's law. For instance, the temporal Taylor's law assumes that the population moments $\sigma_j^2 = \mathrm{Var}(X_{j,t})$ and $\mu_j = \mathbb{E}X_{j,t}$ depend only on $j$ and not on $t$. The corresponding power law is

$$\sigma_j^2 = \alpha \mu_j^\beta, \quad 1 \le j \le n.$$

Asymptotic properties of the temporal Taylor's law can be obtained with similar arguments. This is why we mainly focus on the spatial Taylor's law, equation (2.1).

The main problem here is inference about the two parameters $\alpha$ and $\beta$ of equation (2.1). Since equation (2.1) implies

$$\log \sigma_t^2 = \log(\alpha) + \beta \log(\mu_t), \tag{2.2}$$

a natural way to proceed is to use a linear regression of the logarithm of the empirical spatial variance computed at each time $t$ as a function of the logarithm of the empirical spatial mean computed at each time $t$. The empirical spatial mean and empirical spatial variance are, respectively,

$$\hat{\mu}_t := \frac{1}{n} \sum_{j=1}^{n} X_{j,t} \quad \text{and} \quad \hat{\sigma}_t^2 := \frac{1}{n} \sum_{j=1}^{n} X_{j,t}^2 - \hat{\mu}_t^2.$$

While the variance can be also estimated from the unbiased estimator

$$\hat{\sigma}_t^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_{j,t} - \hat{\mu}_t)^2,$$

our asymptotic results do depend on the choice of a biased or unbiased version of the sample variance. It is natural to estimate the pair $\theta = (\log(\alpha), \beta)'$ from the ordinary least-squares estimator

$$\hat{\theta}_n = \underset{\theta \in \mathbb{R}^2}{\arg\min} \sum_{t=1}^{T} (\log \hat{\sigma}_t^2 - \theta_1 - \theta_2 \log \hat{\mu}_t)^2.$$

Explicitly,

$$\hat{\theta}_n = \hat{D}_T^{-1} \hat{N}_T, \tag{2.3}$$

where

$$\hat{D}_T = \begin{pmatrix} 1 & \frac{1}{T}\sum_{t=1}^{T} \log \hat{\mu}_t \\ \frac{1}{T}\sum_{t=1}^{T} \log \hat{\mu}_t & \frac{1}{T}\sum_{t=1}^{T} \log^2 \hat{\mu}_t \end{pmatrix} \quad \text{and} \quad \hat{N}_T = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} \log \hat{\sigma}_t^2 \\ \frac{1}{T}\sum_{t=1}^{T} \log \hat{\mu}_t \cdot \log \hat{\sigma}_t^2 \end{pmatrix}. \tag{2.4}$$

In the above formula, $\log \hat{\mu}_t$ or $\log \hat{\sigma}_t$ are not defined when $\hat{\mu}_t$ or $\hat{\sigma}_t^2$ vanishes. In this case, one can simply replace these quantities by 0 in the expression of $\hat{\theta}_n$. Such a replacement is a simple convention and does not make any difference asymptotically, since $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ become positive for $n$ large enough. If we set

$$D_T := \begin{pmatrix} 1 & \frac{1}{T}\sum_{t=1}^T \log \mu_t \\ \frac{1}{T}\sum_{t=1}^T \log \mu_t & \frac{1}{T}\sum_{t=1}^T \log^2 \mu_t \end{pmatrix} \quad \text{and} \quad N_T := \begin{pmatrix} \frac{1}{T}\sum_{t=1}^T \log \sigma_t^2 \\ \frac{1}{T}\sum_{t=1}^T \log \mu_t \cdot \log \sigma_t^2 \end{pmatrix},$$

then analogously $\theta = D_T^{-1} N_T$, which makes it natural to introduce the above estimator. Throughout this paper, we assume that the time-series length $T$ depends on the number $n$ of spatial sites, i.e. $T := T_n$. We study the asymptotic properties of $\hat{\theta} := \hat{\theta}_n$ in equation (2.3) as $n \to \infty$. The asymptotic relation between $T$ and $n$ will be important but we will not need other specific assumptions on the mapping $n \mapsto T_n$.

## 3. Consistency properties

We now study the consistency of least-squares estimates. Consistency is defined to mean that $\lim_{n\to\infty} \hat{\theta} = \theta$ in probability (weak consistency) or almost surely (strong consistency).

If $T_+ := \sup_{n\geq 1} T_n < \infty$, consistency properties will roughly follow from the law of large numbers. It is more challenging to study the case $T_n \to \infty$, which is also useful in applications because the dimensions of the arrays of data are often balanced. Though time series with length diverging to infinity are the main focus, our results are also valid for finite $T_+$.

### (a) General results for non-negative random variables

To derive the asymptotic properties of our estimates, we make the following assumptions.

Assumption A1: The $n$ time series of length $T_n$ given by $(X_{j,t})_{1\leq t\leq T}$, $1 \leq j \leq n$, are independent and identically distributed. Assumption A2: There exists $k > 2$ such that $\sup_{t\geq 1} \mathbb{E}|X_{1,t}|^k < \infty$. Assumption A3: There exists some $\delta > 0$ such that

$$p(\delta) := \inf_{t\geq 1} \mathbb{P}(X_{1,t} > \delta) > 0 \quad \text{and} \quad q(\delta) := \inf_{t\geq 1} \mathbb{P}(|X_{1,t} - X_{2,t}| > \delta) > 0.$$

Assumption A4: There exists $c > 0$ such that for any positive integer $T$,

$$\frac{1}{T}\sum_{t=1}^T (\log \mu_t)^2 - \left(\frac{1}{T}\sum_{t=1}^T \log \mu_t\right)^2 \geq c.$$

Let us comment on our assumptions.

**Remark 3.1.** Assumption A1 does not require any specific dependence property across time. Independence across sites $j$ is needed only for technical reasons. However, spatial correlations can be observed in data, as for instance in a study of Hokkaido voles [4]. In theory, it may be possible to remove the assumption of independence across sites $j$ by assuming a stationarity condition and some weak dependence property between the different sites $j$. However, getting consistency properties under such conditions would require additional assumptions, in particular to get the law of large numbers when the locations $j$ are irregularly spaced. We prefer to keep the set-up simple.

**Remark 3.2.** The technical moment assumption A2 is needed to prove limit results as in theorem 3.5. In the electronic supplementary material we provide some numerical experiments which show that the moment assumption on $X_{1,t}$ influences the accuracy of our estimators.

**Remark 3.3.** Assumption A3 roughly means that Dirac masses cannot be cluster points of the marginal distribution, since assumption A3 is not satisfied if and only if at least one of the two sequences $(X_{1,t})_{t\geq 1}$ or $(X_{1,t} - X_{2,t})_{t\geq 1}$ has a subsequence converging to 0 in probability. Excluding

Dirac masses is important to avoid small values for the population mean and the population variance as well as for their empirical versions and then large negative values for their logarithm.

**Remark 3.4.** Assumption A4 ensures that the vectors $(\log \mu_t)_{1\le t\le T}$ vary sufficiently in time. Such a condition is needed to avoid some pathological behaviors such as the convergence of $\mu_t$ when $t \to \infty$. Without this condition, the matrix $\hat{D}_T$ can be asymptotically singular. For a standard linear regression model $y_t = x_t'\beta + \varepsilon_t$, where $x_t'$ denotes the transpose of the column vector $x_t$, our condition is comparable to the existence of a lower bound for the determinant of the matrices $T^{-1}\sum_{t=1}^{T} x_t x_t'$ classical in statistics.

**Theorem 3.5.** *Given assumptions A1–A2–A3–A4, there exists $\gamma > 0$ such that if $T = O(\exp(\gamma n))$, then $\lim_{n\to\infty} \hat{\theta} = \theta$ in probability. If assumption A2 holds with $k > 4$, then $\lim_{n\to\infty} \hat{\theta} = \theta$ almost surely (a.s.).*

Theorem 3.5 requires a limited growth of the time series length $T$ with respect to the number $n$ of sites. It is valid for instance when $T$ and $n$ satisfy a linear relationship or more generally when $T$ is bounded by a power of $n$. Another assumption concerning the marginal distribution of $X_{1,t}$ avoids such a restriction and leads to consistency. Define $\log_-(x) := -\log \min(x, 1)$ for any $x > 0$ and $\log_-^q(x) := (\log_-(x))^q$ for any $q > 0$.

Assumption A5: There exists $q > 2$ such that

$$\sup_{t\ge 1} \mathbb{E}\mathbb{1}_{X_{1,t}>0} \cdot \log_-^q X_{1,t} < \infty \quad \text{and} \quad \sup_{t\ge 1} \mathbb{E}\mathbb{1}_{X_{1,t}\neq X_{2,t}} \cdot \log_-^q |X_{1,t} - X_{2,t}| < \infty.$$

**Theorem 3.6.** *Given assumptions A1–A2–A3–A4–A5, $\hat{\theta}_n$ is weakly consistent if $k > 2$ and strongly consistent if $k > 4$ in assumption A2.*

### (b) A more specific result for integer-valued random variables

For integer-valued random variables, the previous assumptions can be simplified. In particular, it is straightforward to show that assumption A5 is automatically satisfied. Assumption A3 becomes:

Assumption A3′: We have $\inf_{t\ge1} \mathbb{P}(X_{1,t} > 0) > 0$ and $\inf_{t\ge1} \mathbb{P}(X_{1,t} \neq X_{2,t}) > 0$.

**Corollary 3.7.** *Given assumptions A1–A2–A3′–A4, $\hat{\theta}_n$ is weakly consistent if $k > 2$ and strongly consistent if $k > 4$ in assumption A2.*

## 4. Asymptotic normality of parameter estimates

We now derive asymptotic normality of our log-regression ordinary least-squares estimators. To keep the set-up as simple as possible, we also assume independence in the time direction, i.e. for any $j \ge 1$, we assume $X_{j,1}, X_{j,2}, \ldots$ are mutually independent. This assumption can be relaxed but this simple set-up suffices to give insight into the weak convergence of such estimates. For a matrix $C$ of any size, $C'$ denotes its transpose.

Define $Y_{j,t} := (X_{j,t}, X_{j,t}^2)'$ and $m_t := \mathbb{E}Y_{1,t}$ for $1 \le j \le n$ and $1 \le t \le T$. Define the empirical (or sample) version of $m_t$ by

$$\hat{m}_t := \frac{1}{n}\sum_{j=1}^{n} Y_{j,t}.$$

The two components of $m_t$ and $\hat{m}_t$ are denoted, respectively, by $m_{1,t}, m_{2,t}$ and $\hat{m}_{1,t}, \hat{m}_{2,t}$. In terms of our previous notations, we have $m_{1,t} = \mu_t$ and $\hat{m}_{1,t} = \hat{\mu}_t$. In what follows, we use both notations, depending on the context.

We also use a standard notation for weak convergence or convergence in distribution. When $(Z_n)_{n\ge1}$ is a sequence of random vectors of $\mathbb{R}^2$ and $\nu$ is a probability measure on the Borel $\sigma$-field of $\mathbb{R}^2$, notation $Z_n \Rightarrow \nu$ means that $Z_n$ converges weakly to $\nu$, i.e. the distribution of $Z_n$ converges to $\nu$.

The following analysis concerns a function $\phi_\theta$ with two-dimensional parameter $\theta := (\theta_1, \theta_2) \in \mathbb{R}^2$, two-dimensional domain $(0, \infty)^2$, and two-dimensional range $\mathbb{R}^2$. In the two-dimensional argument $(x, y)$ of $\phi_\theta$, $x$ represents the empirical or expected spatial mean of the observations $X_{j,t}$ and $y$ represents the empirical or expected spatial mean of the squared observations $X_{j,t}^2$. With either empirical or expected mean, $y - x^2$ represents the variance, i.e. the mean-squared observation minus the squared mean observation. The two parameters $\theta_1 = \log(\alpha)$, $\theta_2 = \beta$ of $\phi_\theta$ correspond to the intercept and the slope of the log–log form equation (2.2) of Taylor's law. The first component $\phi_{1,\theta}(x, y) := \log(y - x^2) - \theta_1 - \theta_2 \log(x)$ of $\phi_\theta$ is the deviation from equation (2.2), conditional on the two parameter values $\theta_1, \theta_2$. It corresponds to the first component of the normal equations of ordinary least-squares estimation, equations (2.3) and (2.4). The second component $\phi_{2,\theta}(x, y) := \log(x)\log(y - x^2) - \theta_1 \log(x) - \theta_2(\log(x))^2$ similarly corresponds to the second component of the normal equations of ordinary least squares.

Setting $\theta = (\log(\alpha), \beta)'$, the estimation error for the log-regression parameters has the decomposition

$$\hat{\theta} - \theta = \hat{D}_T^{-1}\hat{N}_T - D_T^{-1}N_T = \hat{D}_T^{-1}(\hat{N}_T - N_T + (D_T - \hat{D}_T)\theta).$$

This decomposition implies that

$$U_T := \hat{D}_T(\hat{\theta} - \theta) = \frac{1}{T}\sum_{t=1}^{T}(\phi_\theta(\hat{m}_t) - \phi_\theta(m_t)), \tag{4.1}$$

where

$$\phi_\theta(x, y) := \begin{pmatrix} \phi_{1,\theta}(x, y) \\ \phi_{2,\theta}(x, y) \end{pmatrix},$$

for $x > 0, y > 0$. The idea is simply to make a Taylor expansion of the function $\phi_\theta$ with respect to its two arguments representing the spatial means of the two samples $(X_{j,t})_{1\le j\le n}$ and $(X_{j,t}^2)_{1\le j\le n}$, respectively, computed at any time point $1 \le t \le T$. We then approximate $U_T$ with its first-order Taylor expansion

$$V_T = \frac{1}{T}\sum_{t=1}^{T} J_{\phi_\theta}(m_t) \cdot (\hat{m}_t - m_t), \tag{4.2}$$

where $J_{\phi_\theta}(x, y)$ denotes the Jacobian matrix of $\phi_\theta$ at point $(x, y)$. Intuitively, the convergence rate for $V_T$ is $\sqrt{nT}$ since $V_T$ contains a double summation with $n \times T$ centred terms. We then define

$$\Gamma_T := \mathrm{Var}(\sqrt{nT}V_T) = \frac{1}{T}\sum_{t=1}^{T} J_{\phi_\theta}(m_t) \cdot \mathrm{Var}(Y_{1,t}) \cdot J_{\phi_\theta}(m_t)' \tag{4.3}$$

and $C_T := \Gamma_T^{-1/2}$. It is expected that $\sqrt{nT}C_T V_T \Rightarrow \mathcal{N}_2(0, I_2)$. Here, $\mathcal{N}_2$ denotes the bivariate normal distribution with mean 0 and covariance matrix the $2 \times 2$ identity matrix, $I_2$. The remaining step is to investigate the asymptotic behaviour of $\sqrt{nT}(U_T - V_T)$. When $T/n \to 0$, one can show that under the assumptions of theorem 3.5, $\sqrt{nT}(U_T - V_T) = o_{\mathbb{P}}(1)$. However, when $T/n \to s > 0$, this is not the case, but under the assumptions of theorem 3.5, one can show that

$$\sqrt{nT}(U_T - V_T) = \sqrt{\frac{n}{T}}\sum_{t=1}^{T}\begin{pmatrix} (\hat{m}_t - m_t)'\nabla^{(2)}\phi_{1,\theta}(m_t)(\hat{m}_t - m_t) \\ (\hat{m}_t - m_t)'\nabla^{(2)}\phi_{2,\theta}(m_t)(\hat{m}_t - m_t) \end{pmatrix} + o_{\mathbb{P}}(1),$$

where for a twice-differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$, we denote by $\nabla^{(2)}f(m)$ the Hessian matrix of $f$ at point $m$. The first right-hand term in the previous equality does not vanish asymptotically and has an expectation of order $\sqrt{T/n}$. Hence, this non-negligible bias has to be included in deriving the asymptotic distribution of $\hat{\theta}$.

In what follows, we introduce two asymptotic regimes. In the first, we say that $T/n = o(1)$ if $\lim_{n\to\infty} T_n/n = 0$. In the second, we say that $T/n = O(1)$ if $\sup_{n\ge1} T_n/n < \infty$. The first regime is a special case of the second, which is satisfied when $T$ and $n$ are proportional or asymptotically proportional. However, the second regime requires special attention when the first does not hold.

For this reason, we distinguish between these two regimes in our next results. Our consistency result, theorem 3.5, is valid for these two regimes.

**Theorem 4.1.** *Let $\lambda_-(\Gamma_T)$ denote the smallest eigenvalue of $\Gamma_T$ defined in equation (4.3). Assume that $\underline{\lim}_n \lambda_-(\Gamma_T) > 0$. Suppose that assumptions A1–A2–A3–A4 hold with $k > 4$ in assumption A2 and for any integer $j$ such that $1 \le j \le n$, $(X_{j,t})_{t\ge 1}$ is a sequence of independent random variables.*

*(i) If $T/n = o(1)$, then*

$$\sqrt{nT} C_T \hat{D}_T(\hat{\theta} - \theta) \Rightarrow \mathcal{N}_2(0, I_2).$$

*(ii) If $T/n = O(1)$, then*

$$\sqrt{nT} C_T \hat{D}_T(\hat{\theta} - \theta) - \frac{1}{2}\sqrt{\frac{T}{n}} C_T E_T \Rightarrow \mathcal{N}_2(0, I_2),$$

*where $E_T := (1/T)\sum_{t=1}^{T} G_t$ and*

$$G_t := (Tr(\nabla^{(2)}\phi_{1,\theta}(m_t) Var(Y_{1,t})), Tr(\nabla^{(2)}\phi_{2,\theta}(m_t) Var(Y_{1,t})))'.$$

We comment on these results.

**Remark 4.2.** The existence of a lower bound for the eigenvalues of $\Gamma_T$ is always difficult to check for a specific sequence of marginal distributions, though quite natural. We provide sufficient conditions as well as a way to check it for Poisson random variables in lemma A.7 in appendix A.

**Remark 4.3.** When $T/n \to k > 0$ as $n \to \infty$, the normalized statistics $\sqrt{nT} C_T \hat{D}_T(\hat{\theta} - \theta)$ in theorem 4.1 become asymptotically biased. Neither the matrix $C_T$ nor the vector $E_T$ converges in general when $n \to \infty$ but these sequences remain bounded under our assumptions. This non-centring of the asymptotic distribution is atypical for regular statistical models and is due here to the fact that when the array of data remains balanced, there are not enough replicates of the time series to make the second-order term in the Taylor expansion of the estimates negligible. The asymptotics when $T/n = O(1)$ (theorem 4.1(ii)) also hold when $T/n = o(1)$, since then $\sqrt{T/n} C_T E_T = o(1)$ becomes negligible.

**Remark 4.4.** Figure 1 shows the effect of neglecting the bias for simulated Poisson data when the number of sites equals the number of time points. Without bias correction, the probability distributions of the two coordinates of the normalized statistics $\sqrt{nT} \hat{C}_T \hat{D}_T(\hat{\theta} - \theta)$ are not correctly approximated by that of a standard Gaussian. Here $\hat{C}_T$, which is a natural estimator of $C_T$, is defined in corollary 4.7.

**Remark 4.5.** An important question is whether theorem 4.1 can be extended to a scenario where $(X_{j,t})_{j,t}$ are dependent and non-stationary in time. To do so, it will be necessary to extend lemma A.5, which is a Marcinkiewicz–Zygmund type inequality. Extensions of these inequalities to dependent random fields are available, for instance, for mixing random fields [29, theorem 1] or another measure of dependence [30]. We also need to study the weak convergence of the self-normalized sums $\mathrm{Var}(V_T)^{-1/2} V_T$ to a standard Gaussian vector. A useful result for mixing random fields [31, theorem 3.3.1] does not require stationarity. In that case, the self-normalization $\mathrm{Var}(V_T)$ involves all the covariances between the different variables. Estimation of such normalization was carefully studied in the case of stationary random processes. The authors of [32] review the main approaches for time series. For random fields, the authors of [33] obtain nice results for the univariate case, under $\varphi$-mixing conditions. However, extension to non-stationarity in time is more challenging. First, suitable stationarity assumptions valid only in the locations direction $1 \le j \le n$ have to be carefully formulated to reduce the number of covariances to estimate. These variance estimators are non-parametric. Dealing with non-stationarity in time requires careful attention, certainly a huge amount of data, and development of multiple intermediate tools for dependent random fields. Such apparatus would hide the main and very simple purpose of our paper, which is to improve the practice of testing and interpreting Taylor's law. A reasonable approach to the dependent setting requires another more technical publication.

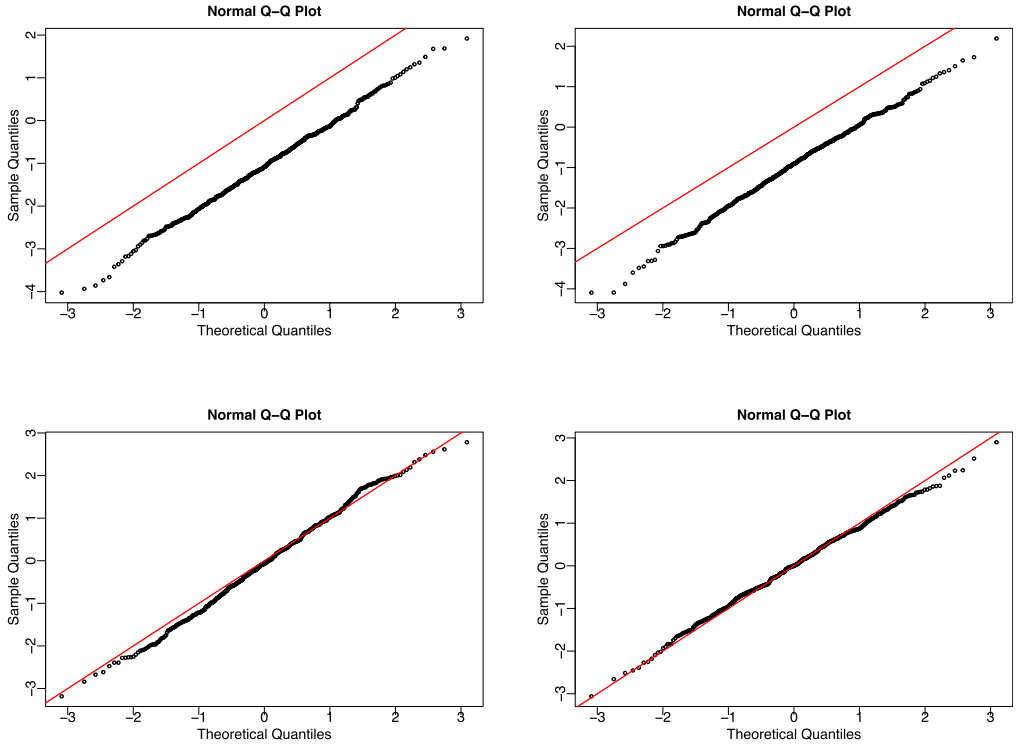


**Figure 1.** For simulated Poisson variates $X_{j,t} \sim \mathcal{P}(\lambda_t)$ with $n = T = 100$ and $\lambda_t = 3 + \cos(t)$, quantile–quantile plots (top) for the two components of statistics $\sqrt{nT}\hat{C}_T\hat{D}_T(\hat{\theta} - \theta)$ and (bottom) for the two components of statistics $\sqrt{nT}\hat{C}_T\hat{D}_T(\hat{\theta} - \theta) - \frac{1}{2}\sqrt{T/n}\hat{C}_T\hat{E}_T$. The abscissa represents the quantiles of the standard Gaussian distribution. The ordinate represents the sample quantiles for the statistics of interest. A good fit should show dots near or on the diagonal line where abscissa equals ordinate, as in the bottom two panels.

**Remark 4.6.** To use these results in practice for constructing tests or confidence intervals, it is necessary to estimate $\Gamma_T$ and $E_T$ from the data. Under our current independence assumption, the following corollary holds without additional requirement:

**Corollary 4.7.** *Suppose that all the assumptions of theorem 4.1 hold true with either the asymptotics* $T/n = o(1)$ *or* $T/n = O(1)$. *Set*

$$\hat{\Gamma}_T := \frac{1}{T}\sum_{t=1}^{T} J_{\phi_{\hat{\theta}}}(\hat{m}_t)\hat{\Sigma}_t J_{\phi_{\hat{\theta}}}(\hat{m}_t)', \quad \hat{C}_T := \hat{\Gamma}_T^{-1/2}$$

*and*

$$\hat{E}_T := \frac{1}{T}\sum_{t=1}^{T} \hat{G}_t, \quad \hat{G}_t := (Tr(\nabla^{(2)}\phi_{1,\hat{\theta}}(\hat{m}_t)\hat{\Sigma}_t), Tr(\nabla^{(2)}\phi_{2,\hat{\theta}}(\hat{m}_t)\hat{\Sigma}_t))'.$$

*Then*

$$\sqrt{nT}\hat{C}_T\hat{D}_T(\hat{\theta} - \theta) - \frac{1}{2}\sqrt{\frac{T}{n}}\hat{C}_T\hat{E}_T \Rightarrow \mathcal{N}_2(0, I_2). \tag{4.4}$$

*Define* $H_n := \hat{D}_T^{-1}\hat{\Gamma}_T\hat{D}_T^{-1}$ *and*

$$M_n := \begin{cases} 0, & \text{if } \frac{T}{n} = o(1), \\ \frac{1}{2}\sqrt{\frac{T}{n}}\hat{D}_T^{-1}\hat{E}_T, & \text{if } \frac{T}{n} = O(1). \end{cases}$$

*Let $q_\nu$ denote the quantile of order $\nu$ of the standard Gaussian distribution. Then the interval*

$$I_{i,n} = \left[\hat{\theta}_i - \frac{\sqrt{H_n(i,i)}q_{1-\nu/2} + M_n(i)}{\sqrt{nT}}, \hat{\theta}_i + \frac{\sqrt{H_n(i,i)}q_{1-\nu/2} - M_n(i)}{\sqrt{nT}}\right]$$

*is a confidence interval with asymptotic level $1-\nu$ for $i=1,2$. When $T/n = O(1)$, the vector $M_n$ corrects the bias of the two confidence intervals.*

**Remark 4.8.** In practice, it is difficult to know which regime is the most appropriate. However, if $T/n = o(1)$, we have $M_n = o(1)$, under our assumptions. So the bias correction can still be used in this case, though it is asymptotically negligible. Moreover, even when $T/n$ seems to be small, the bias correction $M_n$ depends on some empirical averages and can be non-negligible. This is why we recommend using such a bias correction whatever the context, or at least computing both confidence intervals and seeing how they differ from each other. The estimator $\hat{\theta} - M_n/\sqrt{nT}$ can be interpreted as a bias corrected estimator of $\theta$, the bias resulting from the estimation of the nuisance parameters $(\mu_t, \sigma_t)$ in the OLS equations. See [34] for a similar correction in the context of maximum likelihood estimation.

**Remark 4.9.** What asymptotic behaviour could be expected when $T/n \to \infty$? In this case, a higher-order Taylor expansion is needed to control the difference between $U_T$ and the linear part $V_T$ discussed before the statement of theorem 4.1. Suppose, for instance, that $T/n^{k-1} = O(1)$ for some $k \geq 3$. (Our study concerns only the case $k=2$.) Under suitable moment assumptions, one expects that

$$\sqrt{nT}(U_T - V_T) = \sum_{\ell=2}^{k} \frac{1}{\ell!}\sqrt{\frac{n}{T}}\sum_{t=1}^{T} D^{(\ell)}\phi_\theta \cdot [\hat{m}_t - m_t]^\ell + o_{\mathbb{P}}(1),$$

where $D^{(\ell)}$ denotes the formal differential operator of order $\ell$. Indeed, by using lemma A.5, the expectation of $||\hat{m}_t - m_t||^\ell$ is of order $n^{-\ell/2}$, and all the terms of order lower than $k$ in the Taylor expansion are non-negligible, leading to a more important bias. In order to provide a bias correction, it is necessary to estimate these $k-1$ terms. We do not pursue the required additional calculations in this paper, mainly because $T \gg n$ is quite rare for real data, especially for ecological abundance data.

## 5. Example: international bottom Trawl Survey for the North Sea

We analyze some data from the International Council for the Exploration of the Sea (ICES) International Bottom Trawl Survey for the North Sea. The survey estimates the abundance, measured as catch per unit effort (CPUE), of each fish and some invertebrate species binned into length classes, in spatial rectangles, known as subareas, of 0.5° latitude by 1.0° longitude, with an approximate area of 30 $\text{km}^2$. Data included in this study are limited to the years 1977 to 2015. The dataset used here contains only the observations of the nine standard species: *Clupea harengus* (species 1), *Gadus morhua* (species 2), *Melanogrammus aeglefinus* (species 3), *Merlangius merlangus* (species 4), *Pleuronectes platessa* (species 5), *Pollachius virens* (species 6), *Scomber scombrus* (species 7), *Sprattus sprattus* (species 8) and *Trisopterus esmarkii* (species 9). The CPUE is the number of individuals caught per cruise in 30 min of fishing, where only individuals in certain categories of length (measured in mm) were included. Further, data were limited to quarter one only, when survey cruises typically occur from mid-January to mid-February. Thus it excludes seasonal differences in abundance. For each of the nine species, the abundance is recorded at $n = 195$ locations in $T = 39$ years, with one observation per year. More details about this survey and about interpreting the TL slope $\beta$ in ecology are given in [35] and its references. In general, the slope parameter $\beta$ has been shown to be connected to several ecological indicators such as demographic rates, habitat heterogeneity, demographic stochasticity, environmental heterogeneity and sampling procedures. See [36–39] for more general discussions about ecological interpretations of some changes in spatial TL slope.

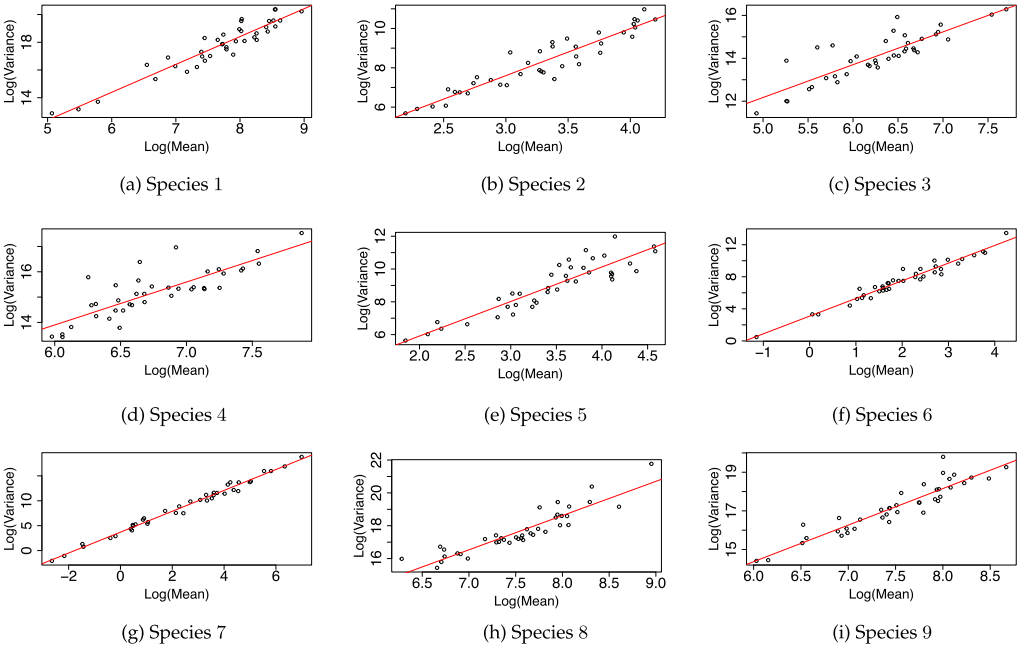


**Figure 2.** For each of nine species, a plot of the natural logarithm of the sample variance as a function of the natural logarithm of the sample mean (black dots) of the catch per unit effort (CPUE). The ordinary least-squares regression lines are red.

Figure 2 gives log–log plots (using the natural logarithm) of the empirical mean and the empirical variance of the CPUE of the nine species.

Table 1 lists the ordinary least-squares estimate of the slope parameter as well as the 95% confidence intervals obtained from standard linear regression methods. The function lm of the sofware R was used to get least-squares estimators of $\theta$ as well as the confidence interval for the slope parameter $\beta$. These confidence intervals are not valid under the present set-up since the input and output variables are, respectively, logarithms of empirical means and empirical variances, a case not covered by the usual asymptotic theory. In contrast, table 2 lists confidence intervals computed from our asymptotic results, with and without bias correction, for the same confidence level. We do not consider species 7 here because more than 75% of these data were zeros and we had numerical difficulties evaluating the quantities in corollary 4.7. Very sparse data could thus be problematic. For such data, log–log regression is not suitable. We could use other methods, e.g. by fitting a parametric zero-inflated distribution to the data [40], or by modelling separately the zero values and the positive ones, e.g. [41]. The estimate of TL slope will then depend on the chosen model. Such alternative methods are beyond the scope of this paper. Here we do not assume a specific parametric model for the $X_{j,t}$ values.

Also, we divided all the abundances by a factor (the mean value of the abundance of the species, computed over the $n \times T$ observations) to avoid numerical issues owing to many large values in the time series. This operation does not change the value of the estimate of the slope parameter $\beta$ because Taylor's law is still valid with the same $\beta$, although it changes the value of the intercept parameter $\alpha$. A pseudo-code summarizing the different steps for computing the confidence intervals for the slope can be found in Section 3 of the electronic supplementary material.

The confidence intervals listed in table 1 are wider than those listed in table 2. This is not surprising since their size is proportional to $1/\sqrt{T}$ while the correct size is proportional to $1/\sqrt{nT}$. As a consequence, after this correction of the width of the confidence intervals, some values of $\beta$ are rejected at the 95% level; for instance, the value $\beta = 2$ for species 4. The two intervals listed table 2 with or without bias correction have exactly the same length. The confidence interval

**Table 1.** Point estimates by ordinary least squares of the slope parameter $\beta$ for the nine species and lower and upper bounds of the 95% confidence interval obtained with conventional linear regression ($n = 195$, $T = 39$).

| species | ordinary least squares $\beta$ | lower bound | upper bound |
|---|---|---|---|
| 1 | 2.009793 | 1.780779 | 2.238807 |
| 2 | 2.390582 | 2.1058999 | 2.675264 |
| 3 | 1.530209 | 1.221486 | 1.838933 |
| 4 | 1.696704 | 1.301093 | 2.092315 |
| 5 | 2.106150 | 1.780474 | 2.431825 |
| 6 | 2.205332 | 2.041850 | 2.368813 |
| 7 | 2.09691 | 2.015611 | 2.178211 |
| 8 | 2.084444 | 1.779885 | 2.389003 |
| 9 | 1.893228 | 1.643332 | 2.143124 |

**Table 2.** Point estimates by ordinary least squares of the slope parameter $\beta$ for eight of the nine species and lower and upper bounds of the 95% confidence interval ($n = 195$, $T = 39$) computed with bias correction (C) and without bias correction (NC). (C) (respectively, (NC)) means that the formula of the confidence interval given in corollary 4.7 is computed under the regime $T/n = O(1)$ (respectively, $T/n = o(1)$).

| species | lower bound NC | upper bound NC | lower bound C | upper bound C |
|---|---|---|---|---|
| 1 | 1.89 | 2.13 | 1.92 | 2.17 |
| 2 | 2.23 | 2.55 | 2.29 | 2.61 |
| 3 | 1.38 | 1.68 | 1.35 | 1.65 |
| 4 | 1.43 | 1.96 | 1.36 | 1.89 |
| 5 | 1.93 | 2.29 | 1.97 | 2.32 |
| 6 | 2.07 | 2.34 | 2.21 | 2.47 |
| 8 | 1.86 | 2.30 | 1.98 | 2.42 |
| 9 | 1.74 | 2.05 | 1.77 | 2.09 |

obtained with bias correction is a simple shift, to the left or to the right, of that obtained without bias correction and it is then no longer centred around the estimate $\hat{\beta}$.

With or without bias correction, the confidence intervals of four species, $1, 5, 8, 9$, include the value $\beta = 2$; three intervals lie entirely above $\beta = 2$; and two intervals lie entirely below $\beta = 2$. Thus, the hypothesis $\beta = 2$ is not rejected for only four species, under the important, probably influential assumption, for each species, that its time series are independently and identically distributed at different spatial locations.

In another application of our results, we test whether the slope $\beta$ of a first species is statistically significantly greater than the slope $\overline{\beta}$ of a second species, i.e. we test $H_0 : \beta \leq \overline{\beta}$ versus $H_1 : \beta > \overline{\beta}$. Assuming that the two arrays are independent and both satisfy the assumptions of corollary 4.7, one can show that the statistic

$$\xi_n := \frac{(\hat{\beta} - \hat{\overline{\beta}} - \beta + \overline{\beta} - \hat{u}_n)}{\hat{v}_n}$$

is asymptotically $\mathcal{N}(0, 1)$-distributed, with

$$\hat{u}_n = \frac{M_n(2) - \overline{M}_n(2)}{\sqrt{nT}} \quad \text{and} \quad \hat{v}_n = \sqrt{\frac{H_n(2,2) + \overline{H}_n(2,2)}{nT}}.$$

Here, $M_n$ and $H_n$ are calculated using the data of the first species with slope $\beta$ as in corollary 4.7, and $\overline{M}_n$ and $\overline{H}_n$ are calculated in identical fashion using the data of the second species with slope $\overline{\beta}$. We then reject $H_0$ at level $\nu$ if $\xi_n > q_{1-\nu}$. For instance, if $(\beta, \overline{\beta})$ denotes the slopes of the pair of species $(6, 5)$, the value of $\xi_n$ is 1.57 and $H_0$ is not rejected at the 5% level, showing that the estimated slope for species 6 is not significantly larger than that of species 5. On the other hand, species 2 has an estimated slope significantly larger than the estimated slope of species 5 because $\xi_n = 2.46$. One can also use the same procedure to test if the slope of spatial TL over one period of time differs significantly from the slope for another time period. In this case, $\hat{\beta}$ and $\hat{\overline{\beta}}$ denote the estimated slopes over the first and second time period, respectively.

Our methods (and most other applications of least-squares regression to estimate the parameters of Taylor's law) ignore the possibility that exogenous environmental fluctuations like the North Atlantic Oscillation influence all locations simultaneously, as well as the possibility that spatially adjacent or spatially remote populations of a species interact.

Visual inspection of figure 2 gives no suggestion, for any species, that log variance is systematically nonlinearly related to log mean. In an exploratory analysis, we fitted a quadratic spatial Taylor's law, that is,

$$\log \sigma_t^2 = \log(\alpha) + \beta \log \mu_t + \gamma \log^2 \mu_t,$$

by ordinary least squares as implemented in the software R, with the standard lm function used to fit linear models. The null hypothesis that $\gamma = 0$ was not rejected for any species and the $p$-values were all large. This finding should be treated with caution as the $t$-tests are not credible in the context of this application. It is necessary to extend our results to this more general parametric model to be more rigorous. Despite these cautions, it seems doubtful that there is any important quadratic effect. Further theoretical developments are required to assess rigorously whether the data reject the null hypothesis of a linear relation between log spatial mean and log spatial variance. For now, we accept the visual evidence that Taylor's law, linear on log–log coordinates, is a plausible model of the variance function of these data.

The present paper does not analyze the stochastic dependence between observations. Recent theoretical analyses of stochastic dependence in Taylor's law include simulations [39] and an analysis of a single trajectory of a time series with serial dependence based on self-normalized quantities that involve Bernstein blocks [42]. In our present context, considering dependence will lead to a more delicate estimation of the normalizing matrix $C_T$ in theorem 4.1, and this would probably require large datasets to get accurate results.

To decide whether this dataset is compatible with the assumption of temporal independence under a non-stationary framework, we compute the residuals $\hat{Z}_{j,t} = (X_{j,t} - \hat{\mu}_t)/\hat{\sigma}_t$ which behave more like a stationary time series since their two first moments approximate 0 and 1, respectively. For each species, we then compute the three first autocorrelation coefficients $\rho(1)$, $\rho(2)$, $\rho(3)$, with lags 1, 2, 3 of each time series $(\hat{Z}_{j,t})_{1 \le t \le T}$ for $j = 1, \ldots, n$. We apply the arctanh function to all the correlation coefficients as a variance-stabilizing function. Van der Vaart [43, section 3.2] defines the corresponding confidence interval of the autocorrelation coefficients, based on this variance-stabilizing function.

Table 3 lists the percentage of confidence intervals that include 0 for each species and each autocorrelation coefficient. The confidence intervals were calculated under the assumption of independence across $t$ of $Z_{j,t} = (X_{j,t} - \mu_t)/\sigma_t$, and by neglecting the sampling error to estimate the local mean and local variance. Since $n = 195$ is much larger than $T = 39$ for this example, one might expect that this sampling error is negligible. However, the results have to be taken with caution since $T$ is not very large. For species 6–9, the assumption of no correlation across time appears as perfectly plausible. Species 3 seems to be the most problematic.

In table 4, we apply the same procedure to test spatial correlation. More precisely, for each of the nine species separately, we compute the sample or empirical correlation $r$ between $Z_{j,t}$ and $Z_{\ell,t}$ for two different locations $j$ and $\ell$ using the two samples $(Z_{j,t})_{1 \le t \le T}$ and $(Z_{\ell,t})_{1 \le t \le T}$. Table 4 lists the percentage of pairs $(j, \ell)$ for which $r = 0$ is not rejected. As for temporal dependence, it

**Table 3.** Percentage of confidence intervals containing the point $\rho(i) = 0$ for serial autocorrelation with lag $i = 1, 2, 3$. The confidence intervals were computed for the $n = 195$ time series using the arctanh variance-stabilizing function.

| species | $\rho(1)$ | $\rho(2)$ | $\rho(3)$ |
|---|---|---|---|
| 1 | 77.4 | 80.0 | 99.0 |
| 2 | 75.4 | 89.2 | 91.3 |
| 3 | 48.2 | 51.3 | 52.8 |
| 4 | 70.0 | 79.0 | 82.0 |
| 5 | 81.5 | 85.6 | 90.3 |
| 6 | 97.9 | 98.0 | 100.0 |
| 7 | 95.9 | 98.5 | 99.0 |
| 8 | 97.4 | 97.4 | 96.9 |
| 9 | 95.4 | 95.4 | 96.9 |

**Table 4.** Percentage of confidence intervals containing the point $r = 0$ for correlation of the abundance between different locations. The confidence intervals were computed using the arctanh variance-stabilizing function.

| species | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| percentage | 83 | 84.5 | 60.4 | 83.8 | 63.4 | 49.4 | 68.1 | 66.7 | 66 |

is difficult to exclude a spatial correlation for some species, such as species 6, whereas for some other species, spatial correlation seems not to be very important.

## 6. Conclusions and perspectives

In this work, we introduced precise assumptions on the data-generating process to study the asymptotic properties of least-squares estimators of Taylor's power law parameters. We have given a first rigorous basis for estimating the parameters of Taylor's law. We have shown that our methods apply to and illuminate real data. Our empirical results suggest that our independence assumptions are a reasonable first approximation for this dataset. Theorems 3.5, 3.6 and 4.1 are our main new results. Our assumptions on the marginal distributions are general enough to cover many situations encountered in practice, excluding data with fat tails and spatial and/or temporal dependence. The asymptotic properties of the regression estimates depend importantly on the asymptotic relation between the time-series length $T$ and the number $n$ of spatial points.

To ecologists interested in using our results to quantify the uncertainty in estimates of the parameters of Taylor's law, we offer the following guidance.

— Always apply the confidence intervals corresponding to the bias correction given in corollary 4.7, with $M_n$ defined as if $T/n = O(1)$. This correction is crucial when $T$ and $n$ are of the same magnitude. Even if the ratio $T/n$ seems to be quite small, it is difficult to know at exactly what threshold the asymptotic regime $T/n = o(1)$ becomes appropriate. The formula with bias correction always works asymptotically when $T/n = o(1)$. Even for finite sample sizes, the confidence intervals should not be too much affected.

— The most important distortion of the confidence intervals could come from spatial or temporal dependence. We ignored such possible dependence in this work. We recommend checking whether the independence assumption is compatible with the dataset, as we did, for instance, in our application to real data, §5. If substantial

dependence is observed, as in [39], further research is needed. Our specific formulas for confidence intervals may or may not apply to data with demonstrable spatial or temporal dependence, but we believe our approach to modelling and analysis of Taylor's law parameter estimation generalizes even to data with dependence.

Some hints for various extensions are given below. When data have fat tails, the limiting distribution of least-squares estimators will probably no longer be Gaussian. Even with thin tails, our contribution appears to be the first to derive an asymptotic distribution for such estimates, which appear in numerous ecological papers without theoretical analysis.

Various extensions could also be investigated. For instance, one can study other estimators of the form

$$\hat{f} = \arg\min_{f \in \mathcal{F}} \sum_{t=1}^{T} (\log \hat{\sigma}_t^2 - f(\log \hat{\mu}_t))^2, \tag{6.1}$$

where $\mathcal{F}$ is a class of functions not necessarily linear. One can then study generalizations of Taylor's law such as those mentioned in §1, including quadratic [5,6] and cubic [7] functions of the log mean, as well as a linear combination of the mean and the square of the mean [8] or higher-degree polynomial functions of the mean.

The case of an infinite-dimensional class of functions is also challenging. Whatever the regularity (e.g. differentiability) imposed on $\mathcal{F}$, an infinite number of functions could map the sequence $(\log \mu_t)_t$ on to the sequence $(\log \sigma_t^2)_t$. This non-parametric extension differs from the usual standard non-parametric estimation problems in statistics.

Another open question related to assumption A1 is to check whether time and space ($t$ and $j$) could be exchanged. Non-stationarity in time of at least both mean and variance (and not only one of them) seems mandatory to get identifiable parameters $\alpha, \beta$ in the spatial Taylor's law equation (2.1). Estimating the normalizing quantities could be more difficult in this non-stationary context. Taking account of the stochastic dependence, spatial or temporal or both, between the observations is also a technical challenge. Many ways to extend theorem 4.1 in the context of dependence have been published; for instance, [44] provide many helpful asymptotic results.

Extensions to dependent random variables are plausible in different settings. When extending our work to allow spatial and temporal dependence, a similar theoretical analysis could be conducted using our proof procedure with a ratio $T/n$ of crucial importance in the statement of the corresponding results. A variant definition of Taylor's law together with a simple extension of Taylor's law under mixing conditions was suggested in [45]. An analysis of the asymptotic behaviour of Taylor's law for processes with heavy tails, correlations and mixtures of different distributions was carried out by [23]. A systematic approach to a weaker condition such as long-range dependence was provided by [46], and a comprehensive approach to long-range dependence under heavy-tailed assumptions was provided by [47]. We plan to address these open questions in future work.

Finally, many stochastic dynamics exhibit a time-varying fluctuation scaling. For instance, [48,49] modelled non-stationary dynamics with a time-varying slope of Taylor's law; these models have been applied to financial data. Cobain *et al.* [35] also discussed possible time-variation of the slope parameters for the same kind of data used in our paper. Extending our results to a time-varying set-up is possible but requires specifying additional assumptions on the slope. In time-series analysis, smooth time-varying parameters can be estimated using the locally stationary framework [50]. This approach is related to non-parametric statistical inference and local estimation using kernels. Break detection when the slope parameter is piecewise constant is another challenge. Andrews [51] developed tests for break points in general parametric models. These extensions require additional investigations. Our contribution in this paper could help to adapt estimation of TL parameters to specific time windows that may be required for using these more complex methods.

## 7. Proofs of the results

Here we prove our main results. Some proofs use technical lemmas in appendix A. The proof is difficult because we did not find adequate a.s. convergence results adapted to the current triangular arrays setting. This led us to a systematic use of the Borel–Cantelli lemma and the bounds for the probability of the events on which the estimates do not have a satisfactory behaviour. Those quantities are controlled through standard exponential inequalities or through the Markov inequality and suitable moment inequalities.

From now on, we denote by $|\cdot|$ the Euclidean norm on $\mathbb{R}^2$: $|(x,y)| := (x^2+y^2)^{1/2}$. We also denote the corresponding operator norm on the set of $2\times 2$ matrices by $|\cdot|$, that is, $|A| := \sup\{|Az|/|z| \mid z \in \mathbb{R}^2\}$. Finally, if $X$ is a random variable, we denote by $||X||_p = \mathbb{E}^{1/p}[|X|^p]$ the $\mathbb{L}_p$-norm of $X$, for any $p \geq 1$.

### (a) Proof of theorem 3.5

The proof will be based on the technical lemma A.1 in appendix A. Lemma A.1 controls the probability of getting small values for empirical means and empirical variances. We then use the constants depending on $\epsilon$ and $L$ defined in lemma A.1. Set $\gamma \in (0, L)$. Define the following event:

$$E_n(\epsilon) := \bigcap_{t=1}^{T} \{\hat{m}_t > \epsilon, \hat{\sigma}_t^2 > \epsilon\}.$$

This sequence of events depending on $n$ takes care of the properties of the estimates when they take small values. Using the union bound and lemma A.1, we have $\mathbb{P}(E_n(\epsilon)^c) \leq 2T\exp(-nL)$.

This shows that $\sum_{n\geq 2} \mathbb{P}(E_n(\epsilon)^c) < \infty$.

We now study the consistency of $\hat{\theta}$. We only need to show that $\hat{D}_T - D_T$ and $\hat{N}_T - N_T$ go to 0 in probability or a.s. Indeed, if we use the decomposition

$$\hat{\theta} - \theta = \hat{D}_T^{-1}(\hat{N}_T - N_T + (D_T - \hat{D}_T)D_T^{-1}N_T),$$

assumptions A2 and A4 guarantee that $N_T = O(1)$ and $D_T^{-1} = O(1)$. There then also exists $\delta > 0$ sufficiently small such that $\sup_{|H|\leq\delta} |(D_T + H)^{-1} - D_T^{-1}| = O(1)$. If $\hat{D}_T - D_T$ and $\hat{N}_T - N_T$ go to 0 in probability or a.s., respectively, we thus conclude that $\hat{D}_T^{-1} - D_T^{-1}$ converges to 0 in probability or a.s. and the weak or strong consistency of $\hat{\theta}$ will then follow. Now, for some $\eta > 0$, we have

$$\mathbb{P}(|\hat{\theta} - \theta| > \eta) \leq \mathbb{P}(\{|\hat{\theta} - \theta| > \eta\} \cap E_n(\epsilon)) + \mathbb{P}(E_n(\epsilon)^c). \tag{7.1}$$

We use the following inequality extensively. If $a, b \geq \epsilon$ and $\kappa \in (0,1)$, we have

$$|\log a - \log b| = \frac{1}{\kappa}|\log a^\kappa - \log b^\kappa| \leq \frac{1}{\kappa\epsilon^\kappa}|a-b|^\kappa.$$

This inequality turns the log-problems across the origin into easier moment bounds. To get this inequality, one can apply the mean value theorem, then use the inequality $|a^\kappa - b^\kappa| \leq |a-b|^\kappa$ for $\kappa \in (0,1)$.

On $E_n(\epsilon)$, we get $|\log\hat{\sigma}_t^2 - \log\sigma_t^2| \leq (1/\kappa\epsilon^\kappa)|\hat{\sigma}_t^2 - \sigma_t^2|^\kappa$.

From assumption A2 and standard moment inequalities for partial sums (lemma A.5), we have

$$\sup_{t\geq 1} \mathbb{E}|\hat{\sigma}_t^2 - \sigma_t^2|^\kappa \leq \sup_{t\geq 1} \mathbb{E}^{2\kappa/k}|\hat{\sigma}_t^2 - \sigma_t^2|^{k/2} = o(1).$$

By choosing $\kappa$ small enough in the previous bounds, for any positive integer $i$, we get

$$\sup_{t\geq 1} \mathbb{E}[|\log\hat{\sigma}_t^2 - \log\sigma_t^2|^i \mathbb{1}_{E_n(\epsilon)}] = o(1) \tag{7.2}$$

and

$$\sup_{t\geq 1} \mathbb{E}[|\log\hat{\mu}_t - \log\mu_t|^i \mathbb{1}_{E_n(\epsilon)}] = o(1). \tag{7.3}$$

On the event $E_n(\epsilon)$, $T^{-1}\sum_{t=1}^{T}[\log\hat{\sigma}_t^2 - \log\sigma_t^2] \to 0$ in $\mathbb{L}^1$ and then in probability when $n \to \infty$. Now, setting $C := \max(\sup_{t\geq 1}|\log\sigma_t^2|, \sup_{t\geq 1}|\log\mu_t|)$, we bound $|\log\hat{\mu}_t\log\hat{\sigma}_t^2 - \log\mu_t\log\sigma_t^2|$ by

$$|\log\hat{\mu}_t - \log\mu_t|\cdot|\log\hat{\sigma}_t^2 - \log\sigma_t^2| + |\log\sigma_t^2|\cdot|\log\hat{\mu}_t - \log\mu_t| + |\log\mu_t|\cdot|\log\hat{\sigma}_t^2 - \log\sigma_t^2|$$
$$\leq \frac{1}{2}|\log\hat{\sigma}_t^2 - \log\sigma_t^2|^2 + \frac{1}{2}|\log\hat{\mu}_t - \log\mu_t|^2 + C|\log\hat{\mu}_t - \log\mu_t| + C|\log\hat{\sigma}_t^2 - \log\sigma_t^2|.$$

From equations (7.2) and (7.3), we also deduce that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\mathbb{1}_{E_n(\epsilon)}|\log\hat{\mu}_t\log\hat{\sigma}_t^2 - \log\mu_t\log\sigma_t^2| = o(1).$$

Hence $\hat{N}_T - N_T \to 0$ in probability on the event $E_n(\epsilon)$. Similarly, $\hat{D}_T - D_T \to 0$ in probability on the event $E_n(\epsilon)$. This leads to the weak consistency of the regression estimates.

For any $\eta > 0$,

$$\sum_{n\geq 1}\mathbb{P}(\{|\hat{\theta} - \theta| > \eta\} \cap E_n(\epsilon)^c) \leq \sum_{n\geq 1}\mathbb{P}(E_n(\epsilon)^c) < \infty$$

and the Borel–Cantelli lemma ensures that $|\hat{\theta} - \theta|\cdot\mathbb{1}_{E_n(\epsilon)^c}$ converges to 0 a.s.

To prove the a.s. convergence, it is enough to prove that $\hat{N}_T - N_T \to 0$ and $\hat{D}_T - D_T \to 0$ a.s. on the event $E_n(\epsilon)$. Using the same strategy as for the convergence in probability, it suffices to show that

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\{|\hat{\sigma}_t^2 - \sigma_t^2| + |\hat{\mu}_t - \mu_t|\} = 0, \quad \text{a.s.} \tag{7.4}$$

Because the function $x \mapsto |x|^{k/2}$ (where $k > 2$ is given in assumption A2) is convex, we get

$$M_T := \left(\frac{1}{T}\sum_{t=1}^{T}|\hat{\sigma}_t^2 - \sigma_t^2|\right)^{k/2} \leq \frac{1}{T}\sum_{t=1}^{T}|\hat{\sigma}_t^2 - \sigma_t^2|^{k/2}.$$

From moment inequalities for partial sums of independent observations (lemma A.5), we get $\mathbb{E}M_T = O(n^{-k/4})$ and by assumption, if $k > 4$, we get $\sum_{n\geq 1}\mathbb{E}M_T < \infty$. Setting $P_T = |\hat{\mu}_t - \mu_t|$, similar arguments yield $\sum_{n\geq 1}\mathbb{E}P_T < \infty$. The Borel–Cantelli lemma ensures that both $M_T$ and $P_T$ go to zero a.s. as $n \to \infty$ and then the validity of equation (7.4), which concludes the proof.

## (b) Proof of theorem 3.6

The proof is similar to that of theorem 3.5. The main difference is in proving the convergence of $\hat{D}_T - D_T$ or $\hat{N}_T - N_T$ to 0. Let $A_t = \{\hat{\mu}_t > \epsilon, \hat{\sigma}_t^2 > \epsilon\}$ where $\epsilon < 1$ satisfies the condition in lemma A.1(ii). Using arguments similar to those in the proof of theorem 1, one can show that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{A_t}[|\log\hat{\sigma}_t^2 - \log\sigma_t^2| + |\log\hat{\mu}_t - \log\mu_t|$$
$$+ |\log^2\hat{\mu}_t - \log^2\mu_t| + |\log\hat{\mu}_t\log\hat{\sigma}_t^2 - \log\mu_t\log\sigma_t^2|]$$

converges to 0 either in probability or a.s., depending on $k$. From lemma A.1,

$$\sum_{n\geq 1}\frac{1}{T}\sum_{t=1}^{T}\mathbb{P}(A_t^c) < \infty,$$

and from the Borel–Cantelli lemma, we deduce that $\lim_{n\to\infty}(1/T)\sum_{t=1}^{T}\mathbb{1}_{A_t^c} = 0$ a.s. Moreover, from lemma A.1, the sequences $(\mu_t)_{t\geq 1}$ and $(\sigma_t^2)_{t\geq 1}$ are upper- and lower-bounded. To show the

required convergence, it suffices to show that the following quantities converge a.s. to 0:

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{0<\hat{\sigma}_t^2<\epsilon}\cdot|\log\hat{\sigma}_t^2|,\quad \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{0<\hat{\sigma}_t^2<\epsilon}\cdot|\log\hat{\sigma}_t^2\cdot\log\hat{\mu}_t|$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{0<\hat{\mu}_t<\epsilon}\cdot|\log\hat{\sigma}_t^2\cdot\log\hat{\mu}_t|,\quad \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{0<\hat{\mu}_t<\epsilon}\cdot|\log\hat{\mu}_t|,\quad \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{0<\hat{\mu}_t<\epsilon}\cdot|\log^2\hat{\mu}_t|.$$

We recall that $\log\hat{\mu}_t$ and $\log\hat{\sigma}_t^2$ are always implicitly defined, respectively, by $\log\hat{\mu}_t\cdot\mathbb{1}_{\hat{\mu}_t>0}$ and $\log\hat{\sigma}_t^2\cdot\mathbb{1}_{\hat{\sigma}_t^2>0}$.

First, because $\log(\epsilon)<0$, we have

$$\log\hat{\mu}_t\cdot\mathbb{1}_{0<\hat{\mu}_t<\epsilon}=\log\hat{\mu}_t\cdot\mathbb{1}_{0<\hat{\mu}_t<\epsilon}\cdot\mathbb{1}_{\cup_{j=1}^{n}\{X_{j,t}>0\}}\geq\sum_{j=1}^{n}\log\left(\frac{X_{j,t}}{n}\right)\cdot\mathbb{1}_{X_{j,t}>0}\cdot\mathbb{1}_{0<\hat{\mu}_t<\epsilon}.$$

We then get the bound

$$|\log\hat{\mu}_t|\cdot\mathbb{1}_{\hat{\mu}_t>0}\leq|\log\hat{\mu}_t|\cdot\mathbb{1}_{\hat{\mu}_t>\epsilon}-\sum_{j=1}^{n}\log\left(\frac{X_{j,t}}{n}\right)\cdot\mathbb{1}_{X_{j,t}>0}\cdot\mathbb{1}_{0<\hat{\mu}_t<\epsilon}.\tag{7.5}$$

Using the same type of arguments for $\hat{\sigma}_t^2=(1/2n^2)\sum_{1\leq i\neq j\leq n}|X_{i,t}-X_{j,t}|^2$, we get the bound

$$\begin{aligned}|\log\hat{\sigma}_t^2|\cdot\mathbb{1}_{\hat{\sigma}_t^2>0}&\leq|\log\hat{\sigma}_t^2|\cdot\mathbb{1}_{\hat{\sigma}_t^2>\epsilon}\\&\quad-\sum_{1\leq i\neq j\leq n}\log\left(\frac{|X_{i,t}-X_{j,t}|^2}{2n^2}\right)\cdot\mathbb{1}_{X_{i,t}\neq X_{j,t}}\cdot\mathbb{1}_{0<\hat{\sigma}_t^2<\epsilon}.\end{aligned}\tag{7.6}$$

We are going to show that, as $n\to\infty$,

$$Z_T:=\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{0<\hat{\sigma}_t^2<\epsilon}\cdot\log\hat{\sigma}_t^2\cdot\log\hat{\mu}_t\to 0\tag{7.7}$$

in probability or a.s. The idea for all the other quantities is exactly the same and details will be omitted.

Let $p=q/2>1$ by assumption A5 and $\alpha=1/(2p)$. From equation (7.5), there exists $C_q>0$ such that

$$||\log\hat{\mu}_t\cdot\mathbb{1}_{\hat{\mu}_t>0}||_{2p}\leq C_q\cdot\mathbb{E}^{\alpha}(\hat{\mu}_t)+\sum_{j=1}^{n}||\log\frac{X_{j,t}}{n}\mathbb{1}_{X_{j,t}>0}||_{2p}.$$

From A5, we deduce the existence of a constant $D>0$ such that the behaviour of $\hat{\mu}_t$ around the origin is not too explosive:

$$||\log\hat{\mu}_t\cdot\mathbb{1}_{\hat{\mu}_t>0}||_{2p}\leq D(1+n\log(n)).\tag{7.8}$$

From equation (7.6) and the Hölder inequality, we deduce that

$$\begin{aligned}&||\log\hat{\mu}_t\cdot\log\hat{\sigma}_t^2\cdot\mathbb{1}_{0<\hat{\sigma}_t^2<\epsilon}||_1\\&\leq\sum_{1\leq i\neq j\leq n}\mathbb{P}(\hat{\sigma}_t^2<\epsilon)^{1-1/p}\cdot||\log\hat{\mu}_t\cdot\mathbb{1}_{\hat{\mu}_t>0}||_{2p}\cdot||\log\left(\frac{|X_{i,t}-X_{j,t}|^2}{2n^2}\right)\cdot\mathbb{1}_{X_{i,t}\neq X_{j,t}}||_{2p}.\end{aligned}$$

From assumption A5, inequality (7.8) and lemma A.1, it is not very difficult to conclude that

$$\sum_{n\geq 1}\frac{1}{T}\sum_{t=1}^{T}||\log\hat{\mu}_t\cdot\log\hat{\sigma}_t^2\cdot\mathbb{1}_{0<\hat{\sigma}_t^2<\epsilon}||_1<\infty.$$

From the Borel–Cantelli lemma, we get equation (7.7). Remembering the discussion at the beginning of the proof, one can deduce the consistency properties of $\hat{\theta}$.

## (c) Proof of theorem 4.1

The second derivatives of the function $\phi_\theta = (\phi_{1,\theta}, \phi_{2,\theta})$ are

$$\frac{\partial^2\phi_{1,\theta}}{\partial x^2}(x,y) = -\frac{2}{y-x^2} - \frac{4x^2}{(y-x^2)^2} + \frac{\theta_2}{x^2},$$

$$\frac{\partial^2\phi_{1,\theta}}{\partial y^2}(x,y) = \frac{-1}{(y-x^2)^2}, \quad \frac{\partial^2\phi_{1,\theta}}{\partial x\partial y}(x,y) = \frac{2x}{(y-x^2)^2},$$

$$\frac{\partial^2\phi_{2,\theta}}{\partial x^2}(x,y) = \frac{-4}{y-x^2} - \frac{\log(y-x^2)}{x^2} - \frac{2\log x}{y-x^2} - \frac{4x^2\log(x)}{(y-x^2)^2} + \frac{\theta_1 - 2\theta_2}{x^2} + \frac{2\theta_2\log(x)}{x^2}$$

and $$\frac{\partial^2\phi_{2,\theta}}{\partial y^2}(x,y) = -\frac{\log(x)}{(y-x^2)^2}, \quad \frac{\partial^2\phi_{2,\theta}}{\partial x\partial y}(x,y) = \frac{1}{x(y-x^2)} + \frac{2x\log(x)}{(y-x^2)^2}.$$

To prove theorem 4.1(i), we first show that

$$\sqrt{nT}(U_T - V_T) = o_{\mathbb{P}}(1). \tag{7.9}$$

Using a Taylor–Lagrange expansion of order 2, we have

$$\sqrt{nT}(U_T - V_T) = \left(\frac{1}{2}\frac{\sqrt{n}}{\sqrt{T}}\sum_{t=1}^{T}(\hat{m}_t - m_t)'H_{1,t}(\hat{m}_t - m_t), \frac{1}{2}\frac{\sqrt{n}}{\sqrt{T}}\sum_{t=1}^{T}(\hat{m}_t - m_t)'H_{2,t}(\hat{m}_t - m_t)\right), \tag{7.10}$$

where

$$H_{1,t} = \nabla^{(2)}\phi_{1,\theta}(\widetilde{m}_t) \quad \text{and} \quad H_{2,t} = \nabla^{(2)}\phi_{2,\theta}(\check{m}_t),$$

for some random points $\widetilde{m}_t, \check{m}_t$ in the segment $[m_t, \hat{m}_t]$. It is only necessary to show equation (7.9) on the event

$$E(\epsilon) = \bigcap_{t=1}^{T}\{\hat{\mu}_t > \epsilon, \hat{\sigma}_t^2 > \epsilon\},$$

which has been defined in the proof of theorem 3.5 and which satisfies $\mathbb{P}(E(\epsilon)^c) \leq 2T\exp(-nL)$ for some $L > 0$. From lemma A.1(ii), on $E(\epsilon)$, we have, setting $w = \min(\epsilon, s)$,

$$\min_{1\leq t\leq T}\min(\mu_t, m_{2,t}, \sigma_t^2, \hat{\mu}_t, \hat{m}_{2,t}, \hat{\sigma}_t^2) \geq w.$$

Moreover, on the same event $E(\epsilon)$, using convexity of the square function, we have for $1 \leq t \leq T$,

$$\min(\widetilde{m}_{2,t} - \widetilde{m}_{1,t}^2, \check{m}_{2,t} - \check{m}_{1,t}^2) \geq \min(\hat{\sigma}_t^2, \sigma_t^2) \geq w.$$

Also $\min(\widetilde{m}_{1,t}, \check{m}_{1,t}) \geq w$. Moreover, there exists a positive constant $\widetilde{C}$ such that for $1 \leq t \leq T$ and $i = 1, 2$,

$$\max(\widetilde{m}_{i,t}, \check{m}_{i,t}) \leq \widetilde{C}(1 + |\hat{m}_{i,t} - m_{i,t}|).$$

It is then sufficient to show that for $i, i' \in \{1, 2\}$,

$$\frac{\sqrt{n}}{\sqrt{T}}\sum_{t=1}^{T} h_{i,i',t}|\hat{m}_{i,t} - m_{i,t}| \cdot |\hat{m}_{i',t} - m_{i',t}| = o_{\mathbb{P}}(1), \tag{7.11}$$

with

$$h_{i,i',t} = \max_{\ell=1,2}\sup_{\substack{w\leq x\leq\widetilde{C}(1+|\hat{\mu}_t-\mu_t|)\\ w\leq z\leq\widetilde{C}(1+|\hat{m}_{2,t}-m_{2,t}|)}}\left|\frac{\partial^2\phi_{\ell,\theta}}{\partial x^i\partial y^{i'}}(x, z + x^2)\right|.$$

A careful inspection of the second partial derivatives of $\phi_\theta$ shows that for any $\delta > 0$, there exists a constant $C'_\delta > 0$ such that

$$h_{1,1,t} \le C'_\delta[1 + |\hat{\mu}_t - \mu_t|^{2+\delta} + |\hat{m}_{2,t} - m_{2,t}|^\delta]$$

and

$$h_{2,2,t} \le C'_\delta[1 + |\hat{\mu}_t - \mu_t|^\delta] \quad \text{and} \quad h_{1,2,t} \le C'_\delta[1 + |\hat{\mu}_t - \mu_t|^{1+\delta}].$$

Take $\delta \in (0,1)$ such that $4 + \delta < k$. Using the moment bound given by lemma A.5, it is not difficult to get that

$$\frac{\sqrt{n}}{\sqrt{T}} \sum_{t=1}^{T} h_{i,i',t} |\hat{m}_{i,t} - m_{i,t}| \cdot |\hat{m}_{i',t} - m_{i',t}| = O_{\mathbb{P}}(\sqrt{T/n}),$$

which yields equation (7.11) and then equation (7.9).

It now remains to show that

$$W_T := \sqrt{nT} C_T V_T \Rightarrow \mathcal{N}_2(0, I_2). \tag{7.12}$$

We recall that $I_2$ denotes the $2 \times 2$ identity matrix.

This follows from the Lindeberg–Feller central limit theorem for triangular arrays [52, Theorem 18.1]. Since $\mathrm{Var}(W_T) = I_2$, we only have to check the Lindeberg condition. Setting $Z_{n,j,t} = (1/\sqrt{nT}) C_T J_{\phi_\theta}(m_t)(Y_{j,t} - m_t)$ and using our assumptions, we have $C_T = O(1)$ and $\max_{1 \le t \le T} |J_{\phi_\theta}(m_t)| = O(1)$. For $\delta > 0$ such that $2(2+\delta) \le k$, we get $\sum_{j=1}^{n} \sum_{t=1}^{T} \mathbb{E}|Z_{n,j,t}|^{2+\delta} = O((nT)^{-\delta/2}) = o(1)$. This proves equation (7.12).

To prove theorem 4.1(ii), we again use the decomposition equation (7.10) and we first show that for $\ell = 1, 2$,

$$\frac{\sqrt{n}}{\sqrt{T}} \sum_{t=1}^{T} (\hat{m}_t - m_t)'[H_{\ell,t} - K_{\ell,t}](\hat{m}_t - m_t) = o_{\mathbb{P}}(1), \tag{7.13}$$

where for $\ell = 1, 2$ and $t \ge 1$, $K_{\ell,t} = \nabla^{(2)} \phi_{\ell,\theta}(m_t)$. It is only necessary to show equation (7.13) on the event $E_n(\epsilon)$. To this end, we will use the bounds $\max\{|\widetilde{\mu}_t - \mu_t|, |\breve{\mu}_t - \mu_t|\} \le |\hat{\mu}_t - \mu_t|$, and

$$\max\{|\widetilde{\sigma}_t^2 - \sigma_t^2|, |\breve{\sigma}_t^2 - \sigma_t^2|\} \le |\hat{m}_{2,t} - m_{2,t}| + (2 \sup_{s \ge 1} \mu_s + |\hat{\mu}_t - \mu_t|) \cdot |\hat{\mu}_t - \mu_t|,$$

where we set $\widetilde{\sigma}_t^2 := \widetilde{m}_{2,t} - \widetilde{m}_{1,t}^2$ and $\breve{\sigma}_t^2 := \breve{m}_{2,t} - \breve{m}_{1,t}^2$. We will also use the fact that if $h : (\epsilon, \infty) \to \mathbb{R}_+$ is any mapping of the form $h(z) = 1/z$, $h(z) = 1/z^2$ or $h(z) = \log(z)$, then for any $\kappa \in (0,1)$, there exists $C_\kappa > 0$ such that $|h(z) - h(z')| \le C_\kappa |z - z'|^\kappa$.

Assertion equation (7.13) can be obtained from these facts and a careful analysis of the second derivatives of $\phi_\theta$. We illustrate this by showing that

$$\frac{\sqrt{n}}{\sqrt{T}} \sum_{t=1}^{T} (\hat{m}_{2,t} - m_{2,t})^2 \left[ \frac{\partial^2 \phi_{1,\theta}}{\partial y^2}(\widetilde{m}_t) - \frac{\partial^2 \phi_{1,\theta}}{\partial y^2}(m_t) \right] = o_{\mathbb{P}}(1). \tag{7.14}$$

For $\kappa \in (0,1)$ small enough to satisfy $4 + 2\kappa \le k$, there exists $D_\kappa > 0$ such that

$$\left| \frac{\widetilde{\mu}_t}{\widetilde{\sigma}_t^4} - \frac{\mu_t}{\sigma_t^4} \right| \le C[|\widetilde{\mu}_t - \mu_t| + |\widetilde{\sigma}_t^2 - \sigma_t^2|^\kappa].$$

To show equation (7.14), it is sufficient to show that

$$\mathbb{E}[(\hat{m}_{2,t} - m_{2,t})^2 |\hat{m}_{i,t} - m_{i,t}|^{\kappa_i}] = O\left(n^{-1-\frac{\kappa_i}{2}}\right),$$

when $(i, \kappa_i) = (2, \kappa), (1, \kappa)$ or $(1, 2\kappa)$. But this point easily follows from lemma A.5 and Hölder's inequality. Assertions similar to equation (7.14) can be obtained in the same way for the other

second-order partial derivatives of $\phi_\theta$. Details are omitted. Then

$$\sqrt{nT}(U_T - V_T) = \left( \frac{\sqrt{n}}{\sqrt{T}} \sum_{t=1}^{T} (\hat{m}_t - m_t)' K_{1,t} (\hat{m}_t - m_t), \frac{\sqrt{n}}{\sqrt{T}} \sum_{t=1}^{T} (\hat{m}_t - m_t)' K_{2,t} (\hat{m}_t - m_t) \right) + o_{\mathbb{P}}(1).$$

One applies lemma A.2 to conclude the proof.

## (d) Proof of corollary 4.7

The result is a consequence of theorem 4.1 and of lemmas A.8 and A.4 in appendix A.

Set $\hat{A}_n := \hat{C}_T \hat{D}_T$ and $A_n := C_T \hat{D}_T$. Using lemma A.8 in appendix A, we have $\hat{C}_T - C_T = o_{\mathbb{P}}(1)$ and from the proof of theorem 3.5, we know that $\hat{D}_T - D_T = o_{\mathbb{P}}(1)$ and both $D_T$ and $D_T^{-1}$ are $O(1)$. We then easily conclude that $\hat{A}_n - A_n = o_{\mathbb{P}}(1)$.

Let us also define $\hat{B}_n := \sqrt{T/n}\hat{C}_T \hat{E}_T$ if $T/n = O(1)$ and $\hat{B}_n := 0$ if $T/n = o(1)$. Once again, using lemma A.8, it is easy to show that $\hat{B}_n - B_n = o_{\mathbb{P}}(1)$ where $B_n := \sqrt{T/n} C_T E_T$ if $T/n = O(1)$ and $B_n := 0$ if $T/n = o(1)$. Setting $W_n = \sqrt{nT}(\hat{\theta} - \theta)$, we already know from theorem 4.1 that $A_n W_n - B_n \Rightarrow \mathcal{N}_2(0, I_2)$. Moreover, using the decomposition

$$Z_n := \hat{A}_n W_n - \hat{B}_n = A_n W_n - B_n + (\hat{A}_n - A_n) A_n^{-1} (A_n W_n - B_n) + (\hat{A}_n A_n^{-1} - I_2) B_n + B_n - \hat{B}_n,$$

we conclude from lemma A.8 that $Z_n \Rightarrow \mathcal{N}_2(0, I_2)$. The result is now a consequence of lemma A.4.

Data accessibility. The data used in the paper along with some numerical codes are provided on Zenodo [25]. The data set is taken from the ICES website and is also accessible at https://www.ices.dk/data/data-portals/Pages/Datras_news_and_updates.aspx.

The data are provided in electronic supplementary material [53].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. L.T.: writing—original draft; J.E.C.: writing—original draft; P.D.: writing—original draft.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This work was funded by CY-AS ('Investissements d'Avenir' ANR-16-IDEX-0008), 'EcoDep' PSI-AAP2020-0000000013.

Acknowledgements. J.E.C. is also affiliated with Department of Statistics and Climate School, Columbia University, New York, NY 10027; and Department of Statistics, University of Chicago, Chicago, IL 60637, USA. J.E.C. thanks Roseanne Benjamin for her help during this work.

# Appendix A

This appendix states and proves several technical lemmas used to prove our main results. Lemma A.1 shows that assumption A3 leads to a lower bound for the population mean and the population variance. It also shows that small values of the sample mean and sample variance occur with very small probability if $n$ is large, as noted at the beginning of the proofs in §7.

**Lemma A.1.** *Suppose that assumption A3 holds, i.e. that there exists some $\delta > 0$ such that*

$$p(\delta) := \inf_{t \geq 1} \mathbb{P}(X_{1,t} > \delta) > 0 \quad \text{and} \quad q(\delta) := \inf_{t \geq 1} \mathbb{P}(|X_{1,t} - X_{2,t}| > \delta) > 0.$$

*Then*

*(i) there exists $s > 0$ such that $\inf_{t \geq 1} \mu_t \geq s$ and $\inf_{t \geq 1} \sigma_t^2 \geq s$.*

*(ii) Let $0 < \epsilon < \min(\delta p(\delta), \delta^2 q(\delta)/4)$. Then there exists $L > 0$ such that for all $n \geq 2$,*

$$\sup_{t \geq 1} \max\{\mathbb{P}(\hat{\mu}_t \leq \epsilon), \mathbb{P}(\hat{\sigma}_t^2 \leq \epsilon)\} \leq \exp(-Ln).$$

*Proof of lemma A.1.* To prove point (i), we use the inequality $x^i \geq \delta^i \mathbb{1}_{\{x \geq \delta\}}$ for any $x \geq 0$ and $i = 1, 2$, which leads to the lower bounds

$$\mu_t \geq \delta \mathbb{P}(X_{1,t} > \delta) \geq \delta p(\delta)$$

and

$$\sigma_t^2 = \frac{1}{2}\mathbb{E}(|X_{1,t} - X_{2,t}|^2) \geq \frac{\delta^2}{2}\mathbb{P}(|X_{1,t} - X_{2,t}| \geq \delta) \geq \frac{\delta^2}{2} q(\delta).$$

To prove point (ii), we observe that $\hat{\mu}_t \geq (\delta/n)\sum_{j=1}^n \mathbb{1}_{X_{j,t} > \delta}$. Then, setting $Z_{j,t} := \mathbb{1}_{X_{j,t} > \delta} - \mathbb{P}(X_{j,t} > \delta)$,

$$\mathbb{P}(\hat{\mu}_t \leq \epsilon) \leq \mathbb{P}\left(\frac{1}{n}\sum_{j=1}^n Z_{j,t} \leq \frac{\epsilon}{\delta} - p(\delta)\right).$$

Using Hoeffding's exponential inequality [54, inequality (2.6)], we conclude that $\sup_{t \geq 1} \mathbb{P}(\hat{\mu}_t \leq \epsilon) \leq \exp(-nL')$ for some $L' > 0$. $L'$ can be taken here as $(\epsilon/\delta - p(\delta))^2$. To control the probability of small values of the empirical variance, we use an immediate consequence of Lagrange's identity:

$$\hat{\sigma}_t^2 = \frac{1}{2n^2}\sum_{1 \leq i \neq j \leq n} (X_{i,t} - X_{j,t})^2.$$

This leads to the lower bound $\hat{\sigma}_t^2 \geq (\delta^2/2n^2)\sum_{1 \leq i \neq j \leq n} \mathbb{1}_{|X_{i,t} - X_{j,t}| > \delta}$. Next, set $Z_{i,j}^{(t)} := \mathbb{1}_{|X_{i,t} - X_{j,t}| > \delta} - \mathbb{P}(|X_{i,t} - X_{j,t}| > \delta)$. Then

$$\mathbb{P}(\hat{\sigma}_t^2 \leq \epsilon) \leq \mathbb{P}\left(\frac{1}{n(n-1)}\sum_{1 \leq i \neq j \leq n} Z_{i,j}^{(t)} \leq \frac{2n^2\epsilon}{\delta^2 n(n-1)} - q(\delta)\right).$$

Since $2n^2/(n(n-1)) \leq 4$ for every $n \geq 2$, an application of Hoeffding's inequality for $U$-statistics [54, bound (5.7)] leads to

$$\mathbb{P}(\hat{\sigma}_t^2 \leq \epsilon) \leq \exp\left(-2\left[\frac{n}{2}\right]h^2\right),$$

where $[n/2]$ denotes the integer-part of $n/2$ and $h = q(\delta) - (4\epsilon/\delta^2)$. It is then always possible to get the conclusion of the lemma for a suitably chosen constant $L \in (0, L')$. ∎

**Lemma A.2.** *Suppose that assumption A1 holds true and for any $j \geq 1$, the variables $X_{j,1}, \ldots, X_{j,T}$ are mutually independent. Let $\{J_t : t \in \mathbb{N}\}$ be a family of $2 \times 2$ matrices such that $\sup_{t \geq 1} |J_t| < \infty$. Suppose that $\sup_{t \geq 1} \mathbb{E}X_{1,t}^{4(1+\kappa)} < \infty$ for some $\kappa \in (0, 1)$. Then*

$$\frac{1}{T}\sum_{t=1}^T (\hat{m}_t - m_t)' J_t (\hat{m}_t - m_t) = \frac{1}{nT}\sum_{t=1}^T Tr(J_t Var(Y_{1,t})) + O_{\mathbb{P}}\left(\left(\frac{T}{n}\right)^{(1/2)-(\kappa/(1+\kappa))} \frac{1}{n^{2\kappa/(1+\kappa)}\sqrt{nT}} + \frac{1}{\sqrt{n}\sqrt{nT}}\right),$$

*where $Y_{j,t} = (X_{j,t}, X_{j,t}^2)'$ for $1 \leq j \leq n$ and $1 \leq t \leq T$.*

*Proof of lemma A.2.* Setting $\overline{Y}_{j,t} := Y_{j,t} - \mathbb{E}Y_{j,t}$ and

$$A := \frac{1}{n^2 T}\sum_{t=1}^T \sum_{j=1}^n \overline{Y}_{j,t}' J_t \overline{Y}_{j,t} \quad \text{and} \quad B := \frac{2}{n^2 T}\sum_{t=1}^T \sum_{1 \leq i < j \leq n} \overline{Y}_{i,t}' J_t \overline{Y}_{j,t},$$

we have $(1/T)\sum_{t=1}^T (\hat{m}_t - m_t)' J_t (\hat{m}_t - m_t) = A + B$. It is straightforward to get $\mathbb{E}(A) = (1/nT)\sum_{t=1}^T \mathrm{Tr}(J_t \mathrm{Var}(Y_{1,t}))$. Moreover, setting $p = 1 + \kappa$, one can apply lemma A.5 to get

$$||A - \mathbb{E}(A)||_p = O(n^{-1-(\kappa/(1+\kappa))} T^{-\kappa/(1+\kappa)}).$$

Then

$$\mathbb{E}(B^2) = \frac{4}{n^4T^2}\sum_{t=1}^{T}\sum_{1\le i<j\le n}\mathbb{E}(|\overline{Y}'_{i,t}J_t\overline{Y}_{j,t}|^2) \le \frac{2n(n-1)}{n^4T}\sup_{t\ge1}|J_t| \times \sup_{t\ge1}\mathbb{E}^2(|\overline{Y}_{1,t}|^2).$$

Hence $B = O_{\mathbb{P}}(1/(\sqrt{n}\sqrt{nT}))$. ■

**Lemma A.3.** *Assume that $(Z_n)_{n\in\mathbb{N}}$ is a sequence of random vectors of $\mathbb{R}^2$ such that $Z_n \Rightarrow \mathcal{N}(0, I_2)$. Then, for any sequence $(u_n)_{n\in\mathbb{N}}$ of deterministic vectors of $\mathbb{R}^2$ such that $|u_n| = 1$, we have*

$$u'_nZ_n \Rightarrow \mathcal{N}(0,1).$$

*Proof of lemma A.3.* It is clear that the sequence $(u'_nZ_n)_{n\in\mathbb{N}}$ is bounded in probability (i.e. tight). We simply have to show that any subsequence has a subsubsequence converging to the distribution $\mathcal{N}(0,1)$. Let $(u'_{\phi_n}Z_{\phi_n})_{n\in\mathbb{N}}$ be a given subsequence. Since the unit sphere is compact, one can find a subsubsequence $(u_{\psi_n})_{n\in\mathbb{N}}$ with $\psi_n \in \{\phi_k : k \in \mathbb{N}\}$ for all $n \in \mathbb{N}$ and such that $\lim_{n\to\infty} u_{\psi_n} = u$ for some $u \in \mathbb{R}^2$ such that $|u| = 1$. Since

$$u'_{\psi_n}Z_{\psi_n} = (u_{\psi_n} - u)'Z_{\psi_n} + u'Z_{\psi_n},$$

our assumption guarantees that $u'Z_{\psi_n} \Rightarrow \mathcal{N}(0,1)$ and we also know that $(u_{\psi_n} - u)'Z_{\psi_n}$ will converge in probability to 0 as $n \to \infty$. From Slutsky's lemma, we conclude that $u'_{\psi_n}Z_{\psi_n} \Rightarrow \mathcal{N}(0,1)$. ■

**Lemma A.4.** *Suppose $Z_n = \hat{A}_nW_n - \hat{B}_n \Rightarrow \mathcal{N}_2(0, I_2)$ with $\hat{A}_n$ a random real matrix of size $2\times 2$, $W_n$ a random real vector of $\mathbb{R}^2$ and $\hat{B}_n$ a random real vector of $\mathbb{R}^2$. Suppose furthermore that $\hat{A}_n - A_n = o_P(1)$, with $A_n$ deterministic and such that $A_n = O(1)$ and $A_n^{-1} = O(1)$. Setting $H_n := \hat{A}_n^{-1}(\hat{A}_n^{-1})'$ and $m_n = \hat{A}_n^{-1}\hat{B}_n$, we have for $i = 1, 2$,*

$$\frac{W_n(i) - m_n(i)}{\sqrt{H_n(i,i)}} \Rightarrow \mathcal{N}(0,1).$$

*Proof of lemma A.4.* To show first that

$$\hat{A}_n^{-1} - A_n^{-1} = o_{\mathbb{P}}(1), \tag{A 1}$$

we write

$$\hat{A}_n^{-1} = (A_n - (A_n - \hat{A}_n))^{-1} = \sum_{k\ge0}(A_n^{-1}(A_n - \hat{A}_n))^kA_n^{-1},$$

which is bounded in probability because $\hat{A}_n - A_n \to 0$ in probability. We then deduce that

$$\hat{A}_n^{-1} - A_n^{-1} = \hat{A}_n^{-1}(A_n - \hat{A}_n)A_n^{-1} = o_{\mathbb{P}}(1),$$

which proves equation (A 1). Next, for $i = 1, 2$, we set $\hat{v}_{n,i} := (\hat{A}_n^{-1})'e_i$, where $e_1 := (1,0)'$, $e_2 := (0,1)'$. Define also $v_{n,i} := (A_n^{-1})'e_i$, $i = 1, 2$. Since $e_i = A'_nv_{n,i}$ and $A_n = O(1)$, $v_{n,i}$ cannot approximate 0. Moreover, $\hat{v}_{n,i} - v_{n,i} = o_{\mathbb{P}}(1)$. Using also that $A_n^{-1} = O(1)$, $v_{n,i}$ is bounded and we deduce that

$$\frac{\hat{v}_{n,i}}{|\hat{v}_{n,i}|} - \frac{v_{n,i}}{|v_{n,i}|} = o_{\mathbb{P}}(1).$$

We then conclude that $(\hat{v}'_{n,i}Z_n/|\hat{v}_{n,i}|) - (v'_{n,i}Z_n/|v_{n,i}|) = o_{\mathbb{P}}(1)$. Using lemma A.3, we know that $v_{n,i}{}'Z_n/|v_{n,i}| \Rightarrow \mathcal{N}(0,1)$. Since $\hat{v}'_{n,i}Z_n/|\hat{v}_{n,i}| = (W_n(i) - m_n(i))/|\hat{v}_{n,i}|$, we only have to show that $|v_{n,i}| = \sqrt{H_n(i,i)}$ to finish the proof. ■

The following lemma is useful to bound moments of partial sums. It is a very special case of the Burkholder inequality for martingales [55, Theorem 2.10].

**Lemma A.5.** *If $\xi_1, \ldots, \xi_n$ are independent random variables with 0 mean and a finite moment of order $p > 1$, then there exists a constant $C_p > 0$ such that*

$$\left\|\sum_{i=1}^{n}\xi_i\right\|_p \le C_p \times \begin{cases} n^{1/p}\max_{1\le i\le n}||\xi_i||_p, & \text{if } p \in (1,2), \\ n^{1/2}\max_{1\le i\le n}||\xi_i||^{1/2}_{p/2}, & \text{if } p \ge 2. \end{cases}$$

**Corollary A.6.** *Suppose that assumption A1 is fulfilled and that for some $k > 4$, $\sup_{t\geq 1} \mathbb{E}|X_{1,t}|^k < \infty$. Define for $i = 1, 2, 3, 4$, $S_{i,t} := (1/n)\sum_{j=1}^{n}(X_{j,t}^i - \mathbb{E}[X_{j,t}^i])$. Suppose that for some integer $d \geq 1$ and $i = 1, \ldots, 4$, $\kappa_i$ is a positive real number, $\alpha_i \in \{1, \ldots, 4\}$ with $\sum_{i=1}^{d} \alpha_i \kappa_i \leq k$.*

*Then*

$$\sup_{t\geq 1} \mathbb{E}\left[\prod_{i=1}^{d} |S_{\alpha_i,t}|^{\kappa_i}\right] = O\left(n^{-\sum_{i=1}^{d} \kappa_i \min\{1/2, 1-\alpha_i/k\}}\right).$$

*Proof of corollary A.6.* Setting $p_i := k/(\alpha_i\kappa_i)$ for $1 \leq i \leq d$, we have $\sum_{i=1}^{d} 1/p_i \leq 1$ and applying the Hölder inequality, we get

$$\mathbb{E}\left[\prod_{i=1}^{d} |S_{\alpha_i,t}|^{\kappa_i}\right] \leq \prod_{i=1}^{d} \mathbb{E}^{1/p_i}[|S_{\alpha_i,t}|^{k/\alpha_i}].$$

The result then follows from lemma A.5. ■

**Lemma A.7.** *Given the assumptions of theorem 4.1, let $\Gamma_T$ be the symmetric, non-negative definite matrix*

$$\Gamma_T := \frac{1}{T}\sum_{t=1}^{T} J_{\phi_\theta}(m_t) Var(Y_{1,t}) J_{\phi_\theta}(m_t)',$$

*where $J_{\phi_\theta}(m_t)$ denotes the Jacobian matrix of $\phi_\theta$ at point $m_t$ and $m_t := \mathbb{E}Y_{1,t}$ and $Y_{j,t} := (X_{j,t}, X_{j,t}^2)'$. Let $\lambda_T$ be the smallest eigenvalue of $\Gamma_T$ and $A_T = (1/T)\sum_{t=1}^{T} J_{\phi_\theta}(m_t)$.*

*(i) There exists $\xi > 0$ such that $\lambda_T \geq \xi \min_{1\leq t\leq T} \det[Var(Y_{1,t})]\det^2(A_T)$.*

*(ii) Suppose that $X_{1,t}$ follows a Poisson distribution with parameter $\mu_t = \mu(t/T)$ with $\mu : [0, 1] \to (0, \infty)$ a continuous function. Then there exists $\zeta > 0$ such that for $T$ large enough, $\lambda_T \geq \zeta$.*

*Proof of lemma A.7.*

(i) Let $\gamma_t$ be the smallest eigenvalue of $\mathrm{Var}(Y_{1,t})$ and $\beta_T$ the smallest eigenvalue of the matrix $A_T A_T'$. If $x \in \mathbb{R}^2$, then

$$\begin{aligned} x'\Gamma_T x &\geq \frac{1}{T}\sum_{t=1}^{T} x' J_{\phi_\theta}(m_t)\mathrm{Var}(Y_{1,t}) J_{\phi_\theta}(m_t)' x \\ &\geq \frac{1}{T}\sum_{t=1}^{T} \gamma_t |J_{\phi_\theta}(m_t)' x|^2 \geq \min_{1\leq t\leq T} \gamma_t (x' A_T)^2 \geq \min_{1\leq t\leq T} \gamma_t \beta_T \\ &\geq \min_{1\leq t\leq T} \frac{\det \mathrm{Var}(Y_{1,t})}{\mathrm{Tr}\mathrm{Var}(Y_{1,t})} \cdot \frac{\det^2 A_T}{\sum_{1\leq i,j\leq 2} |A_T(i,j)|^2}. \end{aligned}$$

The previous three lines assert five inequalities. The second and fourth inequalities are based on the Courant–Fischer theorem. The third one is based on Jensen's inequality. The last inequality is a consequence of the inequality $\lambda \geq \det(B)/\mathrm{Tr}(B)$ for any (non-null) non-negative definite matrix $B$ with an eigenvalue $\lambda$. The result announced follows from the Courant–Fischer theorem and the fact that the assumptions of theorem 4.1 guarantee that the coefficients of the matrices $A_T$ and $\mathrm{Var}(Y_{1,t})$ are bounded with respect to $T$ and $t$, respectively.

(ii) Since $\mu_t = \sigma_t^2$ for the Poisson distribution, using the expressions of the first four moments of the Poisson distribution, we get $\det \mathrm{Var}(Y_{1,t}) = 2\mu_t^3$. The assumption on the function $\mu$

entails that $\inf_{t\geq 1}\mu_t > 0$. Moreover,

$$A_T = \begin{pmatrix} -2 - \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\mu_t} & \frac{1}{T}\sum_{t=1}^{T}\frac{1}{\mu_t} \\ -\frac{1}{T}\sum_{t=1}^{T}\left(2+\frac{1}{\mu_t}\right)\log\mu_t & \frac{1}{T}\sum_{t=1}^{T}\frac{\log\mu_t}{\mu_t} \end{pmatrix}.$$

Then

$$\lim_{T\to\infty}\det(A_T) = 2\int_0^1 \frac{1}{\mu(u)}\,\mathrm{d}u \cdot \int_0^1 \log\mu(u)\,\mathrm{d}u - 2\int_0^1 \frac{\log\mu(u)}{\mu(u)}\,\mathrm{d}u.$$

One can show that the previous limit is positive, since from Jensen's inequality applied first to the probability measure $P(\mathrm{d}u) = \mathbb{1}_{[0,1]}(u)\,\mathrm{d}u/(\mu(u)\int_0^1 \mu(v)^{-1}\,\mathrm{d}v)$ and then to the uniform distribution over $[0,1]$,

$$\left(\int_0^1 \frac{1}{\mu(u)}\,\mathrm{d}u\right)^{-1}\int_0^1 \frac{\log\mu(u)}{\mu(u)}\,\mathrm{d}u < -\log\int_0^1 \frac{1}{\mu(u)}\,\mathrm{d}u < \int_0^1 \log\mu(u)\,\mathrm{d}u.$$

This proves the lemma. ∎

**Lemma A.8.** *Suppose that all the assumptions of theorem 4.1 are valid, with or without the additional condition on $T/n$ given in point (i) or point (ii) of theorem 4.1. Set $\Sigma_t := Var(Y_{1,t})$ and*

$$\hat{\Sigma}_t := \frac{1}{n}\sum_{j=1}^{n} Y_{j,t}Y_{j,t}' - \overline{Y}_{n,t}\overline{Y}_{n,t}' \quad \text{and} \quad \overline{Y}_{n,t} := \frac{1}{n}\sum_{j=1}^{n} Y_{j,t}, \quad \hat{\Gamma}_T := \frac{1}{T}\sum_{t=1}^{T} J_{\phi_{\hat{\theta}}}(\hat{m}_t)\hat{\Sigma}_t J_{\phi_{\hat{\theta}}}(\hat{m}_t)',$$

*where $J_{\phi_\theta}(m_t)$ denotes the Jacobian matrix of $\phi_\theta$ at point $m_t$ and $m_t := \mathbb{E}Y_{1,t}$ and $Y_{j,t} := (X_{j,t}, X_{j,t}^2)'$. Set*

$$\hat{G}_t := (Tr(\nabla^{(2)}\phi_{1,\hat{\theta}}(\hat{m}_t)\hat{\Sigma}_t), Tr(\nabla^{(2)}\phi_{2,\hat{\theta}}(\hat{m}_t)\hat{\Sigma}_t)).$$

*Then, with $\Gamma_T$ defined in the statement of lemma A.7,*

$$\hat{\Gamma}_T - \Gamma_T = o_{\mathbb{P}}(1). \tag{A2}$$

*Moreover, if $C_T := \Gamma_T^{-1/2}$ and $\hat{C}_T := \hat{\Gamma}_T^{-1/2}$, we have*

$$\hat{C}_T - C_T = o_{\mathbb{P}}(1) \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T}(\hat{G}_t - G_t) = o_{\mathbb{P}}(1). \tag{A3}$$

The proof of lemma A.8 is given in the electronic supplementary material.

## References


1. Taylor LR. 1961 Aggregation, variance and the mean. *Nature* **189**, 732–735. (doi:10.1038/189732a0)
2. Bliss CI. 1941 Statistical problems in estimating populations of Japanese beetle larvae. *J. Econ. Entomol.* **34**, 221–232. (doi:10.1093/jee/34.2.221)
3. Taylor RAJ. 2019 *Taylor's power law: order and pattern in nature*. London, UK: Academic Press.
4. Cohen JE, Saitoh T. 2016 Population dynamics, synchrony, and environmental quality of Hokkaido voles lead to temporal and spatial Taylor's laws. *Ecology* **97**, 3402–3413. (doi:10.1002/ecy.1575)
5. Taylor LR, Woiwod IP, Perry JN. 1978 The density-dependence of spatial behaviour and the rarity of randomness. *J. Anim. Ecol.* **47**, 383–406. (doi:10.2307/3790)
6. Xu M, Cohen JE. 2021 Spatial and temporal autocorrelations affect Taylor's law for US county populations: descriptive and predictive models. *PLoS ONE* **16**, e0245062. (doi:10.1371/journal.pone.0245062)

7. Yang Y, Shang HL, Cohen JE. 2022 Temporal and spatial Taylor's law: application to Japanese subnational mortality rates. *J. R. Stat. Soc. Ser. A* **185**, 1979–2006. (doi:10.1111/rssa.12859)
8. Bartlett MS. 1936 Some notes on insecticide tests in the laboratory and in the field. *Suppl. J. R. Stat. Soc.* **3**, 185–194. (doi:10.2307/2983670)
9. Cohen JE, Poulin R, Lagrue C. 2017 Linking Parasite populations in hosts to parasite populations in space through Taylor's law and the negative binomial distribution. *Proc. Natl Acad. Sci. USA* **114**, E47–E56. (doi:10.1073/pnas.1618803114)
10. Tweedie MCK. 1984 An index which distinguishes between some important exponential families. In *Statistics: Applications and New Directions: Proc. Indian Statistical Institute Golden Jubilee International Conference* (eds JK Ghosh, J Roy), pp. 579–604. Calcutta, India: Indian Statistical Institute.
11. Bar-Lev SK, Enis P. 1986 Reproducibility and natural exponential families with power variance functions. *Ann. Stat.* **14**, 1507–1522. (doi:10.1214/aos/1176350173)
12. Bar-Lev SK, Stramer O. 1987 Characterizations of natural exponential families with power variance functions by zero regression properties. *Probab. Theory Relat. Fields* **76**, 509–522. (doi:10.1007/BF00960071)
13. Bar-Lev S. 2020 Independent, tough identical results: the class of Tweedie on power variance functions and the class of Bar-Lev and Enis on reproducible natural exponential families. *Int. J. Stat. Probab.* **9**, 30–35. (doi:10.5539/ijsp.v9n1p30)
14. Cohen JE, Huillet TE. 2022 Taylor's law for some infinitely divisible probability distributions from population models. *J. Stat. Phys.* **188**, 33. (doi:10.1007/s10955-022-02962-y)
15. Clark SJ, Perry JN. 1994 Small sample estimation for Taylor's power law. *Environ. Ecol. Stat.* **1**, 287–302. (doi:10.1007/BF00469426)
16. James A, Plank MJ. 2007 On fitting power laws to ecological data. (https://arxiv.org/abs/0712.0613)
17. Jørgensen B, Demétrio CG, Kristensen E, Banta GT, Petersen HC, Delefosse M. 2011 Bias-corrected Pearson estimating functions for Taylor's power law applied to benthic macrofauna data. *Stat. Probab. Lett.* **81**, 749–758.
18. Perry JN, Woiwod IP. 1992 Fitting Taylor's power law. *Oikos* **65**, 538–542. (doi:10.2307/3545573)
19. Sakoda G, Takayasu H, Takayasu M. 2019 Tracking Poisson parameter for non-stationary discontinuous time series with Taylor's abnormal fluctuation scaling. *Stats* **2**, 55–69. (doi:10.3390/stats2010005)
20. Shi P, Ratkowsky DA, Zhao L, Reddy GVP, Li BL. 2017 Comparison of five methods for parameter estimation under Taylor's power law. *Ecol. Complex.* **32**, 121–130. (doi:10.1016/j.ecocom.2017.10.006)
21. Brown M, Cohen JE, de la Peña VH. 2017 Taylor's law, via ratios, for some distributions with infinite mean. *J. Appl. Probab.* **54**, 657–669. (doi:10.1017/jpr.2017.25)
22. Brown M, Cohen JE, Tang CF, Yam SCP. 2021 Taylor's law of fluctuation scaling for semivariances and higher moments of heavy-tailed data. *Proc. Natl Acad. Sci. USA* **118**, e2108031118. (doi:10.1073/pnas.2108031118)
23. Cohen JE, Davis RA, Samorodnitsky G. 2020 Heavy-tailed distributions, correlations, kurtosis and Taylor's law of fluctuation scaling. *Proc. R. Soc. A* **476**, 20200610. (doi:10.1098/rspa.2020.0610)
24. Cohen JE, Davis RA, Samorodnitsky G. 2022 COVID-19 cases and deaths in the United States follow Taylor's law for heavy-tailed distributions with infinite variance. *Proc. Natl Acad. Sci. USA* **119**, e2209234119. (doi:10.1073/pnas.2209234119)
25. Truquet L. 2025 Fisheries data and R code. Zenodo. (doi:10.5281/zenodo.16947399)
26. Cohen JE. 2014 Stochastic population dynamics in a Markovian environment implies Taylor's power law of fluctuation scaling. *Theor. Popul. Biol.* **93**, 30–37. (doi:10.1016/j.tpb.2014.01.001)
27. Cohen JE. 2014 Taylor's law and abrupt biotic change in a smoothly changing environment. *Theor. Ecol.* **7**, 77–86. (doi:10.1007/s12080-013-0199-z)
28. Fujiwara M, Cohen JE. 2015 Mean and variance of population density and temporal Taylor's law in stochastic stage-structured density-dependent models of exploited fish populations. *Theor. Ecol.* **8**, 175–186. (doi:10.1007/s12080-014-0242-8)
29. Doukhan P. 1995 Mixing. In *Mixing: properties and examples*, pp. 15–23. New York, NY: Springer.

30. Truquet L. 2010 A moment inequality of the Marcinkiewicz–Zygmund type for some weakly dependent random fields. *Stat. Probab. Lett.* **80**, 1673–1679. (doi:10.1016/j.spl.2010.07.011)
31. Guyon X. 1995 *Random fields on a network: modeling, statistics, and applications*. London, UK: Springer Science & Business Media.
32. Doukhan P, Jakubowicz J, León JR. 2010 Variance estimation with applications. In *Dependence in probability, analysis and number theory* (eds I Berkes, RC Bradley, H. Dehling, M. Peligrad, R. Tichy), pp. 203–231. Heber City, UT: Kendrick Press.
33. Prause A, Steland A. 2018 Estimation of the asymptotic variance of univariate and multivariate random fields and statistical inference. *Electron. J. Stat.* **12**, 890–940. (doi:10.1214/18-EJS1398)
34. Li H, Lindsay BG, Waterman RP. 2003 Efficiency of projected score methods in rectangular array asymptotics. *J. R. Statist. Soc. Ser. B* **65**, 191–208. (doi:10.1111/1467-9868.00380)
35. Cobain MRD, Brede M, Trueman CN. 2019 Taylor's power law captures the effects of environmental variability on community structure: an example from fishes in the North Sea. *J. Anim. Ecol.* **88**, 290–301. (doi:10.1111/1365-2656.12923)
36. Anderson R, Gordon D, Crawley M, Hassell M. 1982 Variability in the abundance of animal and plant species. *Nature* **296**, 245–248. (doi:10.1038/296245a0)
37. Zhao L, Sheppard LW, Reid PC, Walter JA, Reuman DC. 2019 Proximate determinants of Taylor's law slopes. *J. Anim. Ecol.* **88**, 484–494. (doi:10.1111/1365-2656.12931)
38. Cohen JE, 2020 Every variance function, including Taylor's power law of fluctuation scaling, can be produced by any location-scale family of distributions with positive mean and variance. *Theor. Ecol.* **13**, 1–5. Corrections: Cohen 2022 *Theoretical Ecology* **15**, 93–94. doi:10.1007/s12080-019-00445-7; and Mallmin, Emil 2022 Comments on: 'Every variance function...can be produced by any location-scale family...' *Theoretical Ecology* **15**, 283–284. https://doi.org/10.1007/s12080-022-00547-9. (doi:10.1007/s12080-019-00445-7)
39. Saitoh T. 2020 Effects of environmental synchrony and density-dependent dispersal on temporal and spatial slopes of Taylor's law. *Popul. Ecol.* **62**, 300–316. (doi:10.1002/1438-390X.12051)
40. Sileshi G, Hailu G, Nyadzi GI. 2009 Traditional occupancy–abundance models are inadequate for zero-inflated ecological count data. *Ecol. Modell.* **220**, 1764–1775. (doi:10.1016/j.ecolmodel.2009.03.024)
41. Lecomte JB, Benoît HP, Ancelet S, Etienne MP, Bel L, Parent E. 2013 Compound Poisson-gamma vs. delta-gamma to handle zero-inflated continuous data under a variable sampling volume. *Methods Ecol. Evol.* **4**, 1159–1166. (doi:10.1111/2041-210X.12122)
42. de la Peña V, Doukhan P, Salhi Y. 2022 A dynamic Taylor's law. *J. Appl. Probab.* **59**, 584–607. (doi:10.1017/jpr.2021.40)
43. Van der Vaart AW. 2000 *Asymptotic statistics*, vol. 3. Cambridge, UK: Cambridge University Press.
44. Dedecker J, Doukhan P, Lang G, Leon JR, Louhichi S, Prieur C. 2007 *Weak dependence*. New York, NY: Springer.
45. de la Peña V, Doukhan P, Salhi Y. 2022 Taylor's law of fluctuation scaling for semivariances and higher moments of heavy-tailed data. *J. Appl. Probab.* **59**, 584–607. (doi:10.1017/jpr.2021.40)
46. Doukhan P, Oppenheim G, Taqqu M. 2002 *Theory and applications of long-range dependence*. Boston, MA: Springer Science and Business Media.
47. Samorodnitsky G. 2007 Long range dependence. *Found. Trends Stoch. Syst.* **1**, E163–E257. (doi:10.1561/0900000004)
48. Abril F, Quimbay C. 2024 Evolution of temporal fluctuation scaling exponent in nonstationary time series using supersymmetric theory of stochastic dynamics. *Phys. Rev. E* **109**, 024112. (doi:10.1103/PhysRevE.109.024112)
49. Eisler Z, Kertesz J, Yook SH, Barabási AL. 2005 Multiscaling and non-universality in fluctuations of driven complex systems. *Europhys. Lett.* **69**, 664. (doi:10.1209/epl/i2004-10384-1)
50. Truquet L. 2019 Local stationarity and time-inhomogeneous Markov chains. *Ann. Stat.* **47**, 2023–2050. (doi:10.1214/18-AOS1739)

51. Andrews DW. 1993 Tests for parameter instability and structural change with unknown change point. *Econometrica* **61**, 821–856. (doi:10.2307/2951764)
52. Billingsley P. 2013 *Convergence of probability measures*. New York, NY: John Wiley & Sons.
53. Truquet L, Cohen JE, Doukhan P. 2025 Inferring the parameters of Taylor's power law of fluctuation scaling. Figshare. (doi:10.6084/m9.figshare.c.8115997)
54. Hoeffding W. 1963 Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58**, 13–30. (doi:10.1080/01621459.1963.10500830)
55. Hall P, Heyde CC. 2014 *Martingale limit theory and its application*. New York, NY: Academic Press.