

On aggregation of strongly dependent time series

Jan Beran¹  | Haiyan Liu^{1,2}  | Sucharita Ghosh³

¹Department of Mathematics Statistics, University of Konstanz

²Department of Statistics, University of Leeds

³Statistics Lab, Swiss Federal Research Institute WSL

Correspondence

Jan Beran, Department of Mathematics Statistics, University of Konstanz, Universitätsstrae 10, 78464 Konstanz, Germany.
Email: jan.beran@uni-konstanz.de

Abstract

We consider cross-sectional aggregation of time series with long-range dependence. This question arises for instance from the statistical analysis of networks where aggregation is defined via routing matrices. Asymptotically, aggregation turns out to increase dependence substantially, transforming a hyperbolic decay of autocorrelations to a slowly varying rate. This effect has direct consequences for statistical inference. For instance, unusually slow rates of convergence for nonparametric trend estimators and nonstandard formulas for optimal bandwidths are obtained. The situation changes, when time-dependent aggregation is applied. Suitably chosen time-dependent aggregation schemes can preserve a hyperbolic rate or even eliminate autocorrelations completely.

KEYWORDS

aggregation, kernel smoothing, long-range dependence, network, time series

1 | INTRODUCTION

Observed time series are often generated by cross-sectional aggregation. Granger (1980) showed that aggregation of short-memory processes can lead to long-range dependence. Generalizations of this result can be found for instance in Gonçalves and Gouriéroux (1988), Ding and Granger (1996), Lippi and Zaffaroni (1998), Igloi and Terdik (1999), Ghosh (2001), Abadir and Talmain (2002), Leipus and Viano (2002), Leipus, Oppenheim, Philippe, and Viano (2006),

Oppenheim and Viano (2004), Zaffaroni (2004, 2007a, 2007b), Kazakevicius, Leipus, and Viano (2004), Davidson and Sibbertsen (2005), Chong (2006), Leipus et al. (2006), Beran, Schützner, and Ghosh (2010), Giraitis, Leipus, and Surgailis (2010), and Pilipauskaite and Surgailis (2014, 2015) (also see, chapter 2.2 in Beran, Feng, Ghosh, & Kulik, 2013). In this paper, we consider instead aggregation of long-memory processes. It is shown that aggregation transforms hyperbolic rates of autocorrelations into slowly decaying rates. This leads to increased uncertainty in statistical inference. The effect can be alleviated using time-dependent aggregation coefficients.

The problem of cross-sectional aggregation of strongly dependent time series is motivated in particular by statistical questions arising in the analysis of computer networks. For example, in the traffic matrix estimation problem, observed data consist of “link flow” time series which are cross-sectional aggregates of a large number of unobservable “origin-destination flows” (OD-flows; e.g., Airolidi & Blocker, 2013; Cao, Davis, Vander Wiel, & Yu, 2000; Castro, Coates, Liang, Nowak, & Yu, 2004; Fang, Vardi, & Zhang, 2007; Nucci & Papagianaki, 2009; Ortuzar & Willumsen, 2011; Tebaldi & West, 1998; Vardi, 1996; Zhang, Roughan, Lund, & Donoho, 2003, and references therein). More specifically, suppose that there are n unobservable processes $(X_{ij})_{j \in \mathbb{N}}$ ($i = 1, \dots, n$). The observed time series at a specific link is then of the form

$$Y_j(n) = \sum_{i=1}^n a_i(j; n) X_{ij}(j = 1, \dots, N), \quad (1)$$

where the coefficients $a_i(j; n)$ are defined by a known routing matrix. Typically, $a_i(j; n) = 1$ if OD-flow number i passes through the link at time j , and $a_i(j; n) = 0$ otherwise.

In this paper, we consider Equation (1) under the assumption that the processes $(X_{ij})_{j \in \mathbb{N}}$ ($i = 1, \dots, n$) are independent of each other and exhibit long-range dependence characterized by randomly generated long-memory parameters $d_i \in (0, \frac{1}{2})$ ($i = 1, \dots, n$). Letting n tends to infinity, we consider autocorrelations and a functional limit theorem for the aggregated series. For coefficients that do not depend on time, the autocorrelations of $Y_j(n)$ turn out to have a rate of decay, that is, slower than hyperbolic. This implies a high degree of uncertainty for statistical inference. For instance, one obtains unusually slow convergence rates for non-parametric trend estimators and nonstandard formulas for optimal bandwidths. The situation changes, when time-dependent coefficients are used. Suitable sequences of random coefficients can remove the adverse effect of aggregation. From the point of view of statistical inference it may thus be recommended that, in systems where the coefficients can be controlled (e.g., Airolidi & Blocker, 2013; Cao et al., 2000), randomized time-dependent designs should be used.

The paper is organized as follows. Asymptotic formulas for autocovariances of aggregated series and a functional limit theorem is derived in Section 2 under the assumption of constant coefficients. In Section 3, these results are used to derive the asymptotic mean squared error and asymptotically optimal bandwidth for kernel estimators of trend functions. The case of time-dependent coefficients is considered in Section 4. In particular, it is shown that a suitable choice of $a_i(j; n)$ ($j \in \mathbb{N}$) can alleviate the effect of aggregation and even remove autocorrelations completely. Simulate examples in Section 5 illustrate the asymptotic results. A data example is discussed in Section 6.

2 | ASYMPTOTIC EFFECT OF AGGREGATION ON AUTOCOVARIANCES

In this section, we address the question which autocovariance structure is obtained for the aggregated process $Y_j(n)$ in Equation (1) when n tends to infinity. Specifically, the following conditions will be used:

- (A1) Let $p \geq 1$ and $\theta = (\theta_1, \dots, \theta_p)$ with $\theta_1 = d$. If $p = 1$, then $\Theta = (0, \frac{1}{2})$. For $p \geq 2$, $\Theta = \Theta_1 \times \Theta_2$ with $\Theta_1 = (0, \frac{1}{2})$, $\Theta_2 \subseteq \mathbb{R}^{p-1}$. Moreover,

$$\mathcal{F}_p = \{\gamma_\theta \mid \gamma_\theta : \mathbb{Z} \rightarrow \mathbb{Z}, \theta \in \Theta\},$$

is a family of autocovariance functions such that, for each $\theta \in \Theta$,

$$\gamma_\theta(k) \underset{k \rightarrow \infty}{\sim} c(\theta)k^{2d-1},$$

where $c(\theta)$ is a positive constant. Here, “ \sim ” means that the ratio of the left- and right-hand side converges to one.

- (A2) $\theta_i \in \Theta$ ($i \in \mathbb{N}$) are iid random vectors with distribution G (on Θ) such that the expected values

$$\gamma^*(k) = E_G[\gamma_\theta(k)] = \int_{\Theta} \gamma_\theta(k) dG(\theta) \quad (k \in \mathbb{Z}),$$

and

$$E_G(|k|^{2d-1}) \quad (k \in \mathbb{Z} \setminus \{0\}),$$

are finite. Moreover, $\gamma^*(0) = \ell_1(0)$ and

$$\gamma^*(k) = \ell_1(k)E_G(|k|^{2d-1}) \quad (k \neq 0), \quad (2)$$

where $\ell_1 : \mathbb{Z} \rightarrow \mathbb{R}_+$ is a symmetric function, that is, slowly varying at infinity and such that $\liminf_{k \rightarrow \infty} \ell_1(k) > 0$. The processes $(X_{ij})_{j \in \mathbb{Z}}$ ($i = 1, \dots, n$) are defined by $X_{ij} = e_i(j)$ ($i = 1, \dots, n$; $j \in \mathbb{N}$) where $(e_i(j))_{j \in \mathbb{N}}$ ($i = 1, 2, \dots$) are independent of each other, with zero mean and, given θ_i , autocovariance functions

$$\text{cov}_{\theta_i}(e_i(j), e_i(j+k)) = \gamma_{\theta_i}(k).$$

The aggregated series is defined by

$$Y_j(n) = \sum_{i=1}^n a_i(j; n)X_{ij} = \sum_{i=1}^n a_i(j; n) e_i(j) \quad (j \in \mathbb{N}), \quad (3)$$

with coefficients $a_i(j; n)$ as specified in Equation (A3).

- (A3) For $\lambda_n \in (0, n]$ ($n \in \mathbb{N}$) with $\lim_{n \rightarrow \infty} \lambda_n = \infty$, let $\xi_{in} \in \{0, 1\}$ ($n \in \mathbb{N}$, $1 \leq i \leq n$) be independent zero-one random variables, independent of θ_i ($i \in \mathbb{N}$) and $(e_i(j))_{j \in \mathbb{N}}$ ($i \in \mathbb{N}$), and such that

$$P(\xi_{in} = 1) = \frac{\lambda_n}{n}. \quad (4)$$

The coefficients $a_i(j; n)$ in Equation (3) are given by

$$a_i(j; n) = \xi_{in}(1 \leq i \leq n, j, n \in \mathbb{N}).$$

Let F_d be the marginal distribution of d and $M_d(u)$ its moment generating function. Then

$$E_G(k^{2d-1}) = M_d(2 \log k)k^{-1},$$

so that condition (2) can be written as

$$\gamma^*(k) = \ell_1(k)M_d(2 \log k)k^{-1} \quad (k \geq 1). \quad (5)$$

Assumption (5) arises naturally (see also Ghosh, 2001), as illustrated by the following examples.

Example 1. Let $\theta = (d, \sigma^2)$ ($d \in (0, \frac{1}{2}), \sigma^2 \in \mathbb{R}_+$) and

$$\gamma_\theta(k) = \frac{\sigma^2}{2} (|k-1|^{2d+1} - 2|k|^{2d+1} + |k+1|^{2d+1}).$$

This is the autocovariance function of fractional Gaussian noise with self-similarity parameter $H = d + \frac{1}{2}$. Let d be uniformly distributed on $(0, \frac{1}{2})$ and independent of σ^2 . Also, assume that $\mu_{\sigma^2} = E(\sigma^2) < \infty$. Then

$$M_d(u) = E_G[\exp(ud)] = 2 \frac{\exp\left(\frac{u}{2}\right) - 1}{u},$$

and, for $k \geq 2$,

$$E_G[\gamma_\theta(k)] = \frac{\mu_{\sigma^2}}{2} [g(k-1) - 2g(k) + g(k+1)],$$

where

$$g(m) = mM_d(2 \log m) = m \frac{m-1}{\log m}.$$

Hence,

$$E_G[\gamma_\theta(k)] = \frac{\mu_{\sigma^2}}{2} \left[\frac{(k-1)(k-2)}{\log(k-1)} - 2 \frac{k(k-1)}{\log k} + \frac{(k+1)k}{\log(k+1)} \right] \underset{k \rightarrow \infty}{\sim} \frac{\mu_{\sigma^2}}{2} \frac{1}{\log k}.$$

Note that

$$E_G(k^{2d-1}) = k^{-1}M_d(2 \log k) = k^{-1} \frac{k-1}{\log k} \underset{k \rightarrow \infty}{\sim} \frac{1}{\log k}.$$

Hence,

$$E_G[\gamma_\theta(k)] = \ell_1(k)E(k^{2d-1}),$$

with

$$\ell_1(k) = \frac{\mu_{\sigma^2}}{2} \frac{k}{k-1} \left[\frac{(k-1)(k-2)}{\log(k-1)} - 2 \frac{k(k-1)}{\log k} + \frac{(k+1)k}{\log(k+1)} \right] \log k \underset{k \rightarrow \infty}{\sim} \frac{\mu_{\sigma^2}}{2}.$$

The reason for the slow decay of $E_G[\gamma_{\theta}(k)]$ is that the random values of d can be arbitrarily close to the nonstationarity boundary of $\frac{1}{2}$. The general result is stated below in Theorem 3. Note in particular that the rate remains the same, if d is uniformly distributed on $(d_{\min}, \frac{1}{2})$ for some $d_{\min} \in (0, \frac{1}{2})$. In contrast, if d is uniformly distributed on $(0, d_{\max})$ for some $d_{\max} \in (0, \frac{1}{2})$, then $E_G(k^{2d-1})$ decays at a hyperbolic rate (see Example 5 below).

Under assumptions (A1)–(A3) the autocovariance function of $Y_j(n)$ is of the form

$$\gamma_{Y(n)}(k) = \text{cov}(Y_j(n), Y_{j+k}(n)) = \lambda_n \gamma^*(k). \quad (6)$$

Thus, we obtain the following result:

Theorem 1. *Suppose that (A1), (A2), and (A3) hold. Then*

$$\lambda_n^{-1} \gamma_{Y(n)}(k) = \gamma^*(k) \in \mathbb{R} \quad (k \in \mathbb{Z}).$$

Remark 1. For a related result in the context of repeated time series see Ghosh (2001).

Remark 2. Assumption (A3) means that, as n tends to infinity, the expected number of nonzero coefficients a_i in Equation (3) diverges to infinity. Using the notation $p = \lim_{n \rightarrow \infty} \lambda_n/n$, two cases are possible: (a) $p = 0$; (b) $p \in (0, 1]$. In case (a), the relative number of processes that are included in the aggregated process $Y_j(n)$ is ultimately much smaller than n , whereas it is proportional to n in case (b).

Theorem 1 can be generalized to weak convergence of finite-dimensional distributions to a Gaussian process:

Theorem 2. *Suppose that (A1), (A2), and (A3) hold. Then*

$$\frac{Y_j(n)}{\sqrt{\lambda_n}} \quad (j \in \mathbb{Z}) \Rightarrow_{\text{fidi}} Z_j \quad (j \in \mathbb{Z}),$$

where Z_j is a zero-mean stationary Gaussian process with autocovariance function γ^* and $\Rightarrow_{\text{fidi}}$ denotes weak convergence of finite-dimensional distributions.

Example 2. Let F_d be the uniform distribution of $\theta = d \in (0, \frac{1}{2})$. Then

$$M_d(2 \log k) = \frac{k-1}{\log k}.$$

Hence,

$$\gamma^*(k) = \frac{1}{\ell(k)},$$

with

$$\ell(k) = \frac{k}{k-1} \frac{\log k}{\ell_1(k)} \quad (k \geq 1), \quad \ell(0) = \frac{1}{\ell_1(0)}.$$

Note that $\ell(k)$ is slowly varying at infinity.

Example 3. Let F_d be the truncated $N(0, 1)$ -distribution with truncation points $a = 0$ and $b = \frac{1}{2}$. Then

$$M_d(u) = e^{\frac{1}{2}u^2} \frac{\Phi\left(\frac{1}{2} - u\right) - \Phi(-u)}{\Phi\left(\frac{1}{2}\right) - \frac{1}{2}} \sim C_M u^{-1} \exp\left(\frac{1}{2}u\right),$$

with $C_M = 1/\sqrt{2\pi} \exp\left(-\frac{1}{8}\right)$. Hence

$$\gamma^*(k) \underset{k \rightarrow \infty}{\sim} \ell_1(k) C_M \frac{1}{\log k}.$$

Example 4. Let F_d be the triangular distribution on $[0, \frac{1}{2}]$ with mode $c \in (0, \frac{1}{2})$. Then

$$M_d(u) = 4 \frac{\left(\frac{1}{2} - c\right) - \frac{1}{2} \exp(cu) + c \exp\left(\frac{1}{2}u\right)}{c \left(\frac{1}{2} - c\right) u^2} \underset{u \rightarrow \infty}{\sim} C_M u^{-2} \exp\left(\frac{1}{2}u\right),$$

where $C_M = 8/(1 - 2c)$. Hence

$$\gamma^*(k) \underset{k \rightarrow \infty}{\sim} \ell_1(k) C_M \frac{1}{(\log k)^2}.$$

Thus $\gamma^*(k)$ converges to zero at a slowly varying rate. Note, however, that the decay is faster than for the uniform and for the truncated normal distribution. A possible reason may be that for the triangular distribution the density tends to zero when approaching the right end point $\frac{1}{2}$.

Example 5. Let F_d be the uniform distribution on $(0, d_{\max})$ with $0 < d_{\max} < \frac{1}{2}$. Then

$$M_d(2 \log k) = \frac{k^{2d_{\max}} - 1}{d_{\max} \log k},$$

so that

$$\gamma^*(k) \underset{k \rightarrow \infty}{\sim} \frac{\ell_1(k)}{d_{\max}} \frac{k^{2d_{\max}-1}}{\log k}.$$

The examples indicate that $\gamma^*(k)$ decays very slowly, unless the support of F_d is bounded away from $\frac{1}{2}$. This is confirmed by the following theorem and corollary.

Theorem 3. Let F_d be a distribution with support $(0, \frac{1}{2})$, and such that for all $x < \frac{1}{2}$,

$$1 - F_d(x) = P(d > x) > 0. \quad (7)$$

Then

$$\sup_{0 < \beta < \frac{1}{2}} \lim_{u \rightarrow \infty} \frac{\exp(\beta u)}{M_d(u)} = 0.$$

Theorem 3 means that the growth of $M_d(u)$ is faster than $\exp(\beta u)$ for any $\beta < \frac{1}{2}$. From Theorem 1, we therefore obtain:

Corollary 1. *Let F_d be a distribution with support $(0, \frac{1}{2})$ and such that Equation (7) holds. Then, under the assumptions of Theorem 1,*

$$\sup_{0 < \alpha < 1} \lim_{k \rightarrow \infty} \frac{k^{-\alpha}}{\gamma^*(k)} = 0.$$

More specifically, the examples discussed above and Corollary 1 suggest the assumption

$$\gamma^*(k) = \frac{1}{\ell_2(k)} \quad (8)$$

where $\ell_2(k)$ is slowly varying at infinity and such that $\lim_{k \rightarrow \infty} \ell_2(k) = \infty$. For statistical inference, a functional limit theorem for

$$S_{n,N}(t) = \ell_2^{1/2}(N) \frac{1}{N\sqrt{\lambda_n}} \sum_{j=1}^{[Nt]} \sum_{i=1}^n a_i(j, n) X_{ij} \quad (t \in [0, 1]),$$

is useful.

Theorem 4. *Suppose that (A1), (A2), and (A3) hold, $\gamma^*(k)$ is of the form Equation (8) and $n, N \rightarrow \infty$. Then*

$$S_{n,N}(t) \quad (t \in [0, 1]) \Rightarrow_{D[0,1]} \zeta(t) \quad (t \in [0, 1]),$$

where “ $\Rightarrow_{D[0,1]}$ ” denotes weak convergence in the space of cadlag functions on $[0, 1]$ equipped with the Skorohod metric, and $\zeta(t)$ is a zero-mean Gaussian process with $P(\zeta(0) = 0) = 1$, stationary increments and autocovariance function

$$\zeta(s, t) = \text{cov}(\zeta(s), \zeta(t)) = st \quad (s, t \in [0, 1]).$$

3 | NONPARAMETRIC TREND ESTIMATION

Theorem 4 has major consequences for statistical inference. To illustrate this, we consider nonparametric trend estimation for aggregated time series. Assumption (A2) is replaced by

- (A2') $\theta_i \in \Theta$ ($i \in \mathbb{N}$) and $(e_i(j))_{j \in \mathbb{N}}$ ($i = 1, 2, \dots$) are as in (A2), ξ_{in} ($n \in \mathbb{N}$, $1 \leq i \leq n$) as in (A3). Moreover, $\mu_i \in C^2[0, 1]$ ($i \in \mathbb{N}$) is a sequence of iid random functions, independent of $(e_i(j))_{j \in \mathbb{N}}$ ($i \in \mathbb{N}$) and ξ_{in} ($n \in \mathbb{N}$, $1 \leq i \leq n$), such that for all $t \in [0, 1]$,

$$\mu_\infty(t) = E[\mu_i(t)] \in C^2[0, 1], \quad (9)$$

and $\text{var}(\mu_i(t)) < \infty$. Given $n \in \mathbb{N}$, the aggregated series is defined by

$$Y_j(n) = \sum_{i=1}^n a_i(j; n) X_{ij} \quad (j = 1, \dots, N), \quad (10)$$

where

$$X_{ij} = \mu_i(t_j) + e_i(j), \quad (11)$$

$t_j = j/N$ and $N = N_n \in \mathbb{N}$.

Since the coefficients a_i are independent of the random trend functions μ_i , assumptions (A1), (A2'), and (A3) lead to

$$\lambda_n^{-1} E[Y_j(n)] = \frac{1}{n} \sum_{i=1}^n E[\mu_i(t_j)] = \mu_\infty(t_j).$$

Moreover, for $n, N_n \rightarrow \infty$ and a sequence $j_n \in \mathbb{N}$ with $1 \leq j_n \leq N_n$ and $t_{j_n} = j_n/N_n \rightarrow t \in (0, 1)$, we have

$$\lim_{n \rightarrow \infty} \lambda_n^{-1} E[Y_{j_n}(n)] = \lim_{n \rightarrow \infty} \mu_\infty(t_{j_n}) = \mu_\infty(t) \in \mathbb{R}.$$

We consider kernel estimation of $\mu_\infty(t)$ defined by

$$\hat{\mu}_\infty(t) = \frac{1}{\lambda_n N b} \sum_{j=1}^N K\left(\frac{t - t_j}{b}\right) Y_j(n)$$

where $b = b_N > 0$ is a bandwidth, and $K \geq 0$ is a symmetric kernel function with support $[-1, 1]$, $\int K(u) du = 1$, and $I_B = \int u^2 K(u) du < \infty$. For simplicity of presentation it will be assumed that λ_n is known. The asymptotic bias and variance of $\hat{\mu}_Y$ under assumptions (A1), (A2'), (A3), and (8) are given by the following theorem.

Theorem 5. *Define*

$$C_1(t) = \frac{\mu_\infty^{(2)}(t) I_B}{2},$$

where $\mu_\infty^{(k)}$ ($k = 1, 2, \dots$) denotes derivatives of μ_∞ . Suppose that (A1), (A2'), (A3), and Equation (8) hold, and

$$n, N \rightarrow \infty, b \rightarrow 0, nb^2 \rightarrow \infty, Nb^3 \rightarrow \infty.$$

Then, for $t \in (0, 1)$,

$$B_{n,N}(t) = E[\hat{\mu}_\infty(t)] - \mu_\infty(t) = C_1(t) b^2 + o(b^2) \quad (12)$$

and

$$V_{n,N}(t) = \text{var}(\hat{\mu}_\infty(t)) = \frac{1}{\lambda_n \ell_2(Nb)} + o\left(\frac{1}{\lambda_n \ell_2(Nb)}\right), \quad (13)$$

The mean squared error and the asymptotically optimal bandwidth are of the following form:

Corollary 2. *Under the assumptions of Theorem 5,*

$$\text{MSE}(\hat{\mu}_\infty(t)) = E[(\hat{\mu}_\infty(t) - \mu_\infty(t))^2] = C_1^2(t) b^4 + \frac{1}{\lambda_n \ell_2(Nb)} + o\left(\frac{1}{\lambda_n \ell_2(Nb)}\right) + o(b^4). \quad (14)$$

More specifically, if ℓ_2 is of the form

$$\ell_2(k) = C_2(\log k)^\beta$$

for some $C_2, \beta > 0$, then the asymptotically optimal bandwidth for estimating $\mu_Y(t)$ is given by

$$b_{opt} = C_{opt} \frac{1}{\lambda_n^{1/4}} \left(\frac{1}{\log N} \right)^{(1+\beta)/4}, \quad (15)$$

with

$$C_{opt} = \left(\frac{\beta}{4C_1^2(t)C_2} \right)^{1/4}.$$

Corollary 2 implies that the optimal mean square error is of the order

$$\text{MSE}_{opt}(\hat{\mu}_\infty(t)) \sim \text{const} \cdot \lambda_n^{-1} (\log N)^{-(1+\beta)}.$$

Thus aggregation leads to a very slow rate of convergence of trend estimators.

4 | ASYMPTOTIC EFFECT OF TIME-VARYING AGGREGATION

So far, we considered time-independent coefficients $a_i(1, n) = \dots = a_i(N, n)$. For time-dependent coefficients, quite different results can be obtained. Specifically, we consider the following modification of (A3):

- (A3') For $\lambda_n \in (0, n]$ ($n \in \mathbb{N}$) with $\lim_{n \rightarrow \infty} \lambda_n = \infty$, let $(\xi_{in}(j))_{j \in \mathbb{Z}}$ ($n \in \mathbb{N}, 1 \leq i \leq n$) be independent stationary processes with $\xi_{in}(j) \in \{0, 1\}$,

$$P(\xi_{in} = 1) = \frac{\lambda_n}{n},$$

$$\text{cov}(\xi_{in}(j), \xi_{in}(j+k)) = \gamma_\xi(k; n) \quad (k \in \mathbb{Z}),$$

and

$$|\gamma_\xi(k; n)| \leq \ell_\xi(k) |k|^{2\delta-1} \frac{\lambda_n}{n} \quad (k \in \mathbb{Z}),$$

where $\delta < \frac{1}{2}$ and ℓ_ξ is a slowly varying function (at infinity). Moreover, the processes $(\xi_{in}(j))_{j \in \mathbb{Z}}$ are independent of θ_i ($i \in \mathbb{N}$) and the processes $(e_i(j))_{j \in \mathbb{N}}$ ($i \in \mathbb{N}$). The coefficients $a_i(j; n)$ are defined by

$$a_i(j; n) = \xi_{in}(j) \quad (1 \leq i \leq n, 1 \leq j \leq N_n).$$

Assumption (A3') means that, for each i and n , the sequence of coefficients $a_i(1; n), \dots, a_i(N, n)$ is not constant. Instead, $a_i(j; n)$ ($j = 1, \dots, N$) is a stationary series with $E[a_i(j; n)] = \lambda_n/n$ and autocorrelation function proportional to $\ell_\xi(k) |k|^{2\delta-1}$. This is in contrast to (A3) where $a_i(1; n) = a_i(2; n) = \dots = a_i(N; n)$. Note also that the memory parameter δ in (A3') is fixed and

characterizes the temporal dependence of the coefficients $a_i(j; n)$ only. In contrast, the dependence parameters θ_i of the processes $(e_i(j))_{j \in \mathbb{Z}}$ ($i = 1, \dots, n$) are random variables, as defined in (A2).

Under (A3'), we have:

Theorem 6. *Suppose that (A1), (A2), and (A3') hold. Then*

$$\left| \lim_{n \rightarrow \infty} \lambda_n^{-1} \gamma_{Y(n)}(k) \right| \leq \ell_3(k) M_d (2 \log k) |k|^{2\delta-2},$$

where ℓ_3 is slowly varying and $\delta \in (0, \frac{1}{2})$ is as specified in (A3').

Under Equation (8) we thus obtain:

Corollary 3. *Under Equation (8) and the assumptions of Theorem 6, we have*

$$\left| \lim_{n \rightarrow \infty} \lambda_n^{-1} \gamma_{Y(n)}(k) \right| \leq \ell_4(k) |k|^{2\delta-1}, \tag{16}$$

where ℓ_4 is slowly varying at infinity and $\delta \in (0, \frac{1}{2})$ as specified in (A3').

Remark 3. Theorem 6 and Corollary 3 mean that suitable time-varying coefficients remove the extreme effect of long-range dependence observed in Theorem 1. The intuitive reason is that at each time point, the aggregated series $Y_j(n)$ includes a different subset of the individual series $(X_{ij})_{j \in \mathbb{Z}}$ ($i = 1, \dots, n$). Since the individual series are independent of each other, varying the selected subset at each time-point weakens the temporal dependence in $Y_j(n)$.

Example 6. Let $m \in \mathbb{N}$, $0 < c_n < n$ such that $c_n \rightarrow \infty$ and $c_n/n \rightarrow 0$, and denote by $Z_i(j; n)$ ($j \in \mathbb{Z}$, $i, n \in \mathbb{N}$) independent random variables with

$$P(Z_i(j; n) = 1) = 1 - P(Z_i(j; n) = 0) = 1 - \frac{c_n}{m+1} n^{-1}. \tag{17}$$

Moreover, define

$$a_i(j; n) = 1 - \prod_{s=0}^m Z_i(j-s; n). \tag{18}$$

Then

$$E[a_i(j; n)] = P(a_i(j; n) = 1) = \frac{\lambda_n}{n},$$

where λ_n is the solution of

$$\frac{\lambda_n}{n} = 1 - \left(1 - \frac{c_n}{m+1} n^{-1} \right)^{m+1},$$

and

$$\text{cov}(a_i(j; n), a_i(j+k; n)) = 0 \quad (k \geq m+1).$$

Therefore,

$$\gamma_{Y(n)}(k) = 0 \quad (k \geq m + 1).$$

Note also that $\lambda_n/n \approx c_n/n$ so that $\lambda_n \rightarrow \infty$ and $\lambda_n/n \rightarrow 0$.

Remark 4. In the context of networks, Corollary 3 implies that it can be advantageous to use time-varying routing matrices. The idea of time-varying routing matrices is related to suggestions discussed in the network engineering literature in the context of hyperbolically decaying dependence (e.g., Grossglauser & Bolot 1999; Eliazar 2007). In the case considered here, passing OD-flows through fixed routing matrices leads to an even stronger dependence structure characterized by Equation (8). This makes randomized procedures with time-varying routing matrices even more important.

5 | SIMULATIONS

A small simulation study illustrates the difference between constant and time-varying coefficients. Set $\lambda_n = 1,000$, $n = 10'000$ and $N = 1,000$. Fixed coefficients a_i ($i = 1, \dots, 10'000$) are generated by iid Bernoulli random variables with $P(a_i = 1) = \lambda_n n^{-1}$. Time-varying coefficients (denoted by a_i^*) are defined by Equations (17) and (18), with $c_n = 1,000$, $m = 20$, and $n = 10'000$. More specifically, we define

$$a_i^*(j; n) = 1 - \prod_{s=0}^m Z_i(j-s; n),$$

with $m = 20$, where

$$P(Z_i(j; n) = 1) = 1 - P(Z_i(j; n) = 0) = 1 - \frac{c_n}{m+1} n^{-1}.$$

Each series X_{ij} ($j = 1, \dots, N$) is generated by fractional Gaussian noise with unit variance and d sampled from a uniform distribution on $(0, \frac{1}{2})$. Figure 1b shows, for one sample path $a_i^*(j; n)$ ($j = 1, \dots, N$), the number of nonzero elements as a function of time $j = 1, \dots, 1,000$. Note that in the context of networks, this corresponds to the number of active links. The trajectories of the first 100 coefficients $a_i^*(\cdot; n)$ ($i = 1, \dots, 100$) for time points $j = 1, \dots, 500$ are visualized in Figure 1c, by drawing horizontal lines between points (j_1, i) and (j_2, i) for which $a_i^*(j; n) = 1$ ($j_1 \leq j \leq j_2$). To compare autocorrelations of $Y_j(n) = \sum_{i=1}^n a_i(j; n)X_{ij}$ and $Y_j^*(n) = \sum_{i=1}^n a_i^*(j; n)X_{ij}$ ($j = 1, \dots, 1,000$), 100 sample paths of $Y_j(n)$ and $Y_j^*(n)$ were simulated and sample autocorrelations were calculated for each series. Figure 1a shows $\bar{\rho}(k)$ and $\bar{\rho}^*(k)$ that were obtained by averaging the simulated sample autocorrelations. More specifically, $\bar{\rho}(k)$ and $\bar{\rho}^*(k)$ are plotted against $1/\log k$. As expected, the values of $\bar{\rho}(k)$ are close to a straight line whereas the values of $\bar{\rho}^*(k)$ are close to zero for $k \geq 21$ (i.e., for $1/\log k \leq 1/\log 21$). This illustrates that fixed coefficients lead to a logarithmic dependence structure, whereas the time-varying coefficients reduce dependence.

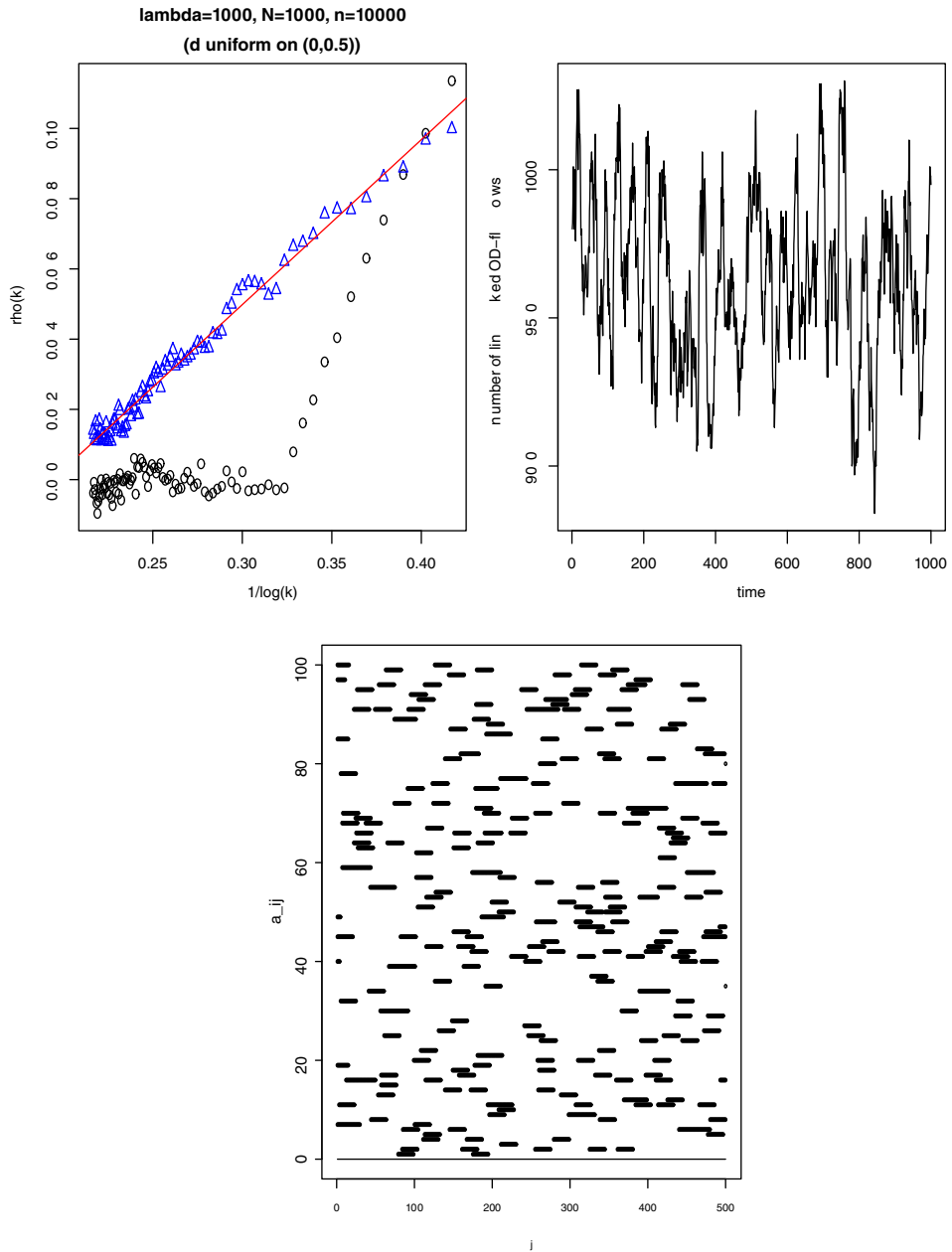


FIGURE 1 (a) Logarithms of average simulated acf for fixed (triangles) and time-varying coefficients (circles) plotted against $1/\log k$ (top left panel); (b) simulated numbers of nonzero time-varying coefficients as a function of time j (top right panel); (c) trajectories of the first 100 coefficients $a_i^*(\cdot; n)$ ($i = 1, \dots, 100$) for time points $j = 1, \dots, 500$ (bottom panel) [Colour figure can be viewed at wileyonlinelibrary.com]

6 | DATA EXAMPLE

Long-range dependence is known to occur frequently in network flows (e.g., Karagiannis, Molle, & Faloutsos, 2004; Leland, Taqqu, Willinger, & Wilson, 1994; Park, Hernández-Campos, Marron, & Smith, 2005). To illustrate this in the context of aggregation, we consider a subset of 38 OD-flows on the Abilene network (Lakhina et al., 2004). The data considered here represent flow volumes measured over consecutive 5-min time intervals for 93 hr, starting on December 25, 2003. Figure 2 shows six typical series (shifted vertically and plotted on a logarithmic for better visibility). The negative slope of the log–log plots of the periodograms in Figure 3 indicate long memory in each of the series. Similar observations can be made for the other series in the data set. Figure 4 shows a histogram of estimated values of d , obtained for each of the 38 series by fitting a FARIMA($p, d, 0$) model using the BIC for choosing the autoregressive order p . The estimated values vary between 0.008 and 0.489. The cross-sectional aggregate of all 38 series is shown in Figure 5a, its periodogram in log–log coordinates in Figure 5b. The very steep slope of the log–log periodogram indicates dependence that is much stronger than for any of the individual series. This observation is confirmed by plotting the sample autocorrelations $\hat{\rho}(k)$ of the individual series (dotted lines) together with the sample autocorrelations for the aggregated series (full line) in Figure 6.

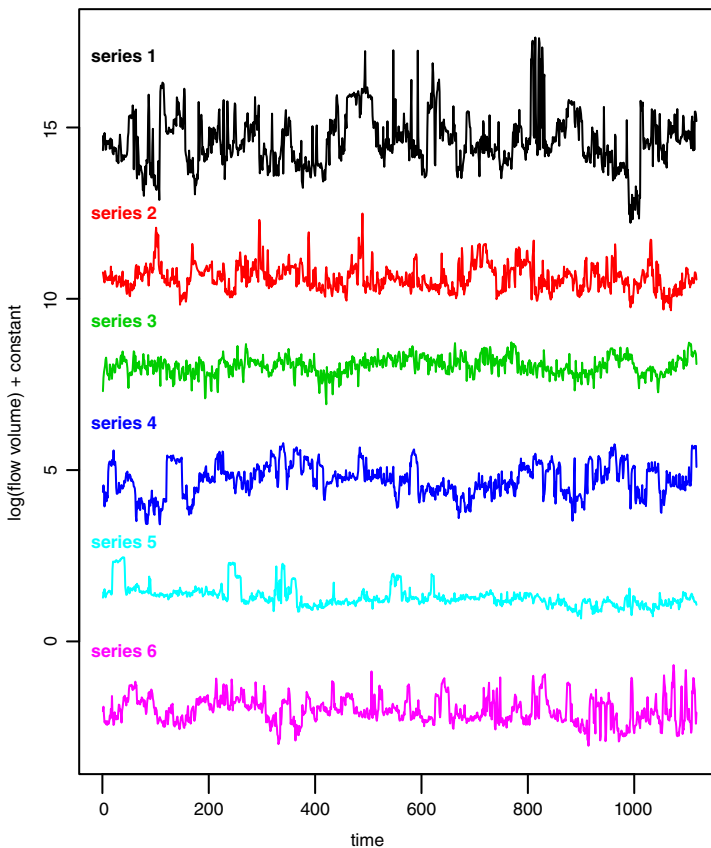


FIGURE 2 Six origin-destination flows on the Abilene network (Lakhina et al. 2004). The data represent flow volumes measured continuously over consecutive 5-min time intervals. For better visibility the series are shifted vertically and plotted on a logarithmic scale [Colour figure can be viewed at wileyonlinelibrary.com]

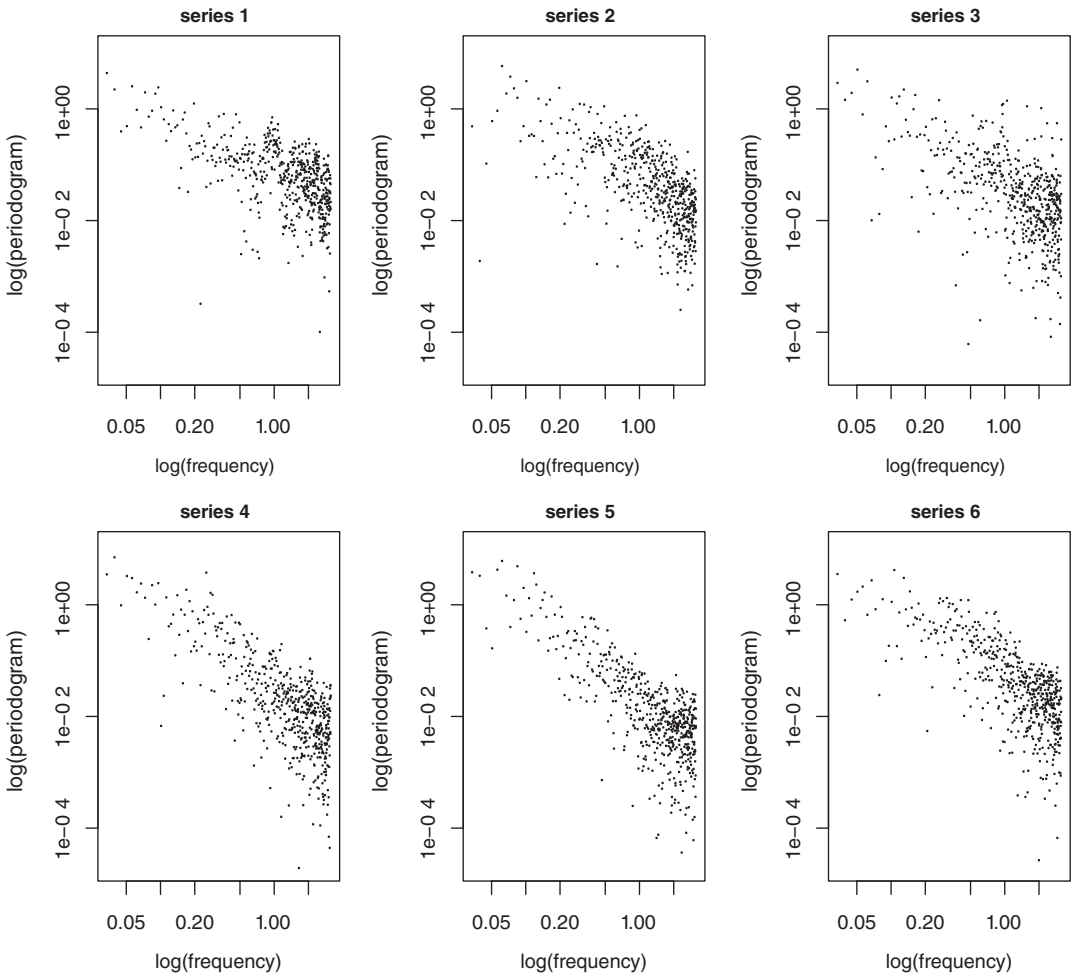


FIGURE 3 Log-log periodogram plots for the series in Figure 2

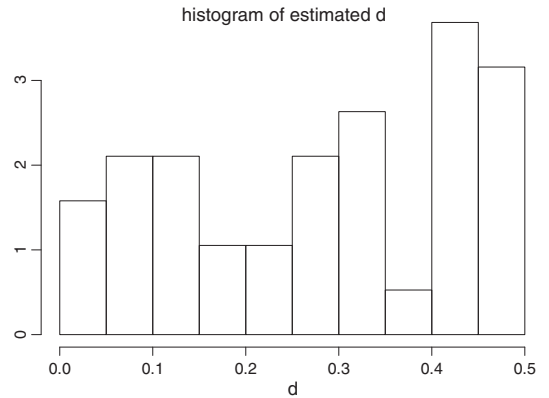


FIGURE 4 Histogram of estimated values of d for 38 origin-destination flows on the Abilene network

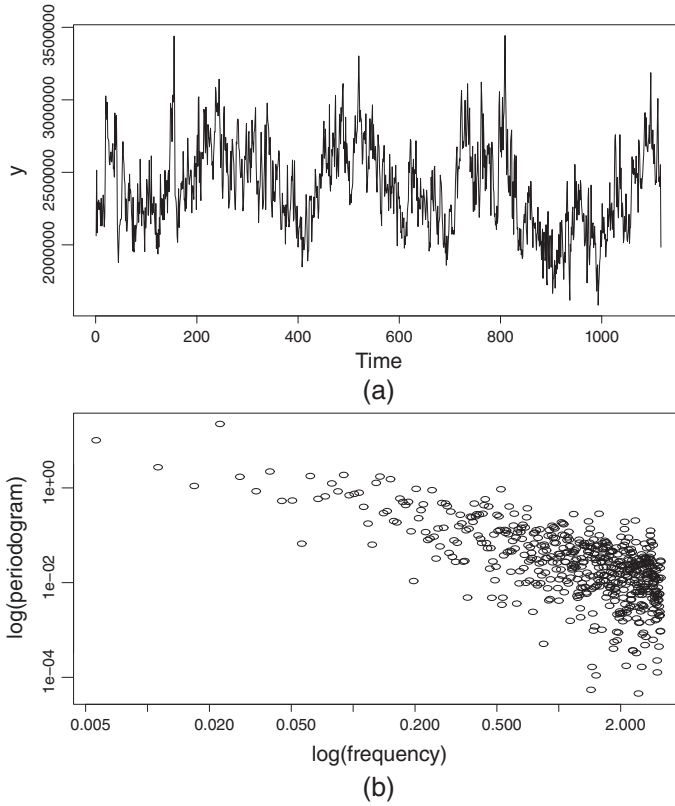


FIGURE 5 (a) Cross-sectional aggregate Y_j of 38 origin-destination flows on the Abilene network; (b) log-log periodogram of Y_j

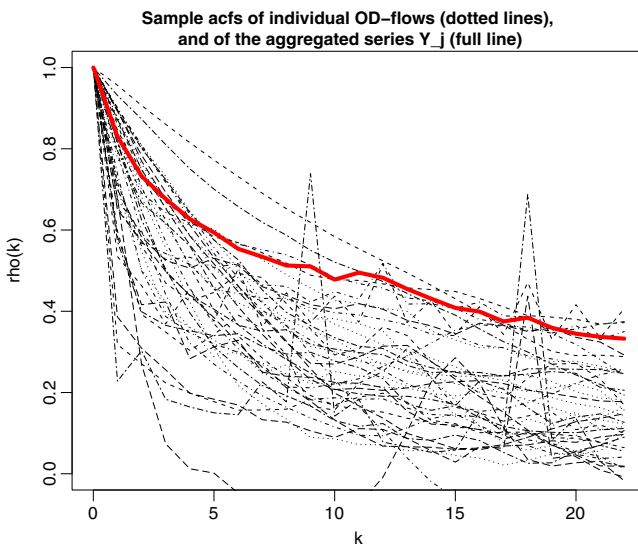


FIGURE 6 Sample autocorrelations of 38 origin-destination flows on the Abilene network (dotted lines), together with sample autocorrelations of the aggregated series (full line) [Colour figure can be viewed at wileyonlinelibrary.com]

7 | FINAL REMARKS

In this paper, we considered cross-sectional aggregation of strongly dependent time series. As it turns out, aggregation with constant coefficients leads to temporal dependence, that is, much stronger than usual long-range dependence. This is a natural extension of results in the literature where, under suitable conditions, aggregation of short-memory processes is shown to induce long-range dependence. On the other hand, suitable time-varying coefficients can reduce dependence. This has important implications for statistical inference. For instance, in the context of networks, it can be advantageous to use time-varying routing matrices.

ACKNOWLEDGEMENTS

The data are due to Lakhina et al. (2004). We would like to thank the referees for their insightful and constructive comments that helped to improve the presentation of the results.

ORCID

Jan Beran  <https://orcid.org/0000-0003-3434-7208>

Haiyan Liu  <https://orcid.org/0000-0002-3442-1722>

REFERENCES

- Abadir, K., & Talmain, G. (2002). Aggregation, persistence and volatility in a macro model. *Review of Economic Studies*, 69(4), 749–779.
- Airoldi, E. M., & Blocker, A. W. (2013). Estimating latent processes on a network from indirect measurements. *Journal of the American Statistical Association*, 108(501), 149–164.
- Beran, J., Feng, Y., Ghosh, S., & Kulik, R. (2013). *Long-memory processes: Probabilistic properties and statistical methods*. Berlin/Heidelberg: Springer.
- Beran, J., Schützner, M., & Ghosh, S. (2010). From short to long memory: Aggregation and estimation. *Computational Statistics & Data Analysis*, 54(11), 2432–2442.
- Billingsley, P. (1968). *Convergence of probability measures*. New York, NY: Wiley.
- Cao, J., Davis, D., Vander Wiel, S., & Yu, B. (2000). Time-varying network tomography: Router link data. *Journal of the American Statistical Association*, 95(452), 1063–1075.
- Castro, R., Coates, M., Liang, G., Nowak, R., & Yu, B. (2004). Network tomography: Recent developments. *Statistical Science*, 19(3), 499–517.
- Chong, T. T. (2006). The polynomial aggregated AR(1) model. *Econometrics Journal*, 9, 98–122.
- Davidson, J., & Sibbertsen, P. (2005). Generating schemes for long-memory processes: Regimes, aggregation and linearity. *Journal of Econometrics*, 128(2), 253–282.
- Ding, Z., & Granger, C. W. J. (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 73(1), 185–215.
- Eliazar, I. (2007). The M/G/∞ system revisited: finiteness, summability, long range dependence, and reverse engineering. *Queueing Systems*, 55(1), 71–82.
- Fang, J., Vardi, Y., & Zhang, C. (2007). *An iterative tomography algorithm for the estimation of network traffic*. In R. Liu, W. Strawderman, & C. Zhang (Eds.), *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*, IMS Lecture Notes—Monograph Series (Vol. 54, pp. 12–23). Beachwood, OH: Institute of Mathematical Statistics.
- Ghosh, S. (2001). Nonparametric trend estimation in replicated time series. *Journal of Statistical Planning and Inference*, 97(2), 263–274.
- Giraitis, L., Leipus, R., & Surgailis, D. (2010). Aggregation of the random coefficient GLARCH(1,1) process. *Econometric Theory*, 26, 406–425.

- Gonçalves, E., & Gouriéroux, C. (1988). Aggrégation de processus autoregressifs d'ordre 1. *Annales d'Économie et de Statistique*, 12, 127–149.
- Granger, C.W.J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14, 227–238.
- Grossglauser, M., & Bolot, J.-C. (1999). On the Relevance of Long-Range Dependence in Network Traffic. *IEEE/ACM Transactions on Networking*, 7(5), 629–640.
- Igloi, E., & Terdik, G. (1999). Long-range dependence through gamma-mixed Ornstein–Uhlenbeck process. *Electronic Journal of Probability*, 4(16), 1–33.
- Karagiannis, T., Molle, M., & Faloutsos, M. (2004). Long-range dependence ten years of Internet traffic modeling. *IEEE Internet Computing*, 8(5), 57–64.
- Kazakevicius, V., Leipus, R., & Viano, M. (2004). Stability of random coefficient ARCH models and aggregation schemes. *Journal of Econometrics*, 120, 139–158.
- Lakhina, A., Paggiannaki, K., Crovella, C., Diot, M., Kolaczyk, E., & Taft, N. (2004). Structural analysis of network traffic flows. *ACM SIGMETRICS Performance Evaluation Review*, 32(1), 61–72.
- Leipus, R., Oppenheim, G., Philippe, A., & Viano, M.-C. (2006). Orthogonal series density estimation in a disaggregation scheme. *Journal of Statistical Planning and Inference*, 136, 2547–2571.
- Leipus, R., & Viano, M.-C. (2002). Aggregation in ARCH models. *Lithuanian Mathematical Journal*, 42, 68–89.
- Leland, W. E., Taqqu, M. S., Willinger, W., & Wilson, D. V. (1994). On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2(1), 1–15.
- Lippi, M. & Zaffaroni, P. (1998). *Aggregation of simple linear dynamics: Exact asymptotic results*, STICERD-LSE Econometrics Discussion Paper 350. London, UK: London School of Economics.
- Nucci, A., & Papagiannaki, K. (2009). *Design, measurement and management of large-scale IP networks*. Cambridge, UK: Cambridge University Press.
- Oppenheim, G., & Viano, M.-C. (2004). Aggregation of random parameters Ornstein–Uhlenbeck or AR processes: Some convergence results. *Journal of Time Series Analysis*, 25, 335–350.
- Ortuzar, D. D., & Willumsen, L. G. (2011). *Modelling transport*. New York, NY: Wiley.
- Park, C., Hernández-Campos, F., Marron, J. S., & Smith, F. D. (2005). Long-range dependence in a changing Internet traffic mix. *Computer Networks*, 48(3), 401–422.
- Pilipauskaite, V., & Surgailis, D. (2014). Joint temporal and contemporaneous aggregation of random-coefficient AR(1) processes. *Stochastic Processes and Their Applications*, 124(2), 1011–1035.
- Pilipauskaite, V., & Surgailis, D. (2015). Joint aggregation of random-coefficient AR(1) processes with common innovations. *Statistics & Probability Letters*, 101, 73–82.
- Tebaldi, C., & West, M. (1998). Bayesian inference on network traffic using link count data. *Journal of the American Statistical Association*, 93(442), 557–573.
- Vardi, Y. (1996). Network tomography: Estimating source-destination traffic intensities from link data. *Journal of the American Statistical Association*, 91, 365–377.
- Zaffaroni, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies. *Journal of Econometrics*, 120, 75–102.
- Zaffaroni, P. (2007a). Aggregation and memory of models of changing volatility. *Journal of Econometrics*, 136, 237–249.
- Zaffaroni, P. (2007b). Contemporaneous aggregation of GARCH processes. *Journal of Time Series Analysis*, 28, 521–544.
- Zhang, Y., Roughan, M., Lund, C. & Donoho, D. (2003). *An information-theoretic approach to traffic matrix estimation*. Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (pp. 301–312). New York, NY: ACM Press.

How to cite this article: Beran J, Liu H, Ghosh S. On aggregation of strongly dependent time series. *Scand J Statist*. 2020;47:690–710. <https://doi.org/10.1111/sjos.12421>

APPENDIX A. PROOFS

Proof of Theorem 1.

$$\begin{aligned}\lambda_n^{-1} \gamma_{Y(n)}(k) &= \lambda_n^{-1} \sum_{i=1}^n E[\xi_{in}^2] E_G[\gamma_\theta(k)] = \lambda_n^{-1} n P(\xi_{in} = 1) \ell_1(k) M_d(2 \log k) k^{-1} \\ &= \frac{\ell_1(k) M_d(2 \log k)}{k}.\end{aligned}$$

Proof of Theorem 2. Note that the processes $(\eta_{i,n}(j))_{j \in \mathbb{Z}}$ ($i = 1, \dots, n$) defined by

$$\eta_{i,n}(j) = a_i(j; n) e_i(j) \quad (j \in \mathbb{Z}),$$

are independent with

$$E[\eta_{i,n}(j)] = 0, \text{var}(\eta_{i,n}(j)) = \frac{\lambda_n}{n} \gamma^*(0).$$

Let

$$\sigma_n^2 = \sum_{i=1}^n \text{var}(\eta_{i,n}(j)) = \lambda_n \gamma^*(0),$$

and recall that $E[e_i^2(j)] = \gamma^*(0) < \infty$, and $a_i(j; n)$ are 0–1–variables that are independent of $e_i(j)$. For $\varepsilon > 0$, we then have

$$\begin{aligned}S_n(\varepsilon) &= \sigma_n^{-2} \sum_{i=1}^n E[\eta_{i,n}^2(j) 1\{|\eta_{i,n}(j)| > \varepsilon \sigma_n\}] \\ &= \sigma_n^{-2} \sum_{i=1}^n E[e_i^2(j) 1\{|e_i(j)| > \varepsilon \sigma_n\} | a_i(j; n) = 1] P(a_i(j; n) = 1) \\ &= \frac{n}{\lambda_n} \cdot E[e_1^2(j) 1\{|e_1(j)| > \varepsilon \sigma_n\}] \frac{\lambda_n}{n} = E[e_1^2(j) 1\{|e_1(j)| > \varepsilon \sqrt{\lambda_n \gamma^*(0)}\}].\end{aligned}$$

Since $E[e_i^2(j)] = \gamma^*(0) < \infty$ ($i = 1, \dots, n$) and $\lambda_n \rightarrow \infty$, we obtain

$$\lim_{n \rightarrow \infty} S_n(\varepsilon) = 0.$$

Thus, the Lindeberg condition and hence also the central limit theorem holds. Analogous arguments can be applied to arbitrary linear combinations $\sum_{l=1}^m c_l Y_{j_l}(n) / \sqrt{\lambda_n}$ ($m \in \mathbb{N}, j_1, \dots, j_m \in \mathbb{Z}$). Thus, for all m and j_1, \dots, j_m , $Y_{j_1, \dots, j_m}(n) / \sqrt{\lambda_n} = (Y_{j_1}(n), Y_{j_2}(n), \dots, Y_{j_m}(n)) / \sqrt{\lambda_n}$ converges in distribution to a normal Gaussian vector $Z_{j_1, \dots, j_m} = (Z_{j_1}, Z_{j_2}, \dots, Z_{j_m})$. The covariance function of Z_{j_1, \dots, j_m} follows from Theorem 1. \blacksquare

Proof of Theorem 3. Let $\beta \in (0, \frac{1}{2})$ and $\beta < \beta^* < \frac{1}{2}$. Then

$$M_d(u) \geq \exp(\beta^* u) P(d > \beta^*).$$

By assumption $\kappa := P(d > \beta^*) > 0$, so that we obtain

$$0 \leq \lim_{u \rightarrow \infty} \frac{\exp(\beta u)}{M_d(u)} \leq \kappa^{-1} \lim_{u \rightarrow \infty} \exp[(\beta - \beta^*)u] = 0,$$

and hence

$$\sup_{0 < \beta < \frac{1}{2}} \lim_{u \rightarrow \infty} \frac{\exp(\beta u)}{M_d(u)} = 0. \quad \blacksquare$$

Proof of Corollary 1. Let F_d be a distribution with support $(0, \frac{1}{2})$, and such that for all $x < \frac{1}{2}$,

$$1 - F_d(x) = P(d > x) > 0.$$

By assumption we have

$$\gamma^*(k) = \ell_1(k)M_d(2 \log k)k^{-1} \quad (k \geq 1).$$

Theorem 3 implies, for any $\beta \in (0, \frac{1}{2})$,

$$\sup_{0 < \beta < \frac{1}{2}} \lim_{k \rightarrow \infty} \frac{\exp(2\beta \log k)}{M_d(2 \log k)} = \sup_{0 < \alpha < 1} \lim_{k \rightarrow \infty} \frac{k^{-\alpha}}{M_d(2 \log k)k^{-1}} = 0.$$

Since ℓ_1 is slowly varying at infinity, this also implies

$$\sup_{0 < \alpha < 1} \lim_{k \rightarrow \infty} \frac{k^{-\alpha}}{\gamma^*(k)} = 0. \quad \blacksquare$$

Proof of Theorem 4. (1) Convergence of finite-dimensional distributions: Let $k \in \mathbb{N}$ and $t_1, \dots, t_m \in [0, 1]$. The processes $(X_{ij})_{j \in \mathbb{Z}}$ are independent copies. Similarly, by assumption, the variables $a_i(1, n)$ ($i = 1, \dots, n$) are iid and, for each i , $a_i(1, n) = a_i(2, n) = \dots = a_i(n, n)$. Theorem 2 then implies weak convergence of $(S_{n,N}(t_1), \dots, S_{n,N}(t_m))$ to a Gaussian random vector. For the autocovariances we have, for $s < t$,

$$E[S_{n,N}(s)S_{n,N}(t)] = E[S_{n,N}^2(s)] + E[S_{n,N}(s)(S_{n,N}(t) - S_{n,N}(s))] \quad (s \leq t).$$

Now,

$$\begin{aligned} \lambda_n \ell_2^{-1}(N) E[S_{n,N}^2(s)] &= N^{-2} \sum_{j_1, j_2=1}^{[Ns]} \sum_{i_1, i_2=1}^n E[a_{i_1}(j_1, n)a_{i_2}(j_2, n)] E[X_{i_1 j_1} X_{i_2 j_2}] \\ &= \sum_{i=1}^n E[a_i^2(j, n)] N^{-2} \sum_{j_1, j_2=1}^{[Ns]} \gamma^*(j_1 - j_2) \\ &= \lambda_n N^{-2} \sum_{k=-([Ns]-1)}^{[Ns]-1} ([Ns] - |k|) \frac{1}{\ell_2(k)} \\ &= \lambda_n (1 + o_N(1)) \frac{s^2}{\ell_2(N)}. \end{aligned}$$

Similarly,

$$\begin{aligned} \lambda_n \ell_2^{-1}(N) E[S_{n,N}(s)(S_{n,N}(t) - S_{n,N}(s))] &= \lambda_n N^{-2} \sum_{j_1=1}^{[Ns]} \sum_{j_2=1}^{[Nt]-[Ns]} \gamma^*(j_2 - j_1) \\ &= \lambda_n (1 + o_N(1)) N^{-2} \frac{[Ns]([Nt] - [Ns])}{\ell_2(N)} \\ &= \lambda_n (1 + o_N(1)) \frac{s(t-s)}{\ell_2(N)}. \end{aligned}$$

Hence,

$$E[S_{n,N}(s)S_{n,N}(t)] = st(1 + o_N(1)).$$

(2) Tightness: For $s < t$, we have

$$E[(S_{n,N}(t) - S_{n,N}(s))^2] = E[S_{n,N}^2(t-s)] = (t-s)^2(1 + o_N(1)).$$

Thus, tightness follows from theorem 15.6 in Billingsley (1968). ■

Proof of Theorem 5. The bias follows by standard arguments. For the variance, Equations (6) and (8) with $\gamma^*(k) = 1/\ell_2(k)$ lead to

$$\begin{aligned} \text{var}(\hat{\mu}_\infty(t)) &= \lambda_n^{-2} \left[(Nb)^{-2} \sum_{j,l=1}^N K\left(\frac{t-t_j}{b}\right) K\left(\frac{t-t_l}{b}\right) \frac{\lambda_n}{\ell_2(j-l)} \right] \\ &= \frac{1}{\lambda_n \ell_2(2Nb)} \left[\int_{-1}^1 \int_{-1}^1 K(x)K(y) dx dy + O((Nb)^{-1}) \right] \\ &= \frac{1}{\lambda_n \ell_2(2Nb)} + o\left(\frac{1}{\lambda_n \ell_2(Nb)}\right). \end{aligned}$$
■

Proof of Corollary 2. Theorem 5 and $Nb^3 \rightarrow \infty$ imply formula (14). Setting the derivative of Equation (14) equal to zero, and noting that

$$(\log N + \log b)^{\beta+1} \sim (\log N)^{\beta+1},$$

yields the asymptotically optimal bandwidth. ■

Proof of Theorem 6.

$$\text{cov}(Y_j(n), Y_{j+k}(n)) = \sum_{i=1}^n E[a_i(j; n)a_i(j+k; n)]E[e_{i_1}(j)e_{i_2}(j+k)] = \sum_{i=1}^n \gamma_\xi(k; n)E_G[\gamma_\theta(k)].$$

Hence,

$$|\text{cov}(Y_j(n), Y_{j+k}(n))| \leq \sum_{i=1}^n |\gamma_\xi(k; n)| |E_G[\gamma_\theta(k)]| \leq \ell_\xi(k) |k|^{2\delta-1} \lambda_n \ell_1(k) M_d(2 \log k) |k|^{-1}.$$

Using the notation

$$\ell_3(k) = \ell_\xi(k)\ell_1(k)$$

leads to

$$|\lambda_n^{-1} \text{cov}(Y_j(n), Y_{j+k}(n))| \leq \ell_3(k)M_d(2 \log k)|k|^{2\delta-2}. \quad \blacksquare$$

Proof of Corollary 3. The result follows directly from Theorem 1. ■