Electronic Journal of Statistics
Vol. 6 (2012) 1670–1714
ISSN: 1935-7524
DOI: 10.1214/12-EJS725

Nonparametric inference for stochastic feedforward networks based on cross-spectral analysis of point processes*

Cornelia Wichelhaus

Institute of Applied Mathematics, University of Heidelberg
Im Neuenheimer Feld 294, 69120 Heidelberg, Germany
e-mail: wichelhaus@statlab.uni-heidelberg.de

and

Roland Langrock

School of Mathematics and Statistics, The Observatory, Buchanan Gardens,
University of St Andrews, St Andrews,
KY16 9LZ, United Kingdom
e-mail: roland@mcs.st-and.ac.uk

Abstract: In this paper we study a nonparametric estimation problem for stochastic feedforward systems with nodes of type G/G/∞. We assume that we have observations of the external arrival and external departure processes of customers of the system, but no information about the movements of the indistinguishable customers in the network. Our aim is the construction of estimators for the characteristic functions and the densities of the service time distributions at the nodes as well as for the routing probabilities. Since the system is only partly observed we have to clarify first if the parameters are identifiable from the given data. The crucial point in our approach is the observation that in our stochastic networks under study the influence of the arrival processes on the departure processes can be described in a linear and time-invariant model. This makes it possible to apply cross-spectral techniques for multivariate point processes. The construction of the estimators is then based on smoothed periodograms. We prove asymptotic normality for the estimators. We present the statistical analysis for a tandem system of two nodes in full details and show afterwards how the results can be generalized to feedforward systems of three or more nodes and to systems with positive feedback probabilities at the nodes.

AMS 2000 subject classifications: Primary 60G55; secondary 62M15, 60K25.
Keywords and phrases: cross-spectral analysis, multivariate point processes, nonparametric statistics for queues, periodogram, self-exciting processes, tandem networks.

Received June 2011.

*This is an original survey paper

1. Introduction

Stochastic networks are complex interacting systems with the typical application fields computer and telecommunication networks, the internet, manufacturing facilities, as well as dynamical population models and models of the neurosciences. In all these fields there is an increasing interest in the statistical inference for system characteristics based on incomplete information of the working systems. For a recent example, have a look at the applied comprehensive modeling and statistical analysis of internet voice communication presented in [33]. In particular, knowledge of the distribution of the operating times at the nodes is essential for specifying the performance behavior and for guaranteeing reliability and quality standards of the systems. It is important to describe the speed and potential monotonicity of service at the nodes and to classify the shape of the service time distribution with respect to unimodality or multimodality. For example, a unimodal service time density shows a homogeneous service behavior whereas a bimodal density may indicate that there are two distinct customer populations or breakdowns of the server. There is interest in both, parametric and nonparametric statistical inference for service time distributions. However, most of the currently applied techniques for service demand estimation in queueing networks focus only on the first and second order moments of the service times obtained e.g. by linear regression of measured utilization values. Some efforts have been done in the recent past to extend queueing networks to more general models with arbitrary distributions and/or nonrenewal processes, but the service distribution estimation problem has remained unsolved up to now.

Important contributions to a parametric analysis of queueing systems are among others the following papers. In [23] the mean service and response times of an IT system are inferred from measurements of the end-to-end delays of the customers. The analysis is based on a mathematical programming formulation. In [22] the traffic intensity in G/G/1 queueing systems is estimated based on samples of the interarrival and service times of customers. The estimators are shown to be consistent and asymptotically normal and corresponding confidence intervals and tests are constructed. Moreover, a comprehensive literature overview on parameter estimation problems in queueing systems is presented. The construction of new confidence intervals for the mean sojourn times in M/G/1-FCFS systems is achieved in [14]. The analysis is based on four different bootstrap procedures. A thorough simulation study is presented. In [15] the analysis is extended to G/G/1-FCFS systems. In [30] traffic parameters like the means of service and waiting times are estimated for various single server systems as well as for queueing networks by sending and observing a test stream. The research is motivated by the need to decide if incoming traffic should be allowed to enter the system in ATM networks. [25] provides a review of general parameter estimation problems for single queues and discusses in particular maximum likelihood estimation, the method of moments and the problem of hypothesis testing in this context. In [21] finite source discrete time Geo/Geo/1/∞/N queues with parallel arrival streams of ordinary customers

and of disasters which empty the system on their arrival are analyzed. Numerical schemes to compute steady state probabilities of the system and performance measures are derived. A numerical study illustrates the sensitivity of the performance measures due to changes of the system's parameters. In [13] a modification of the M/G/∞ system is studied as a model for data traffic generation where the transmission rates may also be random. The statistical analysis in particular concentrates on the influence of changes in the input parameters on the long-range dependence of the accumulated traffic. The connection between geometric coverage processes and queueing systems is exploited for the statistical analysis in [16] and [17](mainly Chapter 2). By interpreting service times in M/G/∞ systems as line segments, busy periods as clumps of segments and idle periods as spacings between clumps, the relation between one dimensional mosaic processes and M/G/∞ systems is established. Based on observations of the busy periods of M/G/∞ systems then in particular estimators of the arrival rate are studied and their efficiency is compared.

Next we give a brief literature overview on nonparametric estimation studies for queueing systems. [18] bases the nonparametric analysis of M/G/∞ systems on data of consecutive sequences of the busy and idle periods. The kernel-based deconvolution approach relies on the trick to expand the characteristic function of the busy period distribution before inversion. The main results are uniform strong consistency for the estimators of the distribution function and the density function of the service time distribution. In a series of papers Pitts and her coauthors contributed to the field of nonparametric inference for queues. They successfully applied a powerful functional approach based on infinite - dimensional versions of the delta method to various settings. [29] studies the problem of estimating the stationary waiting time distribution function in the GI/G/1 queue given random samples from the service and interarrival time distributions. The results are strong consistency and asymptotic normality for the functional estimator. In [2] the arrival rate and the service time distribution function in M/G/ ∞ systems are estimated in a parametric as well as in a nonparametric setting. The analysis is based on data on the queue length process which is either completely observable or just in terms of the idle and busy periods. [3] is devoted to the estimation of the arrival rate and the service time distribution via its generating function in M/G/1 systems given observations of the busy and idle periods of the server. [19] studies nonparametric inference for the service time distribution from the workload process of M/G/1 systems. The analysis exploits the underlying regenerative structure of the sampled workload process and utilizes central limit theorems for regenerative processes and a decompounding approach for geometric random sums. In the early contribution [7] Brillinger studies the problem of nonparametric estimation of the service time distribution in a single G/G/∞ queue based on observations of the input and output processes. His results are asymptotically normal estimators. He applies in this paper the spectral theory of stationary processes. We refer to a thorough presentation to his book [9] written in the context of time series.

To the best of our knowledge up to now there are no results in the important case of nonparametric inference for stochastic networks with two or more nodes

where dependencies between the components due to the moving customers must be taken into consideration. With this paper we contribute to close this gap and present a statistical analysis study of a feedforward network of nodes of type $G/G/\infty$. The external arrival processes of customers to these networks are modelled as general point processes which may in particular show inner dependence. At each node there are infinitely many servers which act independently from each other according to a common general service time distribution. No waiting of customers occurs. Once having left a node customers cannot enter it again on their way through the network. Due to its general interarrival and service times the model is of wide practical interest. Particularly, if nothing is known about the structure of the arrival and service processes this model is often the first choice. If the number of servers is unknown in the application it suggests itself to assume infinitely many servers and thus, to equalize service and sojourn times of customers at the nodes. We assume that we observe over a finite stretch of time at each node the external arrival and external departure times of customers only. It is here not possible to follow individual customer paths through the system and directly assign the departures to the arrivals. Our aim is to provide nonparametric estimators for the service time distributions at the nodes as well as estimators for the unknown routing probabilities. In particular, we will estimate both, the characteristic functions and the densities of the service time distributions. The crucial point in our approach is the observation that in our tandem systems under study the influence of the arrival processes on the departure processes can be described in a linear and time-invariant model. Then the stochastic analysis of the system can be based on models for multivariate point processes and cross-spectral techniques for stationary processes can be applied. The underlying Fourier analysis for stochastic processes originated from the search for hidden periodicities in the time course. As we will see, spectral analysis is a powerful tool for studying inner dependency structures of multivariate processes as well as interrelations between two processes since these properties are well represented in the frequency domain. Our work is in particular a generalization of [7] to the multivariate case with a thorough detailed definition of the mathematical point process framework.

The paper is organized as follows. In Section 2 we summarize the notations used, in particular with respect to point processes. Section 3 describes the network model under consideration (a tandem system of two nodes of type $G/G/\infty$) and its stochastical analysis. Section 4 is devoted to the study of general linear time-invariant models for multivariate point processes. We describe cross-spectral methods for the estimation of characteristics of the model and prove the main results about asymptotic normality for the estimators of the so-called transfer function and the impulse-response function of the model. In Section 5 we show how the results from Section 4 can be applied in the setting of the queueing system presented in Section 3. In particular, we develop estimators for the characteristic functions and the (Lebesgue) densities of the service time distributions at the nodes as well as for the routing probabilities of the model and prove their respective asymptotic normality. Furthermore, we present the multiple coherence as an indicator if service times and routing behavior are

deterministic. In Section 6 we sketch how our statistical approach can be generalized to feedforward systems of more than two nodes and to tandem systems with positive feedback probabilities at the nodes. In Section 7 we show and discuss results from a simulation study.

2. Notation

In the paper we consider point processes on $\mathbb{R}$ to be random counting measures, i.e. we call a measurable mapping D from an underlying probability space $(\Omega, \mathcal{F}, P)$ into the set of all nonnegative integer-valued boundedly finite measures on the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$ a point process. Equivalently, D can be characterized by the sequence of its jump points $D = (\sigma_k^D)_{k\in\mathbb{Z}}$ with the convention $\ldots \leq \sigma_{-2}^D \leq \sigma_{-1}^D < 0 \leq \sigma_0^D \leq \sigma_1^D \leq \ldots$. The number of jumps of a point process D in a Borel time set $B \in \mathcal{B}(\mathbb{R})$ can then be calculated as

$$D(B) = \int \mathbf{1}_B(t)\, dD(t) = \int \mathbf{1}_B(t)\, D(dt) = \sum_{j\in\mathbb{Z}} \mathbf{1}_B(\sigma_j^D).$$

We assume that all considered point processes D are simple, i.e. the jump points are distinct, $\sigma_j^D \neq \sigma_{j+1}^D$ for all j. Further, note that simple point processes do not have any accumulation points so that the sequence of jump points diverges and in finite time sets only a finite number of events occur.

We define n-variate point processes $\mathbf{D}$ on the space $\mathbb{R}$ as vectors whose components are point processes on $\mathbb{R}$, i.e., $\mathbf{D}(B) = (D_1(B), \ldots, D_n(B))^t$ for $B \in \mathcal{B}(\mathbb{R})$, where ^{t} denotes transposition. For further details on point processes we refer the reader to [12].

The conditional expectation $E[D(B)|\mathbf{A}]$ of the random variable $D(B)$ relatively to the multivariate point process $\mathbf{A}$ is defined as the conditional expectation of $D(B)$ relatively to the σ-field generated by the random jump points of all component processes of $\mathbf{A}$. Note that there are at most countably many such jump points since all D_i are simple.

Further, we fix the following notation: $\mathcal{F}[\mu](\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-i\lambda t)\, \mu(dt)$ denotes the Fourier transform of a finite measure μ at frequency $\lambda \in \mathbb{R}$. For real-valued functions h we define the Fourier transform for technical reasons without the factor $1/(2\pi)$, i.e. $\mathcal{F}[h](\lambda) = \int_{-\infty}^{\infty} \exp(-i\lambda t) h(t) dt$. $\mathbf{E}_n$ is the n-dimensional identity matrix, $\mathbf{O}_{n\times m}$ the $n \times m$ matrix with all entries equal to zero and $\mathbf{1}_n$ is the n-dimensional vector with all components equal to 1. $\mathcal{N}_n^{\mathbb{C}}(\mathbf{a}, \mathbf{C})$ indicates an n-variate complex normal distribution with mean vector $\mathbf{a}$ and variance-covariance matrix $\mathbf{C}$. Moreover, for $B \in \mathcal{B}(\mathbb{R})$, let $\mathbf{1}_B(\cdot)$ be the indicator function defined by $\mathbf{1}_B(x) = 1$ if $x \in B$ and $\mathbf{1}_B(x) = 0$ if $x \notin B$. For $B \in \mathcal{B}(\mathbb{R})$ and $u \in \mathbb{R}$ we denote by $B - u$ the set $\{b - u : b \in B\}$. As usual, we set empty sums equal to zero. Furthermore, we assume for all processes and random variables an underlying common probability space $(\Omega, \mathcal{F}, P)$ to be given.

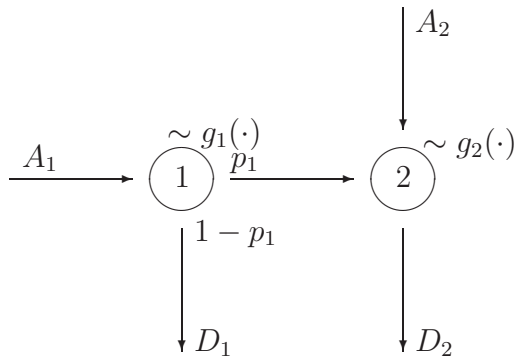

FIG 1. *Tandem network of two nodes.*

3. Tandem models of queues

We present here our approach and results in the framework of a stationary tandem network of two $G/G/\infty$ queues in series according to Figure 1. Generalizations of the model to tandem systems of more than two nodes and to systems with positive feedback probabilities at the nodes are given in Section 6.

We assume that customers which enter the system to obtain service at the nodes are indistinguishable. Customers arrive at nodes 1 and 2 from the outside according to general stationary point processes A_1 and A_2 on the time axis $\mathbb{R}$. We denote the sequences of jump points of the input processes A_1 and A_2 by $(\sigma_j^{A_1})_{j\in\mathbb{Z}}$ and $(\sigma_j^{A_2})_{j\in\mathbb{Z}}$ respectively with $\ldots \le \sigma_{-1}^{A_1} < 0 \le \sigma_0^{A_1} \le \sigma_1^{A_1} \le \ldots$ referring to the arrival times of customers from the outside to node 1 and $\ldots \le \sigma_{-1}^{A_2} < 0 \le \sigma_0^{A_2} \le \sigma_1^{A_2} \le \ldots$ referring to the arrival times of customers from the outside to node 2 respectively. We assume that the point processes A_1 and A_2 are simple, i.e. each jump point $\sigma_j^{A_i}$ of A_i refers to the arrival of exactly one customer and the jump points are distinct, $\sigma_j^{A_i} \neq \sigma_{j+1}^{A_i}$ for all j and $i = 1, 2$.

At each node there are infinitely many servers, no waiting occurs. Service at node i is provided for each customer according to a service time distribution with (Lebesgue) density $g_i(\cdot)$ on $\mathbb{R}_+$ (denoted in Figure 1 by "$\sim g_i(\cdot)$"), $i = 1, 2$. We assume mean service times to be finite, i.e. $\int_0^\infty t\, g_j(t)dt < \infty$ for $j = 1, 2$.

After the completion of service at node 1 a customer decides with probability $1-p_1$ to leave the system and with probability p_1 to move on to node 2 to obtain service there. After being served at node 2 all customers leave the system. We assume here that $p_1 \in (0, 1]$. For $p_1 = 0$ the model would reduce to the single node case considered in [7]. Note that for $p_1 = 1$ we have no observations of the departure process D_1 which is no problem though since then the matrix $\mathbf{H}(\cdot)$ given in (7) shows a simple form which renders the estimation procedure facile.

We denote by $D_1 = (\sigma_j^{D_1})_{j\in\mathbb{Z}}$ and $D_2 = (\sigma_j^{D_2})_{j\in\mathbb{Z}}$ the external output processes from node 1 and node 2 respectively where $(\sigma_j^{D_i})_{j\in\mathbb{Z}}$ is the sequence of

the successive time instants at which customers leave the system from node i, $i = 1, 2$. Clearly, D_1 and D_2 are point processes on $\mathbb{R}$.

We assume that the arrival processes A_1 and A_2 are independent of each other and that all service times at nodes 1 and 2 are independent of each other and of the inter-jump times of the processes A_1 and A_2. Furthermore, the routing decisions of the customers after their respective services at node 1 are taken independently of each other and of the past of the system.

We observe the incoming and outgoing customers of the system over the time interval $(0, T]$, i.e., our observations are part of the sequences of jump points of $A_1 = (\sigma_j^{A_1})_{j\in\mathbb{Z}}$, $A_2 = (\sigma_j^{A_2})_{j\in\mathbb{Z}}$, $D_1 = (\sigma_j^{D_1})_{j\in\mathbb{Z}}$ and $D_2 = (\sigma_j^{D_2})_{j\in\mathbb{Z}}$. We assume the system to be in steady state at time 0. However, the actual number of customers at the nodes at time 0 is not known. The times of movements of customers from node 1 to node 2 are not observable. Our aim is to construct nonparametric estimators of the service time distributions given by the densities $g_1(\cdot)$ and $g_2(\cdot)$ as well as an estimator of the routing probability p_1. Note that we aim high here since we cannot assign the departures to the arrival points and directly measure the individual sojourn times of the customers in the system. If we could, the estimation of $g_1(\cdot)$ and $g_2(\cdot)$ would be straightforward. In particular, for a customer leaving the system via the departure process D_2 we cannot tell if he has entered the system via A_1 or via A_2. Thus, at first it is not clear at all if the unknown parameters are identifiable from the given observations.

The underlying point process structure of the system leads to the following considerations: There is a jump of the output process D_1 in a time Borel set $B \in \mathcal{B}(\mathbb{R})$ if at least for one customer the sum of her arrival time $\sigma_j^{A_1}$ to node 1 and her service time $\gamma_j^{(1)}$ at node 1 is in B and she decided to leave the network after being served at node 1 instead of jumping on to node 2. Therefore, we deduce for the output process D_1:

$$D_1(B) = \sum_{j\in\mathbb{Z}} \mathbf{1}_B(\sigma_j^{A_1} + \gamma_j^{(1)}) \cdot Z_j^{(1-p_1)}, \qquad \text{for } B \in \mathcal{B}(\mathbb{R}), \tag{1}$$

with random Bernoulli variables $Z_j^{(1-p_1)} \sim \text{Bern}(1-p_1)$ and random variables $\gamma_j^{(1)} \sim g_1(\cdot)$ for all $j \in \mathbb{Z}$. All these random variables are independent from each other. Here $Z_j^{(1-p_1)}$ corresponds to the customer's routing decision after the service at node 1 has expired. If $Z_j^{(1-p_1)}$ equals 1 she leaves the network, otherwise she jumps on to node 2.

Taking the conditional expectation of (1) conditioned on the arrival process A_1 we obtain with the theorem of monotone convergence:

$$\begin{aligned} E\Big[D_1(B)\big|A_1\Big] &= \sum_{j\in\mathbb{Z}} E\left(\mathbf{1}_B(\sigma_j^{A_1} + \gamma_j^{(1)}) \,\Big|A_1\right) \cdot E\left(Z_j^{(1-p_1)}\right) \\ &= (1-p_1) \sum_{j\in\mathbb{Z}} \int_0^\infty \mathbf{1}_B(\sigma_j^{A_1} + s)\, g_1(s) ds \end{aligned}$$

$$= (1-p_1) \sum_{j\in\mathbb{Z}} \int_0^\infty \mathbf{1}_{B-\sigma_j^{A_1}}(s) g_1(s) ds$$
$$= (1-p_1) \int\int_0^\infty \mathbf{1}_{B-u}(s) g_1(s) ds\, dA_1(u), \quad \text{for } B \in \mathcal{B}(\mathbb{R}). \quad (2)$$

Analogously we derive for the output process D_2:

$$D_2(B) = \sum_{j\in\mathbb{Z}} \mathbf{1}_B(\sigma_j^{A_1} + \gamma_j^{(1)} + \gamma_j^{(12)}) \cdot (1 - Z_j^{(1-p_1)}) + \sum_{j\in\mathbb{Z}} \mathbf{1}_B(\sigma_j^{A_2} + \gamma_j^{(2)}), \text{ for } B \in \mathcal{B}(\mathbb{R}), \quad (3)$$

with random variables $\gamma_j^{(12)} \sim g_2(\cdot)$ and $\gamma_j^{(2)} \sim g_2(\cdot)$ for all $j \in \mathbb{Z}$, all independent from each other. The reason behind is as follows. There are two different explanations for a jump of D_2 in a time set $B \in \mathcal{B}(\mathbb{R})$: Either for a customer the sum of her arrival time $\sigma_j^{A_1}$ to node 1 and her consecutive service times $\gamma_j^{(1)}$ and $\gamma_j^{(12)}$ at nodes 1 and 2 is in B and she decided after being served at node 1 to jump on to node 2. Or for a customer the sum of her external arrival time $\sigma_j^{A_2}$ to node 2 and her service time $\gamma_j^{(2)}$ at node 2 is in B.

Taking the conditional expectation of (3) conditioned on the arrival processes A_1 and A_2 yields for $B \in \mathcal{B}(\mathbb{R})$:

$$E\left[D_2(B) \middle| \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}\right] = p_1 \int\int_0^\infty \mathbf{1}_{B-u}(s)(g_1 * g_2)(s) ds\, dA_1(u) + \int\int_0^\infty \mathbf{1}_{B-v}(t) g_2(t) dt\, dA_2(v). \quad (4)$$

Summarizing (2) and (4) in a matrix notation gives for $B \in \mathcal{B}(\mathbb{R})$,

$$E\left[\begin{pmatrix} D_1(B) \\ D_2(B) \end{pmatrix} \middle| \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}\right] = \int\int \mathbf{1}_{B-u}(s) \mathbf{h}(s) ds\, d\begin{pmatrix} A_1 \\ A_2 \end{pmatrix}(u) \quad (5)$$

$$\text{with matrix } \mathbf{h}(s) = \begin{pmatrix} g_1(s)\cdot(1-p_1) & 0 \\ (g_1 * g_2)(s)\cdot p_1 & g_2(s) \end{pmatrix}, \; s \in \mathbb{R}_+. \quad (6)$$

For later purposes we compute the Fourier transform of the matrix $\mathbf{h}(\cdot)$ component-by-component,

$$\mathbf{H}(\lambda) = \begin{pmatrix} G_1(\lambda)\cdot(1-p_1) & 0 \\ G_1(\lambda)\cdot G_2(\lambda)\cdot p_1 & G_2(\lambda) \end{pmatrix} \text{ for } \lambda \in \mathbb{R}, \quad (7)$$

where $G_j(\lambda) := \int_0^\infty \exp(-i\lambda t) g_j(t) dt$ for $j = 1, 2$.

We recognize that the influence of the vector $\mathbf{A} := (A_1, A_2)^t$ of arrival processes on the vector $\mathbf{D} := (D_1, D_2)^t$ of departure processes is given in a linear time-invariant manner (for a precise definition see Section 4). We will study general linear time-invariant relations between multivariate point processes in the next section and then in Section 5 apply the results to the particular model given by (5) and (6) to solve our nonparametric estimation problem.

4. Linear time-invariant models for multivariate point processes

This section is devoted to the analysis of functional linear time-invariant relations between two multivariate point processes. These kinds of models were e.g. studied by Brillinger in the context of time series in [9], in the framework of one-dimensional processes with stationary increments in [7] and were sketched in the context of stationary interval functions in [6]. We provide here an analysis in the general framework of multivariate point processes.

Consider a functional relation $\mathcal{R}$ between two multivariate point processes $\mathbf{A} = (A_1, \ldots, A_n)^t$ and $\mathbf{D} = (D_1, \ldots, D_m)^t$ on the space $\mathbb{R}$,

$$\mathbf{D}(B) = \begin{pmatrix} D_1(B) \\ \vdots \\ D_m(B) \end{pmatrix} = \mathcal{R} \begin{pmatrix} A_1 \\ \\ \vdots \\ \\ A_n \end{pmatrix} (B) = \mathcal{R}\left[\mathbf{A}\right](B), \quad \text{for } B \in \mathcal{B}(\mathbb{R}).$$

Thus, the application of the relation $\mathcal{R}$ on the point process $\mathbf{A}$ on $\mathbb{R}$ yields a new point process $\mathbf{D}$ on $\mathbb{R}$ which may have a different dimension.

The relation $\mathcal{R}$ is said to be linear if

$$\mathcal{R}\left[\alpha_1 \mathbf{A_1} + \alpha_2 \mathbf{A_2}\right](B) = \alpha_1 \mathcal{R}\left[\mathbf{A_1}\right](B) + \alpha_2 \mathcal{R}\left[\mathbf{A_2}\right](B) \ \text{ for all } B \in \mathcal{B}(\mathbb{R})$$

for multivariate processes $\mathbf{A_1}, \mathbf{A_2}$ and constants $\alpha_1, \alpha_2 \in \mathbb{R}$. Further, $\mathcal{R}$ is said to be time-invariant if

$$\mathcal{R}\left[\widetilde{\mathbf{A}}^v\right](B) = \mathcal{R}\left[\mathbf{A}\right](B + v) \ \text{ for all } B \in \mathcal{B}(\mathbb{R}),$$

where $\widetilde{\mathbf{A}}^v$ denotes the shifted point process defined by $\widetilde{\mathbf{A}}^v(B) := \mathbf{A}(B + v)$ for all $B \in \mathcal{B}(\mathbb{R})$ and $v \in \mathbb{R}$. A linear, time-invariant operation $\mathcal{R}$ is also called an $m \times n$ linear filter.

We will assume here that a linear filter does only describe the probability structure of the influence of a multidimensional process $\mathbf{A}$ on the process $\mathbf{D}$ up to some additional noise vector ε, which is independent from $\mathbf{A}$, i.e.,

$$\mathbf{D}(B) = \mathcal{R}[\mathbf{A}](B) + \varepsilon(B) \quad \text{with } \mathbb{E}(\varepsilon(B)) = \mathbf{0} \ \text{ for all } B \in \mathcal{B}(\mathbb{R}).$$

Taking conditional expectations conditioned on $\mathbf{A}$ on both sides of the equation yields

$$\mathbb{E}[\mathbf{D}(B)|\mathbf{A}] = \mathcal{R}[\mathbf{A}](B), \ B \in \mathcal{B}(\mathbb{R}).$$

The error process then can be expressed by

$$\varepsilon(B) = \mathbf{D}(B) - \mathbb{E}[\mathbf{D}(B)|\mathbf{A}], \quad B \in \mathcal{B}(\mathbb{R}). \tag{8}$$

We restrict ourselves here to the important class of $m \times n$ matrix linear filters taking the form

$$\mathcal{R}\left[\mathbf{A}\right](B) = \int\int \mathbf{1}_{B-u}(s)\mathbf{h}(s)\, ds\, d\mathbf{A}(u) + \mu, \quad B \in \mathcal{B}(\mathbb{R}), \tag{9}$$

where $\mathbf{h}$ is a nonnegative $m \times n$-matrix with $\int_{-\infty}^{\infty} h_{ij}(u)\, du < \infty$ for all $(i,j) \in \{1,2,\ldots,m\} \times \{1,2,\ldots,n\}$ and μ is an n-dimensional real vector. The vector μ plays the role of an additional linear trend with which in our model the number of expected points of $\mathbf{D}$ given $\mathbf{A}$ may be shifted by a constant. The integral in (9) is to be understood by components, i.e.,

$$\mathcal{R}\begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix}(B)$$
$$= \begin{pmatrix} \int\int \mathbf{1}_{B-u}(s) h_{11}(s)\, ds\, dA_1(u) + \ldots + \int\int \mathbf{1}_{B-u}(s) h_{1n}(s)\, ds\, dA_n(u) + \mu_1 \\ \vdots \\ \int\int \mathbf{1}_{B-u}(s) h_{m1}(s)\, ds\, dA_1(u) + \ldots + \int\int \mathbf{1}_{B-u}(s) h_{mn}(s)\, ds\, dA_n(u) + \mu_n \end{pmatrix}(B).$$

It is obvious that the operation $\mathcal{R}$ is time-invariant, but only linear if μ is the zero vector. However, following Brillinger we call this model in all cases the linear time-invariant model.

Altogether, our model takes the form

$$\mathbb{E}\left[\mathbf{D}(B)\middle|\mathbf{A}\right] = \int\int \mathbf{1}_{B-u}(s)\mathbf{h}(s) ds\, \mathbf{A}(du) + \mu \quad \text{for } B \in \mathcal{B}(\mathbb{R}) \tag{10}$$

with multivariate point processes $\mathbf{A}$ and $\mathbf{D}$.

As in cross-spectral analysis of stationary processes the matrix $\mathbf{h}(\cdot)$ in (9) is called the impulse-response function of the filter $\mathcal{R}$. The Fourier transform of $\mathbf{h}(\cdot)$ taken component-by-component,

$$\mathbf{H}(\lambda) = \int_{-\infty}^{\infty} \mathbf{h}(u) \exp(-i\lambda u) du \quad \text{for } \lambda \in \mathbb{R}, \tag{11}$$

is called transfer function of $\mathcal{R}$.

Our aim is to derive estimators for μ and $\mathbf{H}(\cdot)$ and then for $\mathbf{h}(\cdot)$ from observations of the point processes $\mathbf{A}$ and $\mathbf{D}$ in the time interval $(0,T]$. The approach is as follows: After having introduced the relevant notations we derive a crucial relation between the matrix $\mathbf{H}(\cdot)$ and spectral matrices of the point processes $\mathbf{A}$ and $\mathbf{D}$ (see *(ii)* of Proposition 4.4 below). We then apply smoothed periodograms as estimators for the spectra to plug-in estimators for $\mathbf{H}(\cdot)$ given by equation (19) below. We show under assumptions that these estimators $\mathbf{H}^{(T)}(\cdot)$ are asymptotically normal in Theorem 4.10. Then, we define estimators for $\mathbf{h}(\cdot)$ based on the estimators $\mathbf{H}^{(T)}(\cdot)$ and prove under additional assumptions also their asymptotic normality, see Theorem 4.11.

For the statistical analysis we make the following assumptions:

Assumption 4.1. *The $(n+m)$-variate point process $\mathbf{N} := (\mathbf{A}, \mathbf{D})^t$ on $\mathbb{R}$ is simple, stationary and shows existing moments of all orders, i.e., for all $s \geq 1$,*

$\mathbb{E}\left[N_{a_1}(B_1) \cdot \ldots \cdot N_{a_s}(B_s)\right] < \infty$ *for all $a_1, \ldots, a_s \in \{1, \ldots, n+m\}$ and for all $B_1, \ldots, B_s \in \mathcal{B}(\mathbb{R})$ with finite Lebesgue measures.*

Stationarity of the process $\mathbf{N}$ implies that the cumulant measures of the process decouple into so-called reduced cumulant measures and Lebesgue components which only depend on the lag of the arguments (see e.g. [11] (Chapter 8) and [6]), i.e., for all $k \geq 1$, $a_1, \ldots, a_k \in \{1, \ldots, n+m\}$ and $B_1, \ldots, B_k \in \mathcal{B}(\mathbb{R})$ we have

$$\begin{aligned} &\operatorname{Cum}\left(N_{a_1}(B_1), \ldots, N_{a_{k-1}}(B_{k-1}), N_{a_k}(B_k)\right) \\ &= \int_{B_k} \int_{B_1} \cdots \int_{B_{k-1}} d\breve{C}_{N_{a_1} \ldots N_{a_k}}(t_1 - t_k, \ldots, t_{k-1} - t_k)\, dt_k \end{aligned} \tag{12}$$

with $k-1$-dimensional measures $\breve{C}_{N_{a_1} \ldots N_{a_k}}(du_1, \ldots, du_{k-1})$ on $\mathcal{B}(\mathbb{R}^{k-1})$. Note that the (reduced) cumulant measures are in general only signed measures.

(12) directly implies for the first moment measures that there are constant vectors $\eta^{\mathbf{N}} \in \mathbb{R}^{n+m}$ such that

$$\mathbb{E}\mathbf{N}(B) = \mathbb{E}\left(\int \mathbf{1}_B(s) d\mathbf{N}(s)\right) = \eta^{\mathbf{N}} \cdot \left(\int \mathbf{1}_B(s) ds\right). \tag{13}$$

In particular, it follows that there are vectors $\eta^{\mathbf{A}} \in \mathbb{R}^n$ and $\eta^{\mathbf{D}} \in \mathbb{R}^m$ with

$$\mathbb{E}\mathbf{A}((0,t]) = \eta^{\mathbf{A}}\, t \;\text{ and }\; \mathbb{E}\mathbf{D}((0,t]) = \eta^{\mathbf{D}}\, t.$$

Via the construction of measurable functions (13) can be extended in the usual way for nonnegative measurable functions $f : \mathbb{R} \to \mathbb{R}$:

$$\mathbb{E}\left(\int f(s) d\mathbf{N}(s)\right) = \eta^{\mathbf{N}} \cdot \left(\int f(s) ds\right). \tag{14}$$

Moreover, it follows from (12) that for the covariance measures there exist for $a_1, a_2 \in \{1, 2, \ldots, n+m\}$ reduced covariance measures $\breve{C}_{N_{a_1} N_{a_2}}(dt)$ such that for nonnegative measurable functions $f, g : \mathbb{R} \to \mathbb{R}$ we have

$$\operatorname{Cov}\left(\int f(s) dN_{a_1}(s), \int g(r) dN_{a_2}(r)\right) = \int \int f(s+t) g(s) d\breve{C}_{N_{a_1} N_{a_2}}(t)\, ds. \tag{15}$$

In order to define Fourier transforms of the reduced covariance measures we impose so-called mixing conditions on these measures which are kind of asymptotic independence conditions on the increments of the point process $\mathbf{N}$.

Assumption 4.2. *The $(n+m)$-variate point process $\mathbf{N} := (\mathbf{A}, \mathbf{D})^t$ on $\mathbb{R}$ fulfills Assumption 4.1 and for all $k \geq 2$ and $a_1, \ldots, a_k \in \{1, \ldots, n+m\}$ the reduced cumulant measure of $N_{a_1}, \ldots, N_{a_k}$ satisfies*

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |u_j| \cdot \left|\breve{C}_{N_{a_1} \ldots N_{a_k}}(du_1, \ldots, du_{k-1})\right| < \infty \quad \textit{for all } j \in \{1, \ldots, k-1\}.$$

Particularly, it follows that under Assumption 4.2, $\breve{C}_{N_{a_1} \ldots N_{a_k}}(\cdot)$ are (signed) finite measures on $\mathbb{R}^{k-1}$ for all $k \geq 2$.

Definition 4.3. The second-order spectra $f_{N_jN_k}$ are defined as the Fourier transforms of the reduced covariance measures:

$$f_{N_jN_k}(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-i\lambda u)\breve{C}_{N_jN_k}(du) \quad \text{for } \lambda \in \mathbb{R} \text{ and } j,k \in \{1,2,\ldots,n+m\}.$$

For $j = k$, $f_{N_jN_j}$ is real and nonnegative and called power spectrum, for $j \neq k$, $f_{N_jN_k}$ is complex-valued and called cross spectrum. For $\lambda \in \mathbb{R}$ we summarize spectra of particular importance in the matrices

$$\begin{aligned}
f_{\mathbf{DA}}(\lambda) &:= \big(f_{D_iA_j}(\lambda)\big)_{(i,j)\in\{1,\ldots,m\}\times\{1,\ldots,n\}},\\
f_{\mathbf{AD}}(\lambda) &:= \big(f_{A_jD_i}(\lambda)\big)_{(j,i)\in\{1,\ldots,n\}\times\{1,\ldots,m\}},\\
f_{\mathbf{AA}}(\lambda) &:= \big(f_{A_jA_k}(\lambda)\big)_{(j,k)\in\{1,\ldots,n\}\times\{1,\ldots,n\}},\\
f_{\mathbf{DD}}(\lambda) &:= \big(f_{D_iD_l}(\lambda)\big)_{(i,l)\in\{1,\ldots,m\}\times\{1,\ldots,m\}}.
\end{aligned}$$

We have $\overline{f_{\mathbf{DA}}(\lambda)}^t = f_{\mathbf{AD}}(\lambda)$ for all $\lambda \in \mathbb{R}$, since

$$\begin{aligned}
2\pi\overline{f_{D_iA_j}(\lambda)} &= \int_{-\infty}^{\infty} \overline{\exp(-i\lambda u)}\breve{C}_{D_iA_j}(du) = \int_{-\infty}^{\infty} \exp(i\lambda u)\breve{C}_{D_iA_j}(du)\\
&= \int_{-\infty}^{\infty} \exp(i\lambda(-u))\breve{C}_{D_iA_j}(-du) = \int_{-\infty}^{\infty} \exp(-i\lambda u)\breve{C}_{A_jD_i}(du)\\
&= 2\pi f_{A_jD_i}(\lambda) \text{ for all } i \in \{1,\ldots,m\} \text{ and } j \in \{1,\ldots,n\}.
\end{aligned}$$

Note that the error process ε given in (8) also defines a multivariate random measure on $\mathcal{B}(\mathbb{R})$. With the corresponding matrix $\breve{C}_{\varepsilon\varepsilon}$ of reduced covariance measures we have for $\lambda \in \mathbb{R}$,

$$f_{\varepsilon\varepsilon}(\lambda) = \mathcal{F}\big(\breve{C}_{\varepsilon\varepsilon}\big)(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-i\lambda t)\,\breve{C}_{\varepsilon\varepsilon}(dt).$$

Important properties of the model (10) are now summarized in the next theorem.

Theorem 4.4. *Let Assumption 4.2 be satisfied and suppose that for a fixed λ the matrix $f_{\mathbf{AA}}(\lambda)$ is not singular. For the point processes $\mathbf{A}$ and $\mathbf{D}$ of the model (10) then holds*

$$\begin{aligned}
&(i)\ \ \eta^{\mathbf{D}} = \mathbf{H}(0)\cdot\eta^{\mathbf{A}} + \mu,\\
&(ii)\ \ f_{\mathbf{DA}}(\lambda) = \mathbf{H}(\lambda)\cdot f_{\mathbf{AA}}(\lambda),\\
&(iii)\ \ f_{\varepsilon\varepsilon}(\lambda) = f_{\mathbf{DD}}(\lambda) - f_{\mathbf{DA}}(\lambda)f_{\mathbf{AA}}(\lambda)^{-1}f_{\mathbf{AD}}(\lambda).
\end{aligned}$$

We first note that for a nonnegative function h on $\mathbb{R}$ and a measure $\breve{C}$ on $\mathcal{B}(\mathbb{R})$ the convolution $h * \breve{C}$ is defined by

$$(h * \breve{C})(B) = \int_B \int h(x-t)d\breve{C}(t)dx \text{ for } B \in \mathcal{B}(\mathbb{R}).$$

Thus, $h * \breve{C}$ is a measure on $\mathcal{B}(\mathbb{R})$ with Lebesgue density $\int h(x-t)d\breve{C}(t)$, $x \in \mathbb{R}$.

Proof. The proof of (i) is straightforward: Choosing $B = (0,1]$ and taking the mean on both sides of (10) yields with (14):

$$\eta^{\mathbf{D}} = \mathbb{E}\mathbf{D}((0,1]) = \left(\int\int \mathbf{1}_{(-u,1-u]}(s)\mathbf{h}(s)ds\,du\right)\eta^{\mathbf{A}} + \mu$$
$$= \left(\int \mathbf{h}(s)\int \mathbf{1}_{(-s,1-s]}(u)du\,ds\right)\eta^{\mathbf{A}} + \mu = \left(\int \mathbf{h}(s)ds\right)\eta^{\mathbf{A}} + \mu = \mathbf{H}(0)\eta^{\mathbf{A}} + \mu.$$

Note that the interchange of integral order in the third equation is valid due to the theorem of Tonelli since by assumption each component of the matrix $\int \mathbf{h}(s)ds$ is finite.

For the proof of *(ii)* we have on the one hand for $i = 1,\ldots,m, j = 1,\ldots,n$ and for all $B \in \mathcal{B}(\mathbb{R})$ due to (15):

$$\begin{aligned}
&\mathrm{Cov}(D_i(B), A_j(B)) = \mathrm{Cov}\left(\int \mathbf{1}_B(u)dD_i(u), \int \mathbf{1}_B(t)dA_j(t)\right)\\
&= \int\int \mathbf{1}_B(s+t)\mathbf{1}_B(s)d\breve{C}_{D_iA_j}(t)ds = \int_B\int_{B-s} d\breve{C}_{D_iA_j}(t)ds\\
&= \int_B \breve{C}_{D_iA_j}(B-s)ds. \qquad (16)
\end{aligned}$$

On the other hand we compute for $i = 1,\ldots,m, j = 1,\ldots,n$ and for all $B \in \mathcal{B}(\mathbb{R})$:

$$\begin{aligned}
&\mathrm{Cov}(D_i(B), A_j(B)) = \mathrm{Cov}\left(\mathbb{E}\left(D_i(B)\middle|\mathbf{A}\right), A_j(B)\right)\\
&= \sum_{k=1}^n \mathrm{Cov}\left(\int\int \mathbf{1}_{B-u}(s)h_{ik}(s)ds\,dA_k(u), \int \mathbf{1}_B(t)dA_j(t)\right)\\
&= \sum_{k=1}^n \int \mathbf{1}_B(v)\int\int \mathbf{1}_{B-v-t}(s)h_{ik}(s)ds\,d\breve{C}_{A_kA_j}(t)\,dv\\
&= \sum_{k=1}^n \int_B\int_{B-v}\int h_{ik}(x-t)d\breve{C}_{A_kA_j}(t)dx\,dv\\
&= \int_B \sum_{k=1}^n \left(h_{ik} * \breve{C}_{A_kA_j}\right)(B-v)\,dv. \qquad (17)
\end{aligned}$$

We conclude from (16) and (17) that the measures $\breve{C}_{D_iA_j}(\cdot)$ and $\sum_{k=1}^n \left(h_{ik} * \breve{C}_{A_kA_j}\right)(\cdot)$ are equal on $\mathcal{B}(\mathbb{R})$. This implies

$$\begin{aligned}
\left(f_{\mathbf{DA}}(\lambda)\right)_{ij} &= f_{D_iA_j}(\lambda) = \frac{1}{2\pi}\int \exp(-i\lambda t)\breve{C}_{D_iA_j}(dt)\\
&= \sum_{k=1}^n \frac{1}{2\pi}\int \exp(-i\lambda t)(h_{ik} * \breve{C}_{A_kA_j})(dt) = \sum_{k=1}^n \mathcal{F}\left[h_{ik} * \breve{C}_{A_kA_j}\right](\lambda)\\
&= \sum_{k=1}^n H_{ik}(\lambda)\cdot f_{A_kA_j}(\lambda) = \left(\mathbf{H}(\lambda)\cdot f_{\mathbf{AA}}(\lambda)\right)_{ij}.
\end{aligned}$$

For the proof of *(iii)* we set for a function h, $\widetilde{h}(x) := h(-x)$ for $x \in \mathbb{R}$. Fix $i, j = 1, \ldots, m$ and $B_1, B_2 \in \mathcal{B}(\mathbb{R})$. We first note that due to (8),

$$\mathrm{Cov}(D_i(B_1), D_j(B_2)) = \mathrm{Cov}(\varepsilon_i(B_1), \varepsilon_j(B_2)) + \mathrm{Cov}\Big(\mathbb{E}[D_i(B_1)|\mathbf{A}],\, \mathbb{E}[D_j(B_2)|\mathbf{A}]\Big),$$

since

$$\begin{aligned}
&\mathrm{Cov}\Big(\varepsilon_i(B_1), \mathbb{E}[D_j(B_2)|\mathbf{A}]\Big) \\
&= \mathrm{Cov}\Big(D_i(B_1), \mathbb{E}[D_j(B_2)|\mathbf{A}]\Big) - \mathrm{Cov}\Big(\mathbb{E}[D_i(B_1)|\mathbf{A}], \mathbb{E}[D_j(B_2)|\mathbf{A}]\Big) = 0.
\end{aligned}$$

Next we calculate with (15):

$$\begin{aligned}
\mathrm{Cov}(D_i(B_1), D_j(B_2)) &= \mathrm{Cov}\Big(\int \mathbf{1}_{B_1}(s) dD_i(s), \int \mathbf{1}_{B_2}(t) dD_j(t)\Big) \\
&= \int\int \mathbf{1}_{B_1}(s)\mathbf{1}_{B_2}(s+t)\breve{C}_{D_j D_i}(dt)ds = \int_{B_1} \breve{C}_{D_j D_i}(B_2 - s)ds.
\end{aligned}$$

Similarly, it follows:

$$\mathrm{Cov}(\varepsilon_i(B_1), \varepsilon_j(B_2)) = \int_{B_1} \breve{C}_{\varepsilon_j \varepsilon_i}(B_2 - s)ds.$$

Furthermore we have:

$$\begin{aligned}
&\mathrm{Cov}\Big(\mathbb{E}[D_i(B_1)|\mathbf{A}],\, \mathbb{E}[D_j(B_2)|\mathbf{A}]\Big) \\
&= \sum_{k,l=1}^{n} \mathrm{Cov}\Big(\int\int \mathbf{1}_{B_1-u}(s)h_{ik}(s)ds\, dA_k(u), \int\int \mathbf{1}_{B_2-v}(t)h_{jl}(t)dt\, dA_l(v)\Big) \\
&= \sum_{k,l=1}^{n} \int\int \Big(\int \mathbf{1}_{B_1-u}(s)h_{ik}(s)ds\Big)\Big(\int \mathbf{1}_{B_2-u-v}(t)h_{jl}(t)dt\Big)\, \breve{C}_{A_l A_k}(dv)du \\
&= \sum_{k,l=1}^{n} \int\int \Big(\int_{B_1} \widetilde{h}_{ik}(u-s)ds\Big)\Big(\int \mathbf{1}_{B_2-u}(t)h_{jl}(t-v)dt\Big)\, \breve{C}_{A_l A_k}(dv)du \\
&= \sum_{k,l=1}^{n} \int \Big(\int_{B_1} \widetilde{h}_{ik}(u-s)ds\Big)\, (h_{jl} * \breve{C}_{A_l A_k})(B_2 - u)du \\
&= \sum_{k,l=1}^{n} \int_{B_1} \Big(\int \widetilde{h}_{ik}(u-s)(h_{jl} * \breve{C}_{A_l A_k})(B_2 - u)du\Big)\, ds \\
&= \sum_{k,l=1}^{n} \int_{B_1} \Big(\int_{B_2-s} \int \widetilde{h}_{ik}(x-t)\, d(h_{jl} * \breve{C}_{A_l A_k})(t)\, dx\Big)\, ds \\
&= \sum_{k,l=1}^{n} \int_{B_1} \Big(\widetilde{h}_{ik} * h_{jl} * \breve{C}_{A_l A_k}\Big)\, (B_2 - s)\, ds.
\end{aligned}$$

Putting all together yields that the measures $\breve{C}_{D_jD_i}(\cdot)$ and $\breve{C}_{\varepsilon_j\varepsilon_i}(\cdot) + \sum_{k,l=1}^{n}(h_{jl} * \breve{C}_{A_lA_k} * \widetilde{h}_{ik})(\cdot)$ are equal on $\mathcal{B}(\mathbb{R})$. This implies:

$$\begin{aligned} f_{\varepsilon_j\varepsilon_i}(\lambda) &= \frac{1}{2\pi}\int \exp(-i\lambda t)d\breve{C}_{\varepsilon_j\varepsilon_i}(t) \\ &= f_{D_jD_i}(\lambda) - \sum_{k,l=1}^{n} H_{jl}(\lambda)f_{A_lA_k}(\lambda)\overline{H_{ik}(\lambda)} \\ &= f_{D_jD_i}(\lambda) - \sum_{k=1}^{n}\left(\mathbf{H}(\lambda)f_{\mathbf{AA}}(\lambda)\right)_{jk}\left(\overline{\mathbf{H}(\lambda)}^t\right)_{ki} \\ &= f_{D_jD_i}(\lambda) - \left(\mathbf{H}(\lambda)f_{\mathbf{AA}}(\lambda)\overline{\mathbf{H}(\lambda)}^t\right)_{ji}. \end{aligned}$$

Thus, using *(ii)*, $f_{\varepsilon\varepsilon}(\lambda) = f_{\mathbf{DD}}(\lambda) - f_{\mathbf{DA}}(\lambda)f_{\mathbf{AA}}(\lambda)^{-1}f_{\mathbf{AD}}(\lambda)$. □

Remark 4.5. (a) *(i)* and *(ii)* of Theorem 4.4 guarantee that the model (10) is identifiable under the condition that the matrix $f_{\mathbf{AA}}(\lambda)$ is invertible for $\lambda \in \mathbb{R}$. Then the transfer function $\mathbf{H}(\lambda)$ and the constant shift $\mu \in \mathbb{R}^n$ can be determined provided that we have sufficient information about $\mathbf{A}$ and $\mathbf{D}$. According to *(i)* and *(ii)* the second-order spectral matrices together with the intensity vectors $\eta^{\mathbf{D}}$ and $\eta^{\mathbf{A}}$ represent such information. The mean intensity vectors $\eta^{\mathbf{D}}$ and $\eta^{\mathbf{A}}$ can be estimated by $\frac{E\mathbf{D}((0,T])}{T}$ and $\frac{E\mathbf{A}((0,T])}{T}$ respectively, see Proposition 5.6 below for a discussion of these estimators. For the construction of estimators of the spectral matrices see below. Natural candidates for estimators of $\mathbf{H}(\lambda)$ and μ then are given by

$$\widehat{\mu} = \widehat{\eta}^{\mathbf{D}} - \widehat{\mathbf{H}}(0)\cdot\widehat{\eta}^{\mathbf{A}} \quad \text{and} \quad \widehat{\mathbf{H}}(\lambda) = \widehat{f}_{\mathbf{DA}}(\lambda)\cdot\widehat{f}_{\mathbf{AA}}(\lambda)^{-1}.$$

(b) In case of $n = m = 1$ the gain of the process D over the process A at frequency λ is given by:

$$L(\lambda) := |H(\lambda)|.$$

If the error-process ε is equal to zero then *(ii)* and *(iii)* yield (compare with the discussion of (37) below)

$$f_{DD}(\lambda) = |H(\lambda)|^2 f_{AA}(\lambda) = L(\lambda)^2 f_{AA}(\lambda).$$

This explains the term gain since in the model (10) the amplitudes of the input process $\mathbf{A}$ at particular frequencies are intensified by L.

The most common nonparametric estimators of second-order spectra are based on the so-called periodogram

$$I^{(T)}_{N_jN_k}(\lambda) := \frac{1}{2\pi T}\left(\int_0^T \exp(-i\lambda t)dN_j(t)\right)\cdot\left(\int_0^T \exp(i\lambda t)dN_k(t)\right) \quad \text{for } \lambda \in \mathbb{R},$$

which results from taking the Fourier transform of the natural estimator $\breve{C}^{(T)}_{N_jN_k}$ of the reduced covariance measure $\breve{C}_{N_jN_k}$ given by

$$\breve{C}^{(T)}_{N_jN_k}(du) := \frac{1}{T}\int_0^{T-u} dN_j(t+u)dN_k(t), \quad u \in \mathbb{R}.$$

The periodogram is asymptotically unbiased, but unstable and not consistent. It is known that the periodogram ordinates at distinct frequencies near a frequency λ are asymptotically independent estimates of the spectrum (see e.g. [9], Theorem 5.2.6 for the case of time series). This leads to the fruitful idea of smoothing the periodogram in the neighborhood of λ by means of weighted averages of several periodograms:

Definition 4.6. The smoothed periodogram at frequency $\lambda \in \mathbb{R}$ is given by

$$f^{(T)}_{N_jN_k}(\lambda) := \frac{2\pi}{B_T T}\sum_{s\in\mathbb{Z}\setminus\{0\}} W\left(\frac{\lambda - 2\pi s/T}{B_T}\right) I^{(T)}_{N_jN_k}(2\pi s/T) \tag{18}$$

with a bandwidth $B_T \geq 0$ fulfilling $B_T \to 0$ as $T \to \infty$ and a weight function W which is of bounded variation, standardized, even and fulfills $W(\alpha) = 0$ if $|\alpha| > 0.5$.

As well-known examples for the weight function W we mention the rectangular kernel, the triangular kernel and the Epanechnikov kernel.

We recognize in Definition 4.6 the classical bias-variance trade-off of nonparametric statistics: The sum in (18) is taken over all $s \neq 0$ with $\frac{2\pi s}{T}$ lying in $[\lambda - \frac{B_T}{2}, \lambda + \frac{B_T}{2}]$. Thus, the larger B_T is, the more periodograms are considered for the estimation which reduces the variance. However, the larger B_T is, the more periodograms with frequencies $\frac{2\pi s}{T}$ are considered which show a significant distance to λ. This can produce a higher bias. A way out is to let the bandwidth B_T go to zero in a speed controlled manner by postulating that $B_T \to 0$ and $B_T T \to \infty$ as $T \to \infty$.

The smoothed periodogram given by (18) is stable and consistent. For further details and proofs for periodogram based estimators we refer to [9] (in the context of time series). There are various other ways of estimating the spectrum. We point to [4] for references and for a new approach to construct a shrinkage estimator with a smaller L_2-risk than the smoothed periodogram.

The next theorem shows that vectors of smoothed periodograms as defined in Definition 4.6 are asymptotically normal. For a proof based on the cumulant method see [7], Theorem 4.1.

Theorem 4.7. *Let* $\mathbf{N}$ *be a multivariate point process on* $\mathbb{R}$ *satisfying Assumption 4.2. Assume that for all* j,k, $|f''_{N_jN_k}(\lambda_i)| < \infty$ *for* $i = 1,2,\ldots,M$. *For* $B_T \to 0, B_T T \to \infty$ *and* $B_T^5 T \to 0$ *as* $T \to \infty$ *we have*

$$\sqrt{B_T T}\left[\begin{pmatrix} f^{(T)}_{N_{j_1}N_{k_1}}(\lambda_1) \\ \vdots \\ f^{(T)}_{N_{j_M}N_{k_M}}(\lambda_M)\end{pmatrix} - \begin{pmatrix} f_{N_{j_1}N_{k_1}}(\lambda_1) \\ \vdots \\ f_{N_{j_M}N_{k_M}}(\lambda_M)\end{pmatrix}\right] \xrightarrow{\mathcal{D}} \mathcal{N}^{\mathbb{C}}_M\Big(0, 2\pi\mathbf{\Sigma}\int W(\alpha)^2 d\alpha\Big),$$

where $\mathbf{\Sigma}$ *is an* $M \times M$ *matrix with*

$$\Sigma_{il} = \mathbf{1}_{\{\lambda_i=\lambda_l\}} f_{N_{j_i}N_{j_l}}(\lambda_i)\, f_{N_{k_i}N_{k_l}}(-\lambda_i) + \mathbf{1}_{\{\lambda_i=-\lambda_l\}} f_{N_{j_i}N_{k_l}}(\lambda_i)\, f_{N_{k_i}N_{j_l}}(-\lambda_i)\,.$$

For simplifying computations with matrices and vectors we will use the following two notations: For an $m \times n$ matrix $\mathbf{B} = (b_{ij})_{(i,j)\in\{1,\dots,m\}\times\{1,\dots,n\}}$ with complex-valued entries we define the column vector $vec\,\mathbf{B} \in \mathbb{C}^{m\cdot n}$ as

$$vec\,\mathbf{B} := (b_{11}, \dots, b_{m1}, b_{12}, \dots, b_{m2}, \dots, b_{1n}, \dots, b_{mn})^t.$$

For $\mathbf{B} = (b_{ij})_{(i,j)\in\{1,\dots,m\}\times\{1,\dots,n\}}$ and an $s \times t$ matrix $\mathbf{V}$, the $ms \times nt$ Kronecker matrix-product $\mathbf{B} \otimes \mathbf{V}$ is defined as

$$\mathbf{B} \otimes \mathbf{V} = (b_{ij} \cdot \mathbf{V})_{(i,j)\in\{1,\dots,m\}\times\{j=1,\dots,n\}}.$$

Useful properties of these two matrix notations are given in the following proposition, see [26].

Proposition 4.8. *Let the matrices* $\mathbf{B_1}, \mathbf{B_2}, \mathbf{B_3}, \mathbf{B_4}$ *and* $\mathbf{B_5}$ *have dimensions* $m \times n$, $n \times p$, $o \times u$, $p \times v$ *and* $u \times w$ *respectively. Let* $\mathbf{E}_j$ *denote the* $j \times j$ *identity matrix. Then*

$$\begin{aligned}
&(i)\ \ vec\,(\mathbf{B_1} \cdot \mathbf{B_2}) = (\mathbf{E}_p \otimes \mathbf{B_1})\ vec\,\mathbf{B_2} = (\mathbf{B_2}^t \otimes \mathbf{E}_m)\ vec\,\mathbf{B_1},\\
&(ii)\ \ (\mathbf{B_2} \otimes \mathbf{B_3}) \cdot (\mathbf{B_4} \otimes \mathbf{B_5}) = (\mathbf{B_2} \cdot \mathbf{B_4}) \otimes (\mathbf{B_3} \cdot \mathbf{B_5}),\\
&(iii)\ \ (\mathbf{B_1} \otimes \mathbf{B_3})^t = \mathbf{B_1}^t \otimes \mathbf{B_3}^t.
\end{aligned}$$

To estimate the transfer function $\mathbf{H}$ of the model (10) at frequency $\lambda \in \mathbb{R}$ we use the plug-in estimator

$$\mathbf{H}^{(T)}(\lambda) = f^{(T)}_{\mathbf{DA}}(\lambda)\, f^{(T)}_{\mathbf{AA}}(\lambda)^{-1} \tag{19}$$

with second-order spectral estimators $f^{(T)}_{N_iN_j}(\lambda)$, $i, j \in \{1, 2, n+m\}$, constructed according to (18).

We proceed with a technical result which is essential for the proofs of the two main results in this section:

Lemma 4.9. *With the notations from above the following decomposition holds:*

$$\begin{aligned}
&vec\left(\mathbf{H}^{(T)}(\lambda) - \mathbf{H}(\lambda)\right)\\
&= vec\left(\mathbf{H}(\lambda)\left[f_{\mathbf{AA}}(\lambda) - f^{(T)}_{\mathbf{AA}}(\lambda)\right] f_{\mathbf{AA}}(\lambda)^{-1}\right)\\
&\quad + vec\left(\left[f^{(T)}_{\mathbf{DA}}(\lambda) - f_{\mathbf{DA}}(\lambda)\right] f_{\mathbf{AA}}(\lambda)^{-1}\right)\\
&\quad + (\mathbf{E}_n \otimes \mathbf{H}(\lambda)) \cdot \left((f^{(T)}_{\mathbf{AA}}(-\lambda)^{-1} - f_{\mathbf{AA}}(-\lambda)^{-1}) \otimes \mathbf{E}_n\right)\\
&\quad \cdot vec\left(\left[f_{\mathbf{AA}}(\lambda) - f^{(T)}_{\mathbf{AA}}(\lambda)\right]\right)\\
&\quad + \left((f^{(T)}_{\mathbf{AA}}(-\lambda)^{-1} - f_{\mathbf{AA}}(-\lambda)^{-1}) \otimes \mathbf{E}_m\right) \cdot\ vec\left(\left[f^{(T)}_{\mathbf{DA}}(\lambda) - f_{\mathbf{DA}}(\lambda)\right)\right].
\end{aligned}$$

Proof. Using Proposition 4.8 we have:

$$
\begin{aligned}
&vec\left(\mathbf{H}^{(T)}(\lambda)-\mathbf{H}(\lambda)\right)\\
&= vec\Bigg(f_{\mathbf{DA}}(\lambda)\left[f^{(T)}_{\mathbf{AA}}(\lambda)^{-1}-f_{\mathbf{AA}}(\lambda)^{-1}\right]+\left[f^{(T)}_{\mathbf{DA}}(\lambda)-f_{\mathbf{DA}}(\lambda)\right]f^{(T)}_{\mathbf{AA}}(\lambda)^{-1}\Bigg)\\
&= vec\Bigg(f_{\mathbf{DA}}(\lambda)f_{\mathbf{AA}}(\lambda)^{-1}\left[f_{\mathbf{AA}}(\lambda)-f^{(T)}_{\mathbf{AA}}(\lambda)\right]f^{(T)}_{\mathbf{AA}}(\lambda)^{-1}\\
&\quad+\left[f^{(T)}_{\mathbf{DA}}(\lambda)-f_{\mathbf{DA}}(\lambda)\right]f^{(T)}_{\mathbf{AA}}(\lambda)^{-1}\Bigg)\\
&= vec\left(\mathbf{H}(\lambda)\left[f_{\mathbf{AA}}(\lambda)-f^{(T)}_{\mathbf{AA}}(\lambda)\right]f_{\mathbf{AA}}(\lambda)^{-1}\right)\\
&\quad+vec\left(\left[f^{(T)}_{\mathbf{DA}}(\lambda)-f_{\mathbf{DA}}(\lambda)\right]f_{\mathbf{AA}}(\lambda)^{-1}\right)\\
&\quad+vec\left(\mathbf{H}(\lambda)\left[f_{\mathbf{AA}}(\lambda)-f^{(T)}_{\mathbf{AA}}(\lambda)\right]\cdot\left[f^{(T)}_{\mathbf{AA}}(\lambda)^{-1}-f_{\mathbf{AA}}(\lambda)^{-1}\right]\right)\\
&\quad+vec\left(\left[f^{(T)}_{\mathbf{DA}}(\lambda)-f_{\mathbf{DA}}(\lambda)\right]\cdot\left[f^{(T)}_{\mathbf{AA}}(\lambda)^{-1}-f_{\mathbf{AA}}(\lambda)^{-1}\right]\right)\\
&= vec\left(\mathbf{H}(\lambda)\left[f_{\mathbf{AA}}(\lambda)-f^{(T)}_{\mathbf{AA}}(\lambda)\right]f_{\mathbf{AA}}(\lambda)^{-1}\right)\\
&\quad+vec\left(\left[f^{(T)}_{\mathbf{DA}}(\lambda)-f_{\mathbf{DA}}(\lambda)\right]f_{\mathbf{AA}}(\lambda)^{-1}\right)\\
&\quad+(\mathbf{E}_n\otimes\mathbf{H}(\lambda))\cdot\left((f^{(T)}_{\mathbf{AA}}(-\lambda)^{-1}-f_{\mathbf{AA}}(-\lambda)^{-1})\otimes\mathbf{E}_n\right)\\
&\quad\cdot vec\left(\left[f_{\mathbf{AA}}(\lambda)-f^{(T)}_{\mathbf{AA}}(\lambda)\right]\right)\\
&\quad+\left((f^{(T)}_{\mathbf{AA}}(-\lambda)^{-1}-f_{\mathbf{AA}}(-\lambda)^{-1})\otimes\mathbf{E}_m\right)\cdot\ vec\left(\left[f^{(T)}_{\mathbf{DA}}(\lambda)-f_{\mathbf{DA}}(\lambda)\right)\right].
\end{aligned}
$$

□

We now turn to the first main result in this section. It shows the asymptotic behavior of the estimator for the transfer function of the linear time-invariant model (10). For an analogue in the framework of time series we refer to [9], Theorem 8.8.1.

Theorem 4.10 (Asymptotic Normality for $\mathbf{H}^{(T)}(\lambda)$)**.** *Consider the linear time-invariant model* (10)*. Suppose* $\mathbf{N}=(\mathbf{A},\mathbf{D})^t$ *satisfies Assumption 4.2. Assume for* $\lambda\in\mathbb{R}\backslash\{0\}$ *that* $f_{\mathbf{AA}}(\lambda)$ *is not singular and* $|f''_{N_iN_j}(\lambda)|<\infty$ *for all* $i,j\in\{1,2,\ldots,n+m\}$*. Let* $B_T\to 0$*,* $B_TT\to\infty$ *and* $B_T^5T\to 0$ *as* $T\to\infty$*. Then, as* $T\to\infty$*, we have for the estimator of the transfer function* $\mathbf{H}$ *given by* (19)*, for* $\lambda\in\mathbb{R}\backslash\{0\}$*,*

$$
\begin{aligned}
&\sqrt{B_TT}\ vec\left(\mathbf{H}^{(T)}(\lambda)-\mathbf{H}(\lambda)\right)\\
&\qquad\xrightarrow{\mathcal{D}}\mathcal{N}^{\mathbb{C}}_{mn}\left(\mathbf{0},2\pi\int W(\alpha)^2d\alpha\left(f_{\mathbf{AA}}(-\lambda)^{-1}\otimes f_{\varepsilon\varepsilon}(\lambda)\right)\right).
\end{aligned}
$$

Proof. We divide the proof into 3 steps.

Step 1: Decomposition.

Lemma 4.9 implies:

$$\begin{aligned}
&\sqrt{B_T T}\ vec\left(\mathbf{H}^{(T)}(\lambda) - \mathbf{H}(\lambda)\right)\\
&= vec\left(\mathbf{H}(\lambda)\sqrt{B_T T}\left[f_{\mathbf{AA}}(\lambda) - f_{\mathbf{AA}}^{(T)}(\lambda)\right] f_{\mathbf{AA}}(\lambda)^{-1}\right)\\
&\quad + vec\left(\sqrt{B_T T}\left[f_{\mathbf{DA}}^{(T)}(\lambda) - f_{\mathbf{DA}}(\lambda)\right] f_{\mathbf{AA}}(\lambda)^{-1}\right)\\
&\quad + (\mathbf{E}_n \otimes \mathbf{H}(\lambda)) \cdot \left((f_{\mathbf{AA}}^{(T)}(-\lambda)^{-1} - f_{\mathbf{AA}}(-\lambda)^{-1}) \otimes \mathbf{E}_n\right)\\
&\quad \cdot vec\left(\sqrt{B_T T}\left[f_{\mathbf{AA}}(\lambda) - f_{\mathbf{AA}}^{(T)}(\lambda)\right]\right)\\
&\quad + \left((f_{\mathbf{AA}}^{(T)}(-\lambda)^{-1} - f_{\mathbf{AA}}(-\lambda)^{-1}) \otimes \mathbf{E}_m\right) \cdot\ vec\left(\sqrt{B_T T}\left[f_{\mathbf{DA}}^{(T)}(\lambda) - f_{\mathbf{DA}}(\lambda)\right)\right].
\end{aligned} \tag{20}$$

Step 2: Asymptotic normality.

First we show that all components of the last two summands converge to zero in probability: All components of the vectors $vec\big(\sqrt{B_T T}[f_{\mathbf{AA}}(\lambda) - f_{\mathbf{AA}}^{(T)}(\lambda)]\big)$ and $vec\big(\sqrt{B_T T}[f_{\mathbf{DA}}^{(T)}(\lambda) - f_{\mathbf{DA}}(\lambda)]\big)$ are asymptotically normally distributed according to Theorem 4.7. For the matrix $f_{\mathbf{AA}}^{(T)}(-\lambda)^{-1} - f_{\mathbf{AA}}(-\lambda)^{-1}$ the following inequality holds (Lemma of Dunford-Schwartz, see [31]):

$$\|f_{\mathbf{AA}}^{(T)}(-\lambda)^{-1} - f_{\mathbf{AA}}(-\lambda)^{-1}\| \leq \frac{\|f_{\mathbf{AA}}(-\lambda)^{-1}\|^2 \cdot \|f_{\mathbf{AA}}^{(T)}(-\lambda) - f_{\mathbf{AA}}(-\lambda)\|}{1 - \|f_{\mathbf{AA}}(-\lambda)^{-1}\| \cdot \|f_{\mathbf{AA}}^{(T)}(-\lambda) - f_{\mathbf{AA}}(-\lambda)\|} \tag{21}$$

for some norm $\|\cdot\|$. Since $\|f_{\mathbf{AA}}^{(T)}(-\lambda) - f_{\mathbf{AA}}(-\lambda)\|$ is of order $\mathrm{O}_p({B_T}^{-\frac{1}{2}} T^{-\frac{1}{2}})$ and $f_{\mathbf{AA}}^{(T)}(-\lambda) - f_{\mathbf{AA}}(-\lambda)$ componentwise converges to zero in probability, each component of the matrix $((f_{\mathbf{AA}}^{(T)}(-\lambda)^{-1} - f_{\mathbf{AA}}(-\lambda)^{-1}) \otimes \mathbf{E}_m)$ converges to zero in probability. Thus, each component of the last two summands in (20) is a finite sum of asymptotically normally distributed variates multiplicated with variates which converge in probability to zero. Then, by the Lemma of Slutsky all components of the two summands converge to zero in probability.

The limiting distribution of $\sqrt{B_T T}\ vec(\mathbf{H}^{(T)}(\lambda) - \mathbf{H}(\lambda))$ therefore is completely described by the first two summands in (20). This shows the asymptotic normality of $\sqrt{B_T T}\ vec(\mathbf{H}^{(T)}(\lambda) - \mathbf{H}(\lambda))$ due to Theorem 4.7. The asymptotic mean clearly is the zero vector because of the unbiasedness of $f_{\mathbf{DA}}^{(T)}(\lambda)$ and $f_{\mathbf{AA}}^{(T)}(\lambda)$.

Step 3: Asymptotic variance-covariance matrix.

Using Proposition 4.8 we compute the asymptotic variance-covariance matrix of $\sqrt{B_T T}\ vec\left(\mathbf{H}^{(T)}(\lambda) - \mathbf{H}(\lambda)\right)$ as follows:

$$
\begin{aligned}
&vec\left(-\mathbf{H}(\lambda)\sqrt{B_T\,T}\left(f^{(T)}_{\mathbf{AA}}(\lambda)-f_{\mathbf{AA}}(\lambda)\right)f_{\mathbf{AA}}(\lambda)^{-1}\right)\\
&\quad+vec\left(\sqrt{B_T\,T}\left(f^{(T)}_{\mathbf{DA}}(\lambda)-f_{\mathbf{DA}}(\lambda)\right)f_{\mathbf{AA}}(\lambda)^{-1}\right)\\
&=\left(f_{\mathbf{AA}}(-\lambda)^{-1}\otimes\mathbf{E}_m\right)\left[vec\left(-\mathbf{H}(\lambda)\sqrt{B_T\,T}\left(f^{(T)}_{\mathbf{AA}}(\lambda)-f_{\mathbf{AA}}(\lambda)\right)\right)\right.\\
&\quad\left.+vec\left(\sqrt{B_T\,T}\left(f^{(T)}_{\mathbf{DA}}(\lambda)-f_{\mathbf{DA}}(\lambda)\right)\right)\right]\\
&=\left(f_{\mathbf{AA}}(-\lambda)^{-1}\otimes\mathbf{E}_m\right)\cdot\left(-(\mathbf{E}_n\otimes\mathbf{H}(\lambda)),\quad\mathbf{E}_{mn}\right)\\
&\quad\cdot\begin{pmatrix}vec\left(\sqrt{B_T\,T}(f^{(T)}_{\mathbf{AA}}(\lambda)-f_{\mathbf{AA}}(\lambda))\right)\\ vec\left(\sqrt{B_T\,T}(f^{(T)}_{\mathbf{DA}}(\lambda)-f_{\mathbf{DA}}(\lambda))\right)\end{pmatrix}.
\end{aligned}
$$

This yields the asymptotic covariance matrix given by

$$
\begin{aligned}
&\left(f_{\mathbf{AA}}(-\lambda)^{-1}\otimes\mathbf{E}_m\right)\cdot\left(-(\mathbf{E}_n\otimes\mathbf{H}(\lambda)),\quad\mathbf{E}_{mn}\right)\cdot\mathcal{S}\cdot\begin{pmatrix}-(\mathbf{E}_n\otimes\overline{\mathbf{H}(\lambda)})^t\\ \mathbf{E}_{mn}\end{pmatrix}\\
&\cdot\left(\overline{f_{\mathbf{AA}}(\lambda)}^{-1}\otimes\mathbf{E}_m\right)
\end{aligned}
\tag{22}
$$

with (due to Theorem 4.7)

$$
\mathcal{S}=2\pi\int W(\alpha)^2d\alpha\cdot\begin{pmatrix}f_{\mathbf{AA}}(-\lambda)\otimes f_{\mathbf{AA}}(\lambda) & f_{\mathbf{AA}}(-\lambda)\otimes f_{\mathbf{AD}}(\lambda)\\ f_{\mathbf{AA}}(-\lambda)\otimes f_{\mathbf{DA}}(\lambda) & f_{\mathbf{AA}}(-\lambda)\otimes f_{\mathbf{DD}}(\lambda)\end{pmatrix}.
$$

Hence, (22) is equivalent to

$$
\begin{aligned}
&\left(-(f_{\mathbf{AA}}(-\lambda)^{-1}\otimes\mathbf{H}(\lambda)),\quad f_{\mathbf{AA}}(-\lambda)^{-1}\otimes\mathbf{E}_m\right)\cdot\mathcal{S}\cdot\begin{pmatrix}-(f_{\mathbf{AA}}(-\lambda)^{-1}\otimes\overline{\mathbf{H}(\lambda)}^t)\\ f_{\mathbf{AA}}(-\lambda)^{-1}\otimes\mathbf{E}_m\end{pmatrix}\\
&=2\pi\int W(\alpha)^2d\alpha\left(\mathbf{0}_{nm\times n^2},\quad\left(\mathbf{E}_n\otimes\left(f_{\mathbf{DD}}(\lambda)-f_{\mathbf{DA}}(\lambda)f_{\mathbf{AA}}(\lambda)^{-1}f_{\mathbf{AD}}(\lambda)\right)\right)\right)\\
&\qquad\cdot\begin{pmatrix}-(f_{\mathbf{AA}}(-\lambda)^{-1}\otimes\overline{\mathbf{H}(\lambda)}^t)\\ f_{\mathbf{AA}}(-\lambda)^{-1}\otimes\mathbf{E}_m\end{pmatrix}\\
&=2\pi\int W(\alpha)^2d\alpha\cdot\left(f_{\mathbf{AA}}(-\lambda)^{-1}\otimes\left(f_{\mathbf{DD}}(\lambda)-f_{\mathbf{DA}}(\lambda)f_{\mathbf{AA}}(\lambda)^{-1}f_{\mathbf{AD}}(\lambda)\right)\right)\\
&=2\pi\int W(\alpha)^2d\alpha\cdot\left(f_{\mathbf{AA}}(-\lambda)^{-1}\otimes f_{\varepsilon\varepsilon}(\lambda)\right),
\end{aligned}
$$

applying Theorem 4.4*(iii)* in the last line. □

In the linear time-invariant model it will often be desirable to have an estimate of the impulse-response function $\mathbf{h}(\cdot)$ (time domain), rather than of the transfer function $\mathbf{H}(\cdot)$ (frequency domain) only. This will e.g. be important for the statistical analysis of the stochastic tandem model in the next section. It

follows from the definition of $\mathbf{H}(\lambda)$ in (11) that

$$\mathbf{h}(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(i\lambda u)\mathbf{H}(\lambda)d\lambda.$$

For estimation of $\mathbf{h}(u)$ we follow the suggestion of [7] in the univariate case and use

$$\mathbf{h}^{(T)}(u) = \frac{1-cos(C_T u)}{\pi C_T u^2}\sum_{q=-Q_T}^{Q_T} \exp(iuC_T q)\,\mathbf{H}^{(T)}(C_T q), \quad u \in \mathbb{R}, \tag{23}$$

for small C_T and large Q_T. The factor $\frac{1-cos(C_T u)}{\pi C_T u^2}$ is inserted there and here in the same manner as in [5]. C_T plays the role of the bandwidth of the interpolation intervals. For the proof of the asymptotic normality of the estimator $\mathbf{h}^{(T)}(u)$ the crucial condition is the existence of a constant $L_1 < \infty$ such that $\sup_{\lambda\in\mathbb{R}}\|f_{\mathbf{AA}}(\lambda)^{-1}\|_\infty \le L_1$.

In the following let for a matrix $\mathbf{B}$, $\|\mathbf{B}\|_\infty$ denote the maximum absolute row sum norm which is in particular sub-multiplicative and compatible with the vector maximum norm $\|\cdot\|_\infty$.

Theorem 4.11 (Asymptotic Normality for $\mathbf{h}^{(T)}(u)$)**.** *Consider the linear time-invariant model* (10)*. Suppose that all assumptions of Theorem 4.10 are in force. Further, assume that* $\sup_{\lambda\in\mathbb{R}}\|f_{\mathbf{AA}}(\lambda)^{-1}\|_\infty \le L_1 < \infty, \int |u|^2\|\mathbf{h}(u)\|_\infty du < \infty$ *and* $\int |\lambda|^r\|\mathbf{H}(\lambda)\|_\infty d\lambda < \infty$ *for some* $r > 0$*. Moreover, let* $C_T \to 0, Q_T C_T \to \infty, Q_T^3 B_T^{-1} T^{-1} \to 0$*,* $C_T^4 Q_T B_T T \to 0$ *and* $C_T^{-2r-2} Q_T^{-2r-1} B_T T \to 0$ *as* $T \to \infty$*. Then, as* $T \to \infty$*, we have for the estimator of the impulse-response function* $\mathbf{h}$ *given by* (23)*, for* $u \in \mathbb{R}$*,*

$$\sqrt{C_T^{-2} Q_T^{-1} B_T T}\; vec\left(\mathbf{h}^{(T)}(u) - \mathbf{h}(u)\right) \xrightarrow{\mathcal{D}} \mathcal{N}_{mn}^{\mathbb{C}}\left(\mathbf{0}, 2\pi\int W(\alpha)^2 d\alpha\, \frac{1}{Q_T}\sum_{q=-Q_T}^{Q_T}\left(f_{\mathbf{AA}}(-C_T q)^{-1}\otimes f_{\varepsilon\varepsilon}(C_T q)\right)\right).$$

Note in particular that the asymptotic variance-covariance matrix is independent of the argument u.

Proof. We divide the proof into 4 steps.

Step 1: Decomposition of $vec\left(\mathbf{h}^{(T)}(u)\right)$.

It follows from Lemma 4.9 that for all $u \in \mathbb{R}$:

$$\begin{aligned} vec\left(\mathbf{h}^{(T)}(u)\right) &= \frac{1-cos(C_T u)}{\pi C_T u^2}\sum_{q=-Q_T}^{Q_T} \exp(iuC_T q)\, vec\left(\mathbf{H}(C_T q)\right) \\ &\quad + \frac{1-cos(C_T u)}{\pi C_T u^2}\mathbf{\Gamma}_T(u) \\ &\quad + \frac{1-cos(C_T u)}{\pi C_T u^2}\sum_{q=-Q_T}^{Q_T} \exp(iuC_T q)\,\mathbf{\Phi}_T(C_T q) =: T_1 + T_2 + T_3, \end{aligned}$$

where

$$\mathbf{\Gamma}_T(u) := \sum_{q=-Q_T}^{Q_T} \exp(iuC_Tq)\Bigg[vec\Bigg(\mathbf{H}(C_Tq)\left[f_{\mathbf{AA}}(C_Tq) - f_{\mathbf{AA}}^{(T)}(C_Tq)\right] \cdot f_{\mathbf{AA}}(C_Tq)^{-1}\Bigg) + vec\left(\left[f_{\mathbf{DA}}^{(T)}(C_Tq) - f_{\mathbf{DA}}(C_Tq)\right] f_{\mathbf{AA}}(C_Tq)^{-1}\right)\Bigg]$$

and

$$\begin{aligned}\mathbf{\Phi}_T(C_Tq) := {} & (\mathbf{E}_n \otimes \mathbf{H}(C_Tq)) \cdot \left((f_{\mathbf{AA}}^{(T)}(-C_Tq)^{-1} - f_{\mathbf{AA}}(-C_Tq)^{-1}) \otimes \mathbf{E}_n\right) \\ & \cdot vec\left(\left[f_{\mathbf{AA}}(C_Tq) - f_{\mathbf{AA}}^{(T)}(C_Tq)\right]\right) \\ & + \left((f_{\mathbf{AA}}^{(T)}(-C_Tq)^{-1} - f_{\mathbf{AA}}(-C_Tq)^{-1}) \otimes \mathbf{E}_m\right) \\ & \cdot\ vec\left(\left[f_{\mathbf{DA}}^{(T)}(C_Tq) - f_{\mathbf{DA}}(C_Tq)\right)\right]\end{aligned}$$

Step 2: Asymptotic behavior of term T_1.

We first show that $T_1 = vec(\mathbf{h}(u)) + (O(C_T^3Q_T) + O(C_T^{-r}Q_T^{-r}))\mathbf{1}_{mn}$.

Due to the assumptions we have for some $r > 0$ for each component j of $vec(\mathbf{H}(\lambda))$:

$$\begin{aligned}\infty > \int_{-\infty}^{\infty} |\lambda|^r \, |vec(\mathbf{H}(\lambda))_j| \, d\lambda \ &\geq \int_{-\infty}^{\infty} |\lambda|^r \, |vec(\mathbf{H}(\lambda))_j| \, \mathbf{1}_{\{|\lambda| \geq C_TQ_T\}} \, d\lambda \\ &\geq (C_TQ_T)^r \int_{-\infty}^{\infty} |vec(\mathbf{H}(\lambda))_j| \, \mathbf{1}_{\{|\lambda| \geq C_TQ_T\}} \, d\lambda \\ &= (C_TQ_T)^r \left(\int_{C_TQ_T}^{\infty} |vec(\mathbf{H}(\lambda))_j| \, d\lambda + \int_{-\infty}^{-C_TQ_T} |vec(\mathbf{H}(\lambda))_j| \, d\lambda\right).\end{aligned}$$

This implies with Result 2.2 in [5](page 120) (note that Bohman's term $\varphi''(0)$ here corresponds to $-\int u^2|vec(\mathbf{h}(u))_j|du$):

$$\begin{aligned}&\frac{1 - cos(C_Tu)}{\pi C_Tu^2} \sum_{q=-Q_T}^{Q_T} vec(\mathbf{H}(C_Tq))_j \, \exp(iuC_Tq) \\ &\leq \frac{1}{2\pi} \int_{-C_TQ_T}^{C_TQ_T} \exp(i\lambda u) \, vec(\mathbf{H}(\lambda))_j \, d\lambda + \frac{1}{12\pi} C_T^3Q_T \int u^2 |vec(\mathbf{h}(u))_j| \, du \\ &= vec(\mathbf{h}(u))_j - \frac{1}{2\pi}\left(\int_{C_TQ_T}^{\infty} \exp(i\lambda u) \, vec(\mathbf{H}(\lambda))_j \, d\lambda \right. \\ &\quad \left. + \int_{-\infty}^{-C_TQ_T} \exp(i\lambda u) \, vec(\mathbf{H}(\lambda))_j \, d\lambda\right) + O(C_T^3Q_T) \\ &\leq vec(\mathbf{h}(u))_j + \left(\int_{C_TQ_T}^{\infty} |vec(\mathbf{H}(\lambda))_j| \, d\lambda + \int_{-\infty}^{-C_TQ_T} |vec(\mathbf{H}(\lambda))_j| \, d\lambda\right) + O(C_T^3Q_T) \\ &\leq vec(\mathbf{h}(u))_j + O(C_T^{-r}Q_T^{-r}) + O(C_T^3Q_T).\end{aligned}$$

Thus, we have that $T_1 = vec(\mathbf{h}(u)) + \big(O(C_T^{-r}Q_T^{-r}) + O(C_T^3 Q_T)\big)\,\mathbf{1}_{mn}$. It remains to note that

$$\begin{aligned}\sqrt{C_T^{-2}Q_T^{-1}B_T T}\,\cdot\big(O(C_T^{-r}Q_T^{-r}) + O(C_T^3 Q_T)\big)\\ = O(C_T^{-2r-2}Q_T^{-2r-1}B_T T)^{0.5} + O(C_T^4 Q_T B_T T)^{0.5},\end{aligned}$$

which, by assumption, converges to 0 as $T \to \infty$.

Step 3: Asymptotic behavior of term T_3.
We first show that $\sup_{q=-Q_T}^{Q_T} \|\mathbf{\Phi}_T(C_T q)\|_\infty = o_P\big(Q_T^{2\epsilon} B_T^{-1} T^{-1}\big)$.
It holds:

$$\begin{aligned}&\sup_{q=-Q_T}^{Q_T} \|\mathbf{\Phi}_T(C_T q)\|_\infty\\ &\quad\le \|f_{\mathbf{DA}}(C_T q)\|_\infty \, \|f_{\mathbf{AA}}(C_T q)^{-1}\|_\infty \, \|f_{\mathbf{AA}}^{(T)}(-C_T q)^{-1} - f_{\mathbf{AA}}(-C_T q)^{-1}\|_\infty\\ &\qquad\cdot \|vec\left(f_{\mathbf{AA}}(C_T q) - f_{\mathbf{AA}}^{(T)}(C_T q)\right)\|_\infty\\ &\qquad + \|f_{\mathbf{AA}}^{(T)}(-C_T q)^{-1} - f_{\mathbf{AA}}(-C_T q)^{-1}\|_\infty \, \|vec\left(f_{\mathbf{DA}}^{(T)}(C_T q) - f_{\mathbf{DA}}(C_T q)\right)\|_\infty.\end{aligned}$$

We note that under the assumptions given, we have for any $\epsilon > 0$ as $T \to \infty$ that $\sup_{q=-Q_T}^{Q_T} |f_{N_i N_j}^{(T)}(C_T q) - f_{N_i N_j}(C_T q)| = o_P\big(Q_T^{\epsilon} B_T^{-1/2} T^{-1/2}\big)$ for any components N_i, N_j of the point process $\mathbf{N} = (\mathbf{A}, \mathbf{D})^t$ (see Theorem 4.2 in [7]). Using the Lemma of Dunford-Schwartz as in the proof of Theorem 4.10, see (21), we conclude:

$$\begin{aligned}\|f_{\mathbf{AA}}^{(T)}(-C_T q)^{-1} - f_{\mathbf{AA}}(-C_T q)^{-1}\|_\infty &\le \frac{L_1^2\, o_P\big(Q_T^{\epsilon} B_T^{-1/2} T^{-1/2}\big)}{1 - L_1\, o_P\big(Q_T^{\epsilon} B_T^{-1/2} T^{-1/2}\big)}\\ &= o_P\big(Q_T^{\epsilon} B_T^{-1/2} T^{-1/2}\big),\end{aligned}$$

since $\frac{1}{1+o_P(1)} = O_P(1)$. Further, the mixing conditions (Assumption 4.2) imply that there is a constant $K_1 < \infty$ such that $\sup_{\lambda\in\mathbb{R}} \|f_{\mathbf{DA}}(\lambda)\|_\infty \le K_1$. Thus, it follows for any $\epsilon > 0$:

$$\begin{aligned}\sup_{q=-Q_T}^{Q_T} \|\mathbf{\Phi}_T(C_T q)\|_\infty &\le K_1\, L_1\, o_P\big(Q_T^{\epsilon} B_T^{-1/2} T^{-1/2}\big)\, o_P\big(Q_T^{\epsilon} B_T^{-1/2} T^{-1/2}\big)\\ &\quad + o_P\big(Q_T^{\epsilon} B_T^{-1/2} T^{-1/2}\big)\, o_P\big(Q_T^{\epsilon} B_T^{-1/2} T^{-1/2}\big)\\ &= o_P\big(Q_T^{2\epsilon} B_T^{-1} T^{-1}\big).\end{aligned}$$

Further, this implies for T_3:

$$T_3 = \frac{1 - cos(C_T u)}{\pi C_T u^2} \sum_{q=-Q_T}^{Q_T} \exp(iuC_T q)\, \mathbf{\Phi}_T(C_T q) \le O(C_T Q_T) o_P\big(Q_T^{2\epsilon} B_T^{-1} T^{-1}\big)\mathbf{1}_{mn},$$

since (see [5] page 118)

$$\frac{1-\cos(C_T u)}{\pi C_T^2 u^2} = \frac{1}{2\pi}\int_{-1}^{1} \exp(-itC_T u)(1-|t|)dt \le \frac{1}{2\pi}\int_{-1}^{1} |\exp(-itC_T u)|dt \le \frac{1}{\pi}.$$

Choosing $\epsilon < 0.5$ we conclude for the asymptotic behavior of T_3:

$$\begin{aligned}\sqrt{C_T^{-2} Q_T^{-1} B_T T}\cdot T_3 &= \sqrt{C_T^{-2} Q_T^{-1} B_T T}\cdot o_P\left(C_T Q_T^2 B_T^{-1} T^{-1}\right)\mathbf{1}_{mn}\\ &= o_P(1)\cdot O(Q_T^{1.5} B_T^{-0.5} T^{-0.5})\,\mathbf{1}_{mn},\end{aligned}$$

which converges to 0 as $T\to\infty$ since by assumption then $Q_T^3 B_T^{-1} T^{-1}\to 0$.

Step 4: Asymptotic normality of term T_2.

The term $\mathbf{\Gamma}_T(u)$ determines the asymptotic behavior of $vec(\mathbf{h}^{(T)}(u))$. It directly follows from Theorem 4.7 that $\Gamma^{(T)}(u)$ is asymptotically normal of rate $\sqrt{B_T T}$ with zero mean vector. Applying Theorem 4.7, Proposition 4.8 and the fact that $f^t_{\mathbf{AA}}(\lambda) = f_{\mathbf{AA}}(-\lambda)$ for all λ, the asymptotic variance-covariance matrix of $\mathbf{\Gamma}_T(u)$ can be computed as follows. Note that in the computations below we have left out the term $2\pi\int W(\alpha)^2 d\alpha$ to reduce the length of the equations. Let asCov denote "asymptotic covariance".

$$\begin{aligned}
&\sum_{q_1=-Q_T}^{+Q_T}\sum_{q_2=-Q_T}^{+Q_T} \exp(iuC_T(q_1-q_2))\,B_T T\,\mathrm{asCov}\Big[\\
&\left(f_{\mathbf{AA}}(-C_T q_1)^{-1}\otimes \mathbf{E}_m\right)\mathrm{vec}\left(f^{(T)}_{\mathbf{DA}}(C_T q_1)\right)\\
&-\left(f_{\mathbf{AA}}(-C_T q_1)^{-1}\otimes \mathbf{H}(C_T q_1)\right)\mathrm{vec}\left(f^{(T)}_{\mathbf{AA}}(C_T q_1)\right),\\
&\left(f_{\mathbf{AA}}(-C_T q_2)^{-1}\otimes \mathbf{E}_m\right) vec\left(f^{(T)}_{\mathbf{DA}}(C_T q_2)\right)\\
&-\left(f_{\mathbf{AA}}(-C_T q_2)^{-1}\otimes \mathbf{H}(C_T q_2)\right)\mathrm{vec}\left(f^{(T)}_{\mathbf{AA}}(C_T q_2)\right)\Big]\\
=&\sum_{q_1=-Q_T}^{+Q_T}\sum_{q_2=-Q_T}^{+Q_T} \exp(iuC_T(q_1-q_2))B_T T\Big[\\
&\left(f_{\mathbf{AA}}(-C_T q_1)^{-1}\otimes \mathbf{E}_m\right)\cdot \mathrm{asCov}\left(vec\left(f^{(T)}_{\mathbf{DA}}(C_T q_1)\right), vec\left(f^{(T)}_{\mathbf{DA}}(C_T q_2)\right)\right)\\
&\qquad\cdot\left(f_{\mathbf{AA}}(-C_T q_2)^{-1}\otimes \mathbf{E}_m\right)\\
&-\left(f_{\mathbf{AA}}(-C_T q_1)^{-1}\otimes \mathbf{E}_m\right)\cdot \mathrm{asCov}\left(vec\left(f^{(T)}_{\mathbf{DA}}(C_T q_1)\right), vec\left(f^{(T)}_{\mathbf{AA}}(C_T q_2)\right)\right)\\
&\qquad\cdot\left(f_{\mathbf{AA}}(-C_T q_2)^{-1}\otimes \overline{\mathbf{H}(C_T q_2)}^t\right)\\
&-\left(f_{\mathbf{AA}}(-C_T q_1)^{-1}\otimes \mathbf{H}(C_T q_1)\right)\cdot \mathrm{asCov}\left(vec\left(f^{(T)}_{\mathbf{AA}}(C_T q_1)\right), vec\left(f^{(T)}_{\mathbf{DA}}(C_T q_2)\right)\right)\\
&\qquad\cdot\left(f_{\mathbf{AA}}(-C_T q_2)^{-1}\otimes \mathbf{E}_m\right)\\
&+\left(f_{\mathbf{AA}}(-C_T q_1)^{-1}\otimes \mathbf{H}(C_T q_1)\right)\cdot \mathrm{asCov}\left(vec\left(f^{(T)}_{\mathbf{AA}}(C_T q_1)\right), vec\left(f^{(T)}_{\mathbf{AA}}(C_T q_2)\right)\right)\\
&\qquad\cdot\left(f_{\mathbf{AA}}(-C_T q_2)^{-1}\otimes \overline{\mathbf{H}(C_T q_2)}^t\right)\Big]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{q_1=-Q_T}^{+Q_T} \sum_{q_2=-Q_T}^{+Q_T} \exp(iuC_T(q_1-q_2))\,\mathbf{1}_{q_1=q_2} \Big\{ \\
&\quad \big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes\mathbf{E}_m\big)\cdot\big(f_{\mathbf{AA}}(-C_Tq_1)\otimes f_{\mathbf{DD}}(C_Tq_1)\big)\cdot\big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes\mathbf{E}_m\big) \\
&\quad - \big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes\mathbf{E}_m\big)\cdot\big(f_{\mathbf{AA}}(-C_Tq_1)\otimes f_{\mathbf{DA}}(C_Tq_1)\big) \\
&\qquad \cdot\big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes f_{\mathbf{AA}}(C_Tq_1)^{-1}f_{\mathbf{AD}}(C_Tq_1)\big) \\
&\quad - \big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes\mathbf{H}(C_Tq_1)\big)\cdot\big(f_{\mathbf{AA}}(-C_Tq_1)\otimes f_{\mathbf{AD}}(C_Tq_1)\big) \\
&\qquad \cdot\big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes\mathbf{E}_m\big) \\
&\quad + \big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes\mathbf{H}(C_Tq_1)\big)\cdot\big(f_{\mathbf{AA}}(-C_Tq_1)\otimes f_{\mathbf{AA}}(C_Tq_1)\big) \\
&\qquad \cdot\Big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes\overline{\mathbf{H}(C_Tq_1)}^{t}\Big)\Big\} \\
&\quad + \sum_{q_1=-Q_T}^{+Q_T} \sum_{q_2=-Q_T}^{+Q_T} \exp(iuC_T(q_1-q_2))\,\mathbf{1}_{q_1=-q_2} \Big\{ \\
&\quad \big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes\mathbf{E}_m\big) \\
&\qquad \cdot\big(f_{\mathbf{AD}}(-C_Tq_1)\otimes f_{\mathbf{D}A_1}(C_Tq_1),\ldots,f_{\mathbf{AD}}(-C_Tq_1)\otimes f_{\mathbf{D}A_n}(C_Tq_1)\big) \\
&\qquad \cdot\big(f_{\mathbf{AA}}(C_Tq_1)^{-1}\otimes\mathbf{E}_m\big) \\
&\quad - \big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes\mathbf{E}_m\big) \\
&\qquad \cdot\big(f_{\mathbf{AA}}(-C_Tq_1)\otimes f_{\mathbf{D}A_1}(C_Tq_1),\ldots,f_{\mathbf{AA}}(-C_Tq_1)\otimes f_{\mathbf{D}A_n}(C_Tq_1)\big) \\
&\qquad \cdot\big(f_{\mathbf{AA}}(C_Tq_1)^{-1}\otimes f_{\mathbf{AA}}(-C_Tq_1)^{-1}f_{\mathbf{AD}}(-C_Tq_1)\big) \\
&\quad - \big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes\mathbf{H}(C_Tq_1)\big) \\
&\qquad \cdot\big(f_{\mathbf{AD}}(-C_Tq_1)\otimes f_{\mathbf{A}A_1}(C_Tq_1),\ldots,f_{\mathbf{AD}}(-C_Tq_1)\otimes f_{\mathbf{A}A_n}(C_Tq_1)\big) \\
&\qquad \cdot\big(f_{\mathbf{AA}}(C_Tq_1)^{-1}\otimes\mathbf{E}_m\big) \\
&\quad + \big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes\mathbf{H}(C_Tq_1)\big) \\
&\qquad \cdot\big(f_{\mathbf{AA}}(-C_Tq_1)\otimes f_{\mathbf{A}A_1}(C_Tq_1),\ldots,f_{\mathbf{AA}}(-C_Tq_1)\otimes f_{\mathbf{A}A_n}(C_Tq_1)\big) \\
&\qquad \cdot\big(f_{\mathbf{AA}}(C_Tq_1)^{-1}\otimes f_{\mathbf{AA}}(-C_Tq_1)^{-1}f_{\mathbf{AD}}(-C_Tq_1)\big)\Big\} \\
&= \sum_{q_1=-Q_T}^{+Q_T} \sum_{q_2=-Q_T}^{+Q_T} \exp(iuC_T(q_1-q_2))\,\mathbf{1}_{q_1=q_2} \Big\{ \\
&\quad \big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes f_{\mathbf{DD}}(C_Tq_1)\big) \\
&\quad - 2\,\big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes f_{\mathbf{DA}}(C_Tq_1)f_{\mathbf{AA}}(C_Tq_1)^{-1}f_{\mathbf{AD}}(C_Tq_1)\big) \\
&\quad + \big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}\otimes f_{\mathbf{DA}}(C_Tq_1)f_{\mathbf{AA}}(C_Tq_1)^{-1}f_{\mathbf{AD}}(C_Tq_1)\big)\Big\} \\
&\quad + \sum_{q_1=-Q_T}^{+Q_T} \sum_{q_2=-Q_T}^{+Q_T} \exp(iuC_T(q_1-q_2))\,\mathbf{1}_{q_1=-q_2} \Big\{ \\
&\quad \big(f_{\mathbf{AA}}(-C_Tq_1)^{-1}f_{\mathbf{AD}}(-C_Tq_1)\otimes f_{\mathbf{D}A_1}(C_Tq_1),\ldots, \\
&\qquad f_{\mathbf{AA}}(-C_Tq_1)^{-1}f_{\mathbf{AD}}(-C_Tq_1)\otimes f_{\mathbf{D}A_n}(C_Tq_1)\big)\cdot\big(f_{\mathbf{AA}}(C_Tq_1)^{-1}\otimes\mathbf{E}_m\big)
\end{aligned}
$$

$$
\begin{aligned}
&- \big(\mathbf{E}_n \otimes f_{\mathbf{D}A_1}(C_T q_1), \dots, \mathbf{E}_n \otimes f_{\mathbf{D}A_n}(C_T q_1)\big) \\
&\qquad \cdot \big(f_{\mathbf{AA}}(C_T q_1)^{-1} \otimes f_{\mathbf{AA}}(-C_T q_1)^{-1} f_{\mathbf{AD}}(-C_T q_1)\big) \\
&- \big(f_{\mathbf{AA}}(-C_T q_1)^{-1} f_{\mathbf{AD}}(-C_T q_1) \otimes f_{\mathbf{D}A_1}(C_T q_1), \dots, \\
&\qquad f_{\mathbf{AA}}(-C_T q_1)^{-1} f_{\mathbf{AD}}(-C_T q_1) \otimes f_{\mathbf{D}A_n}(C_T q_1)\big) \cdot \big(f_{\mathbf{AA}}(C_T q_1)^{-1} \otimes \mathbf{E}_m\big) \\
&+ \big(\mathbf{E}_n \otimes f_{\mathbf{D}A_1}(C_T q_1), \dots, \mathbf{E}_n \otimes f_{\mathbf{D}A_n}(C_T q_1)\big) \\
&\qquad \cdot \big(f_{\mathbf{AA}}(C_T q_1)^{-1} \otimes f_{\mathbf{AA}}(-C_T q_1)^{-1} f_{\mathbf{AD}}(-C_T q_1)\big) \Big\} \\
= &\sum_{q_1=-Q_T}^{+Q_T} \sum_{q_2=-Q_T}^{+Q_T} \exp(iuC_T(q_1-q_2))\, \mathbf{1}_{q_1=q_2} \Big(f_{\mathbf{AA}}(-C_T q_1)^{-1} \otimes f_{\varepsilon\varepsilon}(C_T q_1)\Big) \\
= &\sum_{q_1=-Q_T}^{+Q_T} \Big(f_{\mathbf{AA}}(-C_T q_1)^{-1} \otimes f_{\varepsilon\varepsilon}(C_T q_1)\Big).
\end{aligned}
$$

Using again that $\frac{1-\cos(C_T u)}{\pi C_T u^2} = O(C_T)$, the proof is complete. □

5. Nonparametric estimation for the tandem model of two nodes

In this section we apply the results of the preceding section to the estimation problem in the tandem network model (Figure 1) given by (5) and (6) as discussed in Section 2. We will in particular construct estimators for the Fourier transforms G_1 and G_2, for the routing probability p_1, for the densities g_1 and g_2 and for the multiple coherence of the model. For all estimators we will prove their respective asymptotic normality under appropriate conditions.

We see that the tandem model is represented by the linear time-invariant model stated in (10) with $\mu = 0$, the matrix $\mathbf{h}$ given by (6), and $\mathbf{A}$ and $\mathbf{D}$ being two-dimensional stationary point processes. Due to Remark 4.5 the unknown parameters $G_1(\lambda), G_2(\lambda)$ and p_1 are identifiable provided that the spectral matrix $f_{\mathbf{AA}}(\lambda)$ is invertible, $\lambda \in \mathbb{R}$.

Since by assumption the arrival processes A_1 and A_2 are stochastically independent of each other, $f_{A_1A_2}(\lambda) = 0$ for all $\lambda \in \mathbb{R}$, which yields the spectral matrix $f_{\mathbf{AA}}(\lambda)$, $\lambda \in \mathbb{R}$, to be given by:

$$
f_{\mathbf{AA}}(\lambda) = \begin{pmatrix} f_{A_1A_1}(\lambda) & 0 \\ 0 & f_{A_2A_2}(\lambda) \end{pmatrix}, \ \lambda \in \mathbb{R}.
$$

Furthermore, the arrival process A_2 at the second node has no influence on the external departure process D_1 of the first node and vice versa, thus $f_{D_1A_2}(\lambda) = 0$ for all $\lambda \in \mathbb{R}$. This implies:

$$
\begin{aligned}
\mathbf{H}(\lambda) &= \begin{pmatrix} H_{11}(\lambda) & H_{12}(\lambda) \\ H_{21}(\lambda) & H_{22}(\lambda) \end{pmatrix} = f_{\mathbf{DA}}(\lambda) \cdot (f_{\mathbf{AA}}(\lambda))^{-1} \\
&= \begin{pmatrix} f_{D_1A_1}(\lambda) & 0 \\ f_{D_2A_1}(\lambda) & f_{D_2A_2}(\lambda) \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{f_{A_1A_1}(\lambda)} & 0 \\ 0 & \frac{1}{f_{A_2A_2}(\lambda)} \end{pmatrix} = \begin{pmatrix} \frac{f_{D_1A_1}(\lambda)}{f_{A_1A_1}(\lambda)} & 0 \\ \frac{f_{D_2A_1}(\lambda)}{f_{A_1A_1}(\lambda)} & \frac{f_{D_2A_2}(\lambda)}{f_{A_2A_2}(\lambda)} \end{pmatrix}.
\end{aligned} \tag{24}
$$

Estimation of the Fourier transforms $G_1(\cdot)$ and $G_2(\cdot)$

A comparison of (24) with (7) yields for the Fourier transform $G_2(\cdot)$ of the service density at the second node

$$G_2(\lambda) = H_{22}(\lambda) = \frac{f_{D2A2}(\lambda)}{f_{A2A2}(\lambda)} \quad \text{for } \lambda \in \mathbb{R}, \tag{25}$$

and for the Fourier transform $G_1(\cdot)$ of the service density at the first node

$$\begin{aligned} G_1(\lambda) &= H_{11}(\lambda) + H_{21}(\lambda) \cdot (H_{22}(\lambda))^{-1} \\ &= \frac{f_{D1A1}(\lambda)}{f_{A1A1}(\lambda)} + \frac{f_{D2A1}(\lambda)}{f_{A1A1}(\lambda)} \cdot \frac{f_{A2A2}(\lambda)}{f_{D2A2}(\lambda)} \quad \text{for } \lambda \in \mathbb{R}. \end{aligned} \tag{26}$$

Remark 5.1. Considering the spectrum as a measure of coherence between the respective processes the forms of $G_1(\cdot)$ and $G_2(\cdot)$ can be explained in an intuitive way: $G_2(\lambda)$ describes at frequency λ the coherence between the external input process A_2 and the external output process D_2 at node 2 after subtraction of the inner coherence of the arrival process A_2. $G_1(\cdot)$ can be interpreted similarly.

In analogy to the last section we construct estimators of $G_1(\cdot)$ and $G_2(\cdot)$ based on smoothed periodograms as estimators for the spectra according to (18):

$$\text{estimator for } G_2(\lambda): \quad G_2^{(T)}(\lambda) = \frac{f_{D2A2}^{(T)}(\lambda)}{f_{A2A2}^{(T)}(\lambda)}, \quad \lambda \in \mathbb{R}, \tag{27}$$

$$\text{estimator for } G_1(\lambda): \quad G_1^{(T)}(\lambda) = \frac{f_{D1A1}^{(T)}(\lambda)}{f_{A1A1}^{(T)}(\lambda)} + \frac{f_{D2A1}^{(T)}(\lambda)}{f_{A1A1}^{(T)}(\lambda)} \cdot \frac{f_{A2A2}^{(T)}(\lambda)}{f_{D2A2}^{(T)}(\lambda)}, \quad \lambda \in \mathbb{R}. \tag{28}$$

Theorem 5.2 (Asymptotic Normality for $G_1^{(T)}$ and $G_2^{(T)}$)**.** *Consider the tandem model of queues given by (5) and (6) as described in Section 2. Assume that mean service times at nodes 1 and 2 are finite, i.e., $\int_0^\infty t\, g_i(t)dt < \infty$ for $i = 1, 2$. Assume further that the arrival processes A_1 and A_2 are stationary, show existing moments of all orders and satisfy the mixing conditions on their reduced cumulant measures as stated in Assumption 4.2. Then the 4-dimensional point process $\mathbf{N} = (\mathbf{A}, \mathbf{D})^t$ on $\mathbb{R}$ satisfies Assumption 4.2. Under the conditions that for $\lambda \neq 0$, $f_{A_1A_1}(\lambda) \neq 0$ and $f_{A_2A_2}(\lambda) \neq 0$, $|f_{D2A2}(\lambda)|^2 \neq 0$ and $|f''_{N_iN_j}(\lambda)| < \infty$ for all $i, j = 1, 2, 3, 4$, the asymptotic behavior of the estimators $G_1^{(T)}(\lambda)$ and $G_2^{(T)}(\lambda)$ given by (27) and (28) for $T \to \infty$ is then characterized by*

$$\sqrt{B_T T}\left(\begin{pmatrix} G_1^{(T)}(\lambda) \\ G_2^{(T)}(\lambda) \end{pmatrix} - \begin{pmatrix} G_1(\lambda) \\ G_2(\lambda) \end{pmatrix}\right) \xrightarrow{\mathcal{D}} \mathcal{N}_2^{\mathbb{C}}\left(0, 2\pi \int W(\alpha)^2 d\alpha \cdot \mathbf{\Sigma}(\lambda)\right)$$

for a sequence of bandwidths $(B_T)_{T\geq 0}$ satisfying $B_T \to 0, B_T T \to \infty$ and $B_T^5 T \to 0$ as $T \to \infty$. The asymptotic variance-covariance matrix is given by

$$\Sigma_{11}(\lambda) = \frac{f_{\varepsilon_1\varepsilon_1}(\lambda)}{f_{A_1A_1}(\lambda)} + \frac{f_{\varepsilon_2\varepsilon_1}(\lambda)+f_{\varepsilon_1\varepsilon_2}(\lambda)}{f_{A_1A_1}(\lambda)\,H_{22}(\lambda)} + \frac{f_{\varepsilon_2\varepsilon_2}(\lambda)}{f_{A_1A_1}(\lambda)H_{22}^2(\lambda)} + \frac{f_{\varepsilon_2\varepsilon_2}(\lambda)}{f_{A_2A_2}(\lambda)}\frac{H_{21}^2(\lambda)}{H_{22}^4(\lambda)},$$
$$\Sigma_{12}(\lambda) = \Sigma_{21}(\lambda) = -\frac{f_{\varepsilon_2\varepsilon_2}(\lambda)}{f_{A_2A_2}(\lambda)}\frac{H_{21}(\lambda)}{H_{22}^2(\lambda)} \quad \textit{and} \quad \Sigma_{22}(\lambda) = \frac{f_{\varepsilon_2\varepsilon_2}(\lambda)}{f_{A_2A_2}(\lambda)}.$$

Proof. In analogy to the proof of Theorem 3.1 in [7] it follows that the point process $\mathbf{N} = (\mathbf{A}, \mathbf{D})^t$ on $\mathbb{R}$ fulfills Assumption 4.2 since the mean sojourn time in the network is finite. By the delta method (see e.g. [32], Theorem 3.1) asymptotic normality follows from Theorem 4.10 with the fact that

$$\begin{aligned}\phi(vec(\mathbf{H}(\lambda))) &= \phi((H_{11}(\lambda), H_{21}(\lambda), H_{12}(\lambda), H_{22}(\lambda))^t)\\ &= (H_{11}(\lambda) + H_{21}(\lambda)\cdot H_{22}(\lambda)^{-1}, H_{22}(\lambda))^t = (G_1(\lambda), G_2(\lambda))^t\end{aligned}$$

with the differentiable function $\phi : \mathbb{C}^4\backslash(\mathbb{C}^3 \times \{0\}) \to \mathbb{C}^2$ given by $\phi(x_1, x_2, x_3, x_4)^t = (x_1 + x_2 \cdot x_4^{-1}, x_4)^t$. The total derivative of the function ϕ at $vec(\mathbf{H}(\lambda))$ is given by

$$\phi^{'}(vec(\mathbf{H}(\lambda))) = \begin{pmatrix} 1 & \frac{1}{H_{22}(\lambda)} & 0 & -\frac{H_{21}(\lambda)}{H_{22}^2(\lambda)} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The asymptotic variance-covariance matrix can then be computed by

$$2\pi \int W(\alpha)^2 d\alpha \cdot \phi^{'}(vec(\mathbf{H}(\lambda))) \cdot \left(f_{\mathbf{AA}}(-\lambda)^{-1} \otimes f_{\varepsilon\varepsilon}(\lambda)\right) \cdot \phi^{'}(vec(\mathbf{H}(\lambda)))^t$$

which yields the stated terms. □

Remark 5.3. The importance of Theorem 5.2 for practical applications lies in particular in the possible construction of approximate confidence intervals for the parameters of interest. For example, under the conditions of Theorem 5.2, an approximate $(1-\alpha)$-confidence interval for $\mathrm{Re}(G_1(\lambda))$ is given by

$$\left[\mathrm{Re}(G_1^{(T)}(\lambda)) - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\pi\int W(\alpha)^2 d\alpha\,\mathrm{Re}(\Sigma_{11}(\lambda))}{B_T T}},\right.$$
$$\left.\mathrm{Re}(G_1^{(T)}(\lambda)) + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\pi\int W(\alpha)^2 d\alpha\,\mathrm{Re}(\Sigma_{11}(\lambda))}{B_T T}}\right].$$

Here $u_{1-\frac{\alpha}{2}}$ denotes the $(1-\frac{\alpha}{2})$-quantile of the real-valued standard normal distribution. Clearly, the goodness of the approximation increases as $T \to \infty$.

Example 5.4. Consider the case that the external arrival processes A_1 and A_2 are Poisson processes with constant intensities κ_1 and κ_2, i.e., the tandem system under study consists of two nodes of type M/G/∞. It follows then for the reduced covariance measures $\breve{C}_{A_iA_i}(du) = \kappa_i\delta_0(du)$ and thus the power spectra are given by $f_{A_iA_i}(\lambda) \equiv \frac{\kappa_i}{2\pi}$, $\lambda \in \mathbb{R}$, (see [11], Example 8.2(a)). Since the intensities κ_i can directly be estimated by the observed arrival times of the processes A_i, the estimation of the power spectra $f_{A_iA_i}(\lambda)$ is here straightforward without use of the periodograms, $i = 1, 2$. However, the estimation of the cross

spectra (like $f_{D_1A_1}(\lambda)$) cannot be simplified, although the departure processes D_1 and D_2 are also Poisson processes with constant intensities (for a proof see [1], Theorem 1.3 in Chapter IV.1).

Estimation of the routing probability p_1

Now we turn to the estimation of the routing probability p_1 in the tandem network from Figure 1 in steady state. We first show how we can use the developed spectral theory to derive an asymptotically normal estimator. Equation (7) directly yields an equation for the routing probability p_1 in terms of the components of the transfer function $\mathbf{H}(\cdot)$:

$$p_1 = \frac{H_{21}(\lambda)}{H_{21}(\lambda) + H_{11}(\lambda)H_{22}(\lambda)},\ \lambda \in \mathbb{R}.$$

It is interesting to see that the right hand side thus is independent of λ. However, an estimator of the form

$$\widehat{p}_1^{(T)} = \widehat{p}_1^{(T)}(\lambda) = \frac{H_{21}^{(T)}(\lambda)}{H_{21}^{(T)}(\lambda) + H_{11}^{(T)}(\lambda)H_{22}^{(T)}(\lambda)},\ \lambda \in \mathbb{R}, \tag{29}$$

formally depends on the frequency λ under consideration. Therefore, we make a smoothing approach and define our estimator for p_1 by the minimum of the quadratic loss, i.e., $p_1^{(T)}$ shall minimize for a finite set $\mathcal{L} \subset \mathbb{R}\backslash\{0\}$ of λ values

$$\sum_{\lambda\in\mathcal{L}} \left| H_{11}^{(T)}(\lambda) - G_1^{(T)}(\lambda)\,(1-p_1)\right|^2 + \sum_{\lambda\in\mathcal{L}} \left| \frac{H_{21}^{(T)}(\lambda)}{H_{22}^{(T)}(\lambda)} - G_1^{(T)}(\lambda)\,p_1\right|^2. \tag{30}$$

Taking the derivative of (30) with respect to p_1 yields:

$$\begin{aligned}
&2\sum_{\lambda\in\mathcal{L}} \Bigg[\mathrm{Re}\left(H_{11}^{(T)}(\lambda) - G_1^{(T)}(\lambda)(1-p_1)\right)\cdot \mathrm{Re}(G_1^{(T)}(\lambda)) \\
&\qquad + \mathrm{Im}\left(H_{11}^{(T)}(\lambda) - G_1^{(T)}(\lambda)(1-p_1)\right)\cdot \mathrm{Im}(G_1^{(T)}(\lambda))\Bigg] \\
&-2\sum_{\lambda\in\mathcal{L}} \Bigg[\mathrm{Re}\left(\frac{H_{21}^{(T)}(\lambda)}{H_{22}^{(T)}(\lambda)} - G_1^{(T)}(\lambda)\cdot p_1\right)\cdot \mathrm{Re}(G_1^{(T)}(\lambda)) \\
&\qquad + \mathrm{Im}\left(\frac{H_{21}^{(T)}(\lambda)}{H_{22}^{(T)}(\lambda)} - G_1^{(T)}(\lambda)\cdot p_1\right)\cdot \mathrm{Im}(G_1^{(T)}(\lambda))\Bigg] \\
&= 2\sum_{\lambda\in\mathcal{L}} \Bigg[\mathrm{Re}\left(H_{11}^{(T)}(\lambda)\cdot G_1^{(T)}(-\lambda)\right) \\
&\qquad - |G_1^{(T)}(\lambda)|^2 - \mathrm{Re}\left(\frac{H_{21}^{(T)}(\lambda)}{H_{22}^{(T)}(\lambda)}\cdot G_1^{(T)}(-\lambda)\right) + 2\,p_1\,|G_1^{(T)}(\lambda)|^2\Bigg]
\end{aligned}$$

$$= 2\sum_{\lambda\in\mathcal{L}}\left[\mathrm{Re}\left(\left(H_{11}^{(T)}(\lambda) - \frac{H_{21}^{(T)}(\lambda)}{H_{22}^{(T)}(\lambda)}\right)\cdot G_1^{(T)}(-\lambda)\right) - |G_1^{(T)}(\lambda)|^2 + 2\,p_1|G_1^{(T)}(\lambda)|^2\right]$$
$$= 4\sum_{\lambda\in\mathcal{L}}\left[-|G_1^{(T)}(\lambda)|^2 + \mathrm{Re}\left(H_{11}^{(T)}(\lambda)\cdot G_1^{(T)}(-\lambda)\right) + p_1\,|G_1^{(T)}(\lambda)|^2\right],$$

where the last equality is due to $H_{11}^{(T)}(\lambda) - \frac{H_{21}^{(T)}(\lambda)}{H_{22}^{(T)}(\lambda)} = -G_1^{(T)}(\lambda) + 2H_{11}^{(T)}(\lambda)$. The second derivative of the quadratic loss (30) with respect to p_1 is positive and thus, the optimal $p_1^{(T)}$ is given by

$$p_1^{(T)} = 1 - \frac{\sum_{\lambda\in\mathcal{L}}\mathrm{Re}\Big(G_1^{(T)}(\lambda)\cdot\frac{f_{D_1A_1}^{(T)}(-\lambda)}{f_{A_1A_1}^{(T)}(-\lambda)}\Big)}{\sum_{\lambda\in\mathcal{L}}|G_1^{(T)}(\lambda)|^2} \tag{31}$$

with $G_1^{(T)}(\cdot)$ from (28) and $f_{A_1A_1}^{(T)}(\cdot)$ and $f_{D_1A_1}^{(T)}(\cdot)$ constructed according to (18). Note that $p_1^{(T)}$ is by definition real valued. Since for all $\lambda\in\mathcal{L}$

$$\begin{aligned}|G_1(\lambda)|^2 - \mathrm{Re}\big(G_1(\lambda)\cdot H_{11}(-\lambda)\big) &= |G_1(\lambda)|^2 - \mathrm{Re}\big(G_1(\lambda)\cdot G_1(-\lambda)\big)(1-p_1)\\ &= p_1\cdot|G_1(\lambda)|^2,\end{aligned}$$

the estimator $p_1^{(T)}$ indeed quantifies the underlying true routing behavior of customers in the tandem network.

Theorem 5.5 (Asymptotic Normality for $p_1^{(T)}$)**.** *Let the conditions of Theorem 5.2 be satisfied for all $\lambda\in\mathcal{L}$. Then we have for the estimator $p_1^{(T)}$ from (31) with a finite constant $s_2\in\mathbb{R}_+$ as $T\to\infty$,*

$$\sqrt{B_T T}\,\left(p_1^{(T)} - p_1\right) \xrightarrow{\mathcal{D}} \mathcal{N}_1\left(0, \frac{2\pi\, s_2\int W(\alpha)^2 d\alpha}{\left(\sum_{\lambda\in\mathcal{L}}|G_1(\lambda)|^2\right)^2}\right).$$

Proof. We have

$$p_1^{(T)} = \frac{\sum_{\lambda\in\mathcal{L}}\big[|G_1^{(T)}(\lambda)|^2 - \mathrm{Re}\big(G_1^{(T)}(\lambda)\cdot H_{11}^{(T)}(-\lambda)\big)\big]}{\sum_{\lambda\in\mathcal{L}}|G_1^{(T)}(\lambda)|^2}. \tag{32}$$

Consider first the denominator of (32). Since $\sqrt{B_T T}(G_1^{(T)}(\lambda) - G_1(\lambda))$ is asymptotically normal according to Theorem 5.2, $|G_1^{(T)}(\lambda)|^2$ converges to $|G_1(\lambda)|^2$ in probability for each $\lambda\in\mathcal{L}$ by the continuous mapping theorem. Thus, the denominator converges in probability to $\sum_{\lambda\in\mathcal{L}}|G_1(\lambda)|^2$.

Each summand of the numerator is presentable as function value of a differentiable mapping φ of a finite set of spectral estimates $f_{D_iA_j}^{(T)}(\lambda_k)$ with $i,j\in\{1,2\}, \lambda_k\in\mathbb{R}, k\in\mathbb{N}$, appropriate. Due to Theorem 4.7 the delta method yields that

$$\sqrt{B_T T}\,\left(|G_1^{(T)}(\lambda)|^2 - \mathrm{Re}\big(G_1^{(T)}(\lambda)\cdot H_{11}^{(T)}(-\lambda)\big)\right)$$

converges in distribution to a normal distribution with mean

$$|G_1(\lambda)|^2 - \mathrm{Re}\big(G_1(\lambda)\cdot H_{11}(-\lambda)\big) = p_1 \cdot |G_1(\lambda)|^2$$

for each $\lambda \in \mathcal{L}$. Thus, the numerator converges in distribution to a normal distribution with mean $p_1 \cdot \sum_{\lambda\in\mathcal{L}} |G_1(\lambda)|^2$. The asymptotic variance term s_2 can be computed based on the derivative φ' of the mapping φ involved following the delta method by $\varphi'\mathbf{\Sigma}\varphi'^t$ with $\mathbf{\Sigma}$ from Theorem 4.7. However, the term s_2 is too lengthy and complicated to be given here in explicit terms. By the lemma of Slutsky the proof is complete. □

An alternative estimator for the routing probability p_1 based on observations of the tandem system in the time interval $(0,T]$ can be constructed as

$$\widetilde{p_1}^{(T)} = 1 - \frac{D_1((0,T])}{A_1((0,T])}. \tag{33}$$

Intuitively, as $T \to \infty$, the observed relative frequency of arrivals at node 1 which depart from the system after having been served at node 1 should tend to the true routing probability $1-p_1$. Due to Theorem 4.4 we have

$$\mathbb{E}D_1((0,t]) = \eta_1^{\mathbf{D}}\, t = (1-p_1)\,\eta_1^{\mathbf{A}}\, t = (1-p_1)\,\mathbb{E}A_1((0,t])$$

since $H_{11}(0) = G_1(0)(1-p_1)$ and $H_{12}(0) = 0$. It follows: $1-p_1 = \frac{\mathbb{E}D_1((0,t])}{\mathbb{E}A_1((0,t])}$.

The following result holds, see Theorem 1 in [8]:

Proposition 5.6. *Let* $\mathbf{N}$ *be an* $m+n$*-variate point process on* $\mathbb{R}$ *satisfying Assumption 4.2. Set*

$$q_{N_a}^{(T)} := \frac{N_a((0,T])}{T} \quad \textit{for } a \in \{1,2,\ldots,m+n\}.$$

Then we have for all $k \geq 1$ *and* $a_1,\ldots,a_k \in \{1,2,\ldots,m+n\}$ *as* $T\to\infty$*:*

$$\sqrt{T}\begin{pmatrix} q_{N_{a_1}}^{(T)} - E(q_{N_{a_1}}^{(T)}) \\ \vdots \\ q_{N_{a_k}}^{(T)} - E(q_{N_{a_k}}^{(T)}) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_k\left(0, 2\pi\mathbf{\Gamma}\right),$$

with matrix $\mathbf{\Gamma}$ *given by the entries* $\Gamma_{a_i a_j} = f_{N_{a_i}N_{a_j}}(0) = \frac{1}{2\pi}\breve{C}_{N_{a_i}N_{a_j}}(\mathbb{R})$.

Note that due to stationarity the mean $E(q_{N_{a_j}}^{(T)})$ is independent of T for all j. With $q_{A_1}^{(T)} := \frac{A_1((0,T])}{T}$ and $q_{D_1}^{(T)} := \frac{D_1((0,T])}{T}$ we thus have under the conditions from Proposition 5.6 as $T\to\infty$

$$\sqrt{T}\begin{pmatrix} q_{D_1}^{(T)} - \eta_1^{\mathbf{D}} \\ q_{A_1}^{(T)} - \eta_1^{\mathbf{A}} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_2\left(0, 2\pi\begin{pmatrix} f_{D_1D_1}(0) & f_{D_1A_1}(0) \\ f_{D_1A_1}(0) & f_{A_1A_1}(0)\end{pmatrix}\right).$$

It follows an asymptotic result for the estimator $\widetilde{p_1}^{(T)}$ of the routing probability given by (33):

Theorem 5.7 (Asymptotic Normality for $\widetilde{p}_1^{(T)}$)**.** *Let* $\mathbf{N} = (D_1, A_1)$ *satisfy the conditions of Proposition 5.6. Then we have as* $T \to \infty$*:*

$$\sqrt{T}\ \left(\widetilde{p}_1^{(T)} - p_1\right) \xrightarrow{\mathcal{D}} \mathcal{N}_1\left(0,\ 2\pi\ \frac{f_{D_1D_1}(0) + (1-p_1)^2 f_{A_1A_1}(0)}{(\eta_1^{\mathbf{A}})^2}\right).$$

Proof. The proof relies on a Taylor expansion analogously to the proof of Theorem 4.10:

$$\sqrt{T}\left(\widetilde{p_1}^{(T)} - p_1\right) = \sqrt{T}(q_{D_1}^{(T)} - \eta_1^{\mathbf{D}})\frac{1}{\eta_1^{\mathbf{A}}} - \sqrt{T}(q_{A_1}^{(T)} - \eta_1^{\mathbf{A}})\frac{\eta_1^{\mathbf{D}}}{(\eta_1^{\mathbf{A}})^2} + o_p(1).$$

□

We recognize for the asymptotic behavior of the estimator $\widetilde{p}_1^{(T)}$ the optimal parametric rate $\sqrt{T}$.

Estimation of the densities $g_1(\cdot)$ and $g_2(\cdot)$

We now turn to the estimation of the service time densities $g_1(\cdot)$ and $g_2(\cdot)$. It is clear from the structure of the impulse-response function $\mathbf{h}(u)$ defined in (6) that without knowing the routing probability p_1 the estimation of $g_1(\cdot)$ and $g_2(\cdot)$ is a hard problem due to the convolution involved. As we have seen this was not the case for the estimation of the Fourier transforms $G_1(\cdot)$ and $G_2(\cdot)$, a big well-known advantage of estimation in the frequency domain. In this subsection, we assume that p_1 is known, which is no crucial restriction given that from the previous subsection we have an estimate of p_1 with rate $\sqrt{T}$ at hand.

Then, clearly, appropriate estimators for $g_1(u)$ and $g_2(u)$ are

$$g_1^{(T)}(u) := \frac{h_{11}^{(T)}(u)}{1-p_1} = \frac{1 - cos(C_T u)}{\pi C_T u^2}\frac{1}{1-p_1}\sum_{q=-Q_T}^{Q_T} \exp(iuC_T q)\frac{f_{D_1A_1}^{(T)}(C_T q)}{f_{A_1A_1}^{(T)}(C_T q)}, \tag{34}$$

$$g_2^{(T)}(u) := h_{22}^{(T)}(u) = \frac{1 - cos(C_T u)}{\pi C_T u^2}\sum_{q=-Q_T}^{Q_T} \exp(iuC_T q)\frac{f_{D_2A_2}^{(T)}(C_T q)}{f_{A_2A_2}^{(T)}(C_T q)}, \tag{35}$$

for $u \in \mathbb{R}$ with small C_T and large Q_T.

As a direct corollary from Theorem 4.11 we obtain:

Theorem 5.8 (Asymptotic Normality for $g_1^{(T)}$ and $g_2^{(T)}$)**.** *Consider the tandem model of queues given by* (5) *and* (6) *as described in Section 2. Suppose that all assumptions of Theorem 5.2 are in force. Moreover, assume for* $j = 1, 2$ *that* $\inf_{\lambda\in\mathbb{R}} f_{A_jA_j}(\lambda) \geq \frac{1}{L_1}$ *for a constant* $L_1 < \infty$*. In addition, let* $\int |u|^2 \|\mathbf{h}(u)\|_\infty du < \infty$ *and* $\int |\lambda|^r \|\mathbf{H}(\lambda)\|_\infty d\lambda < \infty$ *for some* $r > 0$*. Further, consider* $C_T \to 0, Q_T C_T \to \infty, Q_T^3 B_T^{-1} T^{-1} \to 0$*,* $C_T^4 Q_T B_T T \to 0$ *and* $C_T^{-2r-2} Q_T^{-2r-1} B_T T \to 0$ *as* $T \to \infty$*. Then, as* $T \to \infty$*, we have for the esti-*

mators of the service time densities $g_1(\cdot)$ and $g_2(\cdot)$ given by (34) and (35) for $u \in \mathbb{R}$,

$$\sqrt{C_T^{-2} Q_T^{-1} B_T T} \left(\begin{pmatrix} g_1^{(T)}(u) \\ g_2^{(T)}(u) \end{pmatrix} - \begin{pmatrix} g_1(u) \\ g_2(u) \end{pmatrix} \right)$$
$$\xrightarrow{\mathcal{D}} \mathcal{N}_2^{\mathbb{C}} \left(\mathbf{0}, 2\pi \int W(\alpha)^2 d\alpha \, \frac{1}{Q_T} \sum_{q=-Q_T}^{Q_T} \begin{pmatrix} \frac{f_{\varepsilon_1\varepsilon_1}(C_T q)}{f_{A_1A_1}(C_T q)} \frac{1}{(1-p_1)^2} & 0 \\ 0 & \frac{f_{\varepsilon_2\varepsilon_2}(C_T q)}{f_{A_2A_2}(C_T q)} \end{pmatrix} \right).$$

Estimation of the multiple coherence

In the rest of this section we will take a closer look at the role of the error process ε in our setting,

$$\varepsilon(B) = \begin{pmatrix} \varepsilon_1(B) \\ \varepsilon_2(B) \end{pmatrix} = \begin{pmatrix} D_1(B) - E[D_1(B)|\mathbf{A}] \\ D_2(B) - E[D_2(B)|\mathbf{A}] \end{pmatrix}, \quad B \in \mathcal{B}(\mathbb{R}).$$

The multiple coherence of the output process D_i with the input processes A_1 and A_2 at frequency $\lambda \in \mathbb{R}$ is defined as

$$|R_{D_i\mathbf{A}}(\lambda)|^2 := \frac{f_{D_i\mathbf{A}}(\lambda) f_{\mathbf{AA}}(\lambda)^{-1} f_{\mathbf{A}D_i}(\lambda)}{f_{D_iD_i}(\lambda)}, \qquad \text{for } i = 1, 2,$$

compare with [9] p.296, for the time series case.

The notion multiple coherence for processes is an analogy to the correlation of random variables and represents a measure of a model's goodness of fit. It expresses the extent to which the output process D_i is determined by the input processes A_1 and A_2 through linear time-invariant operations. Theorem 4.4(*iii*) implies that for $i = 1, 2$ and $\lambda \in \mathbb{R}$,

$$f_{\varepsilon_i\varepsilon_i}(\lambda) = \left(1 - |R_{D_i\mathbf{A}}(\lambda)|^2\right) f_{D_iD_i}(\lambda). \tag{36}$$

Since power spectral matrices are non-negative definite we have $0 \le |R_{D_i\mathbf{A}}(\lambda)|^2 \le 1$ for all $\lambda \in \mathbb{R}$, $i = 1, 2$. We construct an estimator for the quantity $|R_{D_i\mathbf{A}}(\lambda)|^2$ in the obvious plug-in manner analogously to the estimator of the transfer function,

$$|R^{(T)}_{D_i\mathbf{A}}(\lambda)|^2 := \frac{f^{(T)}_{D_i\mathbf{A}}(\lambda) f^{(T)}_{\mathbf{AA}}(\lambda)^{-1} f^{(T)}_{\mathbf{A}D_i}(\lambda)}{f^{(T)}_{D_iD_i}(\lambda)}, \qquad \text{for } i = 1, 2,$$

with spectral estimators as defined in (18). We have the following asymptotic result:

Theorem 5.9. *If in addition to the conditions of Theorem 5.2 it holds $f_{D_iD_i}(\lambda) > 0$ for $\lambda \in \mathbb{R}$, then the asymptotic behavior of the estimator $|R^{(T)}_{D_i\mathbf{A}}(\lambda)|^2$ for $i = 1, 2$ is as $T \to \infty$ given by:*

$$\sqrt{B_T T} \left(|R^{(T)}_{D_i\mathbf{A}}(\lambda)|^2 - |R_{D_i\mathbf{A}}(\lambda)|^2 \right)$$
$$\xrightarrow{\mathcal{D}} \mathcal{N}_1 \left(0, 4\pi \int W(\alpha)^2 d\alpha \cdot |R_{D_i\mathbf{A}}(\lambda)|^2 \cdot \left(1 - |R_{D_i\mathbf{A}}(\lambda)|^2\right)^2 \right).$$

Proof. The proof runs analogously to the proof of Theorem 4.10. The main step to show is:

$$\begin{aligned}
&\sqrt{B_T T}\left(|R^{(T)}_{D_i\mathbf{A}}(\lambda)|^2 - |R_{D_i\mathbf{A}}(\lambda)|^2\right)\\
&= \sqrt{B_T T}\left[f^{(T)}_{D_i\mathbf{A}}(\lambda) - f_{D_i\mathbf{A}}(\lambda)\right] f_{\mathbf{AA}}(\lambda)^{-1} f_{\mathbf{A}D_i}(\lambda)\, f_{D_iD_i}(\lambda)^{-1}\\
&-f_{D_i\mathbf{A}}(\lambda)\, f_{\mathbf{AA}}(\lambda)^{-1}\sqrt{B_T T}\left[f^{(T)}_{\mathbf{AA}}(\lambda) - f_{\mathbf{AA}}(\lambda)\right] f_{\mathbf{AA}}(\lambda)^{-1} f_{\mathbf{A}D_i}(\lambda)\, f_{D_iD_i}(\lambda)^{-1}\\
&+f_{D_i\mathbf{A}}(\lambda)\, f_{\mathbf{AA}}(\lambda)^{-1}\sqrt{B_T T}\left[f^{(T)}_{\mathbf{A}D_i}(\lambda) - f_{\mathbf{A}D_i}(\lambda)\right] f_{D_iD_i}(\lambda)^{-1}\\
&-|R_{D_i\mathbf{A}}(\lambda)|^2\sqrt{B_T T}\left[f^{(T)}_{D_iD_i}(\lambda) - f_{D_iD_i}(\lambda)\right] f_{D_iD_i}(\lambda)^{-1} + o_p(1).
\end{aligned}$$

□

In case of $|R_{D_i\mathbf{A}}(\lambda)|^2 \approx 0$, relation (36) implies that the power spectrum of the error process ε_i at frequency λ is nearly as large as the corresponding value of the power spectrum of the output process D_i. This case is called incoherent, since the error variance is not reduced by the input processes A_1 and A_2.

In case of $|R_{D_i\mathbf{A}}(\lambda)|^2 \approx 1$, called the perfectly coherent case, there is a strong linear relationship between the input processes A_1 and A_2 and the output process D_i at frequency $\lambda \in \mathbb{R}$ and the spectrum of the error process ε_i at λ is close to zero.

Thus, the multiple coherence $|R_{D_i\mathbf{A}}(\lambda)|^2$ is seen to provide a measure of the degree of linear relationship between the process D_i and the input processes A_1 and A_2 at frequency λ.

Recall the classical spectral representation theorem of Cramér for stationary processes, see e.g. [9] or [10]. In a modified version it says that the stationary process ε_i can be written in the form, for $B \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
&\varepsilon_i((0,t]) - \lim_{\substack{s\uparrow t\\ s\neq t}} \varepsilon_i((0,s])\\
&= \underbrace{\mathbb{E}\left(\varepsilon_i((0,t]) - \lim_{\substack{s\uparrow t\\ s\neq t}} \varepsilon_i((0,s])\right)}_{=0} + \int_{-\infty}^{+\infty} \exp(i\lambda t)\, K(\lambda)\, dZ(\lambda),
\end{aligned} \tag{37}$$

with $|K(\lambda)|^2 = f_{\varepsilon_i\varepsilon_i}(\lambda)$ for all $\lambda \in \mathbb{R}$ and a complex-valued stochastic process Z with orthonormal increments. Note that $\varepsilon_i((0,t]) - \lim_{\substack{s\uparrow t\\ s\neq t}} \varepsilon_i((0,s])$ is the jump level of ε_i at time point t. The spectral representation shows a decomposition of the process into periodic components (sine and cosine) of frequencies λ and random amplitudes with variance $|K(\lambda)|^2 = f_{\varepsilon_i\varepsilon_i}(\lambda)$. It enlightens the meaning of the name power spectrum: The value $f_{\varepsilon_i\varepsilon_i}(\lambda)$ gives the power of the component of frequency λ.

Assuming that $|R_{D_i\mathbf{A}}(\lambda)|^2$ is close to 1 for almost all $\lambda \in \mathbb{R}$, the error process ε_i itself is reduced to a small quantity close to zero according to the spectral representation (37). By definition of ε_i this implies the approximation,

$$D_i(B) \approx \mathbb{E}[D_i(B)|\mathbf{A}] \quad \text{for } B \in \mathcal{B}(\mathbb{R}).$$

Thus, for all time instants t the number of external departures from node i is nearly equal to the expected number of external departures from node i

upon complete knowledge of the arrival processes A_1 and A_2. Equality would mean that the output process D_i is entirely determined by the knowledge of A_1 and A_2 which would imply that the service times are deterministic, i.e. the service densities are Dirac measures, and that also the routing decisions are deterministic, i.e. p_1 equals 0 or 1. Clearly, the observed data of the external arrival and departure processes would also indicate this situation.

6. Generalizations of the tandem model of two nodes

We sketch here how our statistical approach of the tandem model studied so far can be extended to more general queueing models. We always assume that we are able to observe all external arrival and external departure processes at the G/G/∞ nodes of the networks, but have no information about the movements of the customers between the nodes in the system. The aim is to estimate the service time distributions at the nodes as well as the routing probabilities.

Feedforward models of more than two nodes

Let us consider a tandem system of three nodes according to Figure 2. The structure of customer movements leads to the following equations for Borel time sets $B \in \mathcal{B}(\mathbb{R})$:

$$
\begin{aligned}
D_1(B) &= \sum_{j\in\mathbb{Z}} \mathbf{1}_B(\sigma_j^{A_1} + \gamma_j^{(1)}) \cdot Z_j^{(1-p_{12})},\\
D_2(B) &= \sum_{j\in\mathbb{Z}} \mathbf{1}_B(\sigma_j^{A_1} + \gamma_j^{(1)} + \gamma_j^{(12)}) \cdot (1 - Z_j^{(1-p_{12})}) \cdot Z_j^{(1-p_{23})}\\
&\quad + \sum_{j\in\mathbb{Z}} \mathbf{1}_B(\sigma_j^{A_2} + \gamma_j^{(2)}) \cdot \widetilde{Z}_j^{(1-p_{23})},\\
D_3(B) &= \sum_{j\in\mathbb{Z}} \mathbf{1}_B(\sigma_j^{A_1} + \gamma_j^{(1)} + \gamma_j^{(12)} + \gamma_j^{(123)}) \cdot (1 - Z_j^{(1-p_{12})}) \cdot (1 - Z_j^{(1-p_{23})})\\
&\quad + \sum_{j\in\mathbb{Z}} \mathbf{1}_B(\sigma_j^{A_2} + \gamma_j^{(2)} + \gamma_j^{(23)}) \cdot (1 - \widetilde{Z}_j^{(1-p_{23})}) + \sum_{j\in\mathbb{Z}} \mathbf{1}_B(\sigma_j^{A_3} + \gamma_j^{(3)}),
\end{aligned}
$$

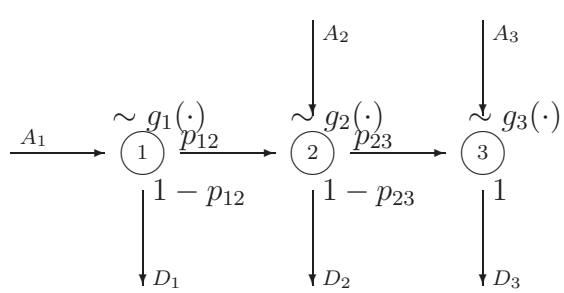

FIG 2. *Tandem network of three nodes.*

with Bernoulli variables $Z_j^{(1-p_{12})} \sim \text{Bern}(1-p_{12})$ and $Z_j^{(1-p_{23})}, \widetilde{Z}_j^{(1-p_{23})} \sim \text{Bern}(1-p_{23})$ as well as random variables $\gamma_j^{(1)} \sim g_1(\cdot)$, $\gamma_j^{(12)}, \gamma_j^{(2)} \sim g_2(\cdot)$ and $\gamma_j^{(123)}, \gamma_j^{(23)}, \gamma_j^{(3)} \sim g_3(\cdot)$, for all $j \in \mathbb{Z}$. All random variables are independent from each other. In a matrix notation this yields

$$E\left[\begin{pmatrix} D_1(B) \\ D_2(B) \\ D_3(B) \end{pmatrix} \middle| \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}\right] = \int\int \mathbf{1}_{B-u}(s)\mathbf{h}(s)ds\, d\begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}(u), \quad B \in \mathcal{B}(\mathbb{R}),$$

with 3×3 matrix $\mathbf{h}(\cdot)$ having the component-by-component-wise computed Fourier transform $\mathbf{H}(\cdot)$:

$$\mathbf{H}(\lambda) = \begin{pmatrix} G_1(\lambda)(1-p_{12}) & 0 & 0 \\ G_1(\lambda)G_2(\lambda)p_{12}(1-p_{23}) & G_2(\lambda)(1-p_{23}) & 0 \\ G_1(\lambda)G_2(\lambda)G_3(\lambda)p_{12}p_{23} & G_2(\lambda)G_3(\lambda)p_{23} & G_3(\lambda) \end{pmatrix} \quad \text{for } \lambda \in \mathbb{R}. \tag{38}$$

(38) directly implies the following identification relations:

$$\begin{aligned} G_3(\lambda) &= H_{33}(\lambda) \\ G_2(\lambda) &= H_{22}(\lambda) + \frac{H_{32}(\lambda)}{H_{33}(\lambda)} \qquad & p_{23} &= \frac{H_{32}(\lambda)}{H_{32}(\lambda) + H_{22}(\lambda)H_{33}(\lambda)} \\ G_1(\lambda) &= H_{11}(\lambda) + \frac{H_{21}(\lambda)}{H_{22}(\lambda)} \qquad & p_{12} &= \frac{H_{21}(\lambda)}{H_{21}(\lambda) + H_{11}(\lambda)H_{22}(\lambda)}. \end{aligned} \tag{39}$$

Note that the equations for $G_1(\lambda)$ and p_{12} are exactly the same as in case of the tandem system of two nodes (compare with Section 5). This is not surprising since for these parameters the existence of a third node does not provide any further information. From (39) estimators for the routing probabilities p_{12} and p_{23} as well as for $G_1(\lambda), G_2(\lambda)$ and $G_3(\lambda)$, for $\lambda \in \mathbb{R}$, then can be constructed and their respective asymptotic normality can be proved in analogy to Section 5. A further generalization to feedforward systems of more than three nodes is possible in an obvious straightforward way.

Tandem models with feedback

Next we consider queueing systems with positive feedback probabilities at the nodes, i.e., a customer may be served at the same node several times consecutively before jumping on to another node. Let $B \in \mathcal{B}(\mathbb{R})$. The underlying point process structure of the system given by Figure 3 implies that (note that clearly $p_{11} + p_{10} + p_{12} = 1$ and $p_{22} + p_{20} = 1$):

$$D_1(B) = \sum_{j\in\mathbb{Z}} \mathbf{1}_B \left(\sigma_j^{A_1} + \gamma_j^{(1)} + \sum_{l=1}^{W_j^{p_{11}}} \gamma_{j,l}^{(11)}\right) \cdot Z_j^{\frac{p_{10}}{p_{10}+p_{12}}},$$

$$D_2(B) = \sum_{j\in\mathbb{Z}} \mathbf{1}_B \left(\sigma_j^{A_1} + \gamma_j^{(1)} + \sum_{l=1}^{W_j^{p_{11}}} \gamma_{j,l}^{(11)} + \gamma_j^{(12)} + \sum_{s=1}^{W_j^{p_{22}}} \gamma_{j,s}^{(122)}\right) \cdot \left(1 - Z_j^{\frac{p_{10}}{p_{10}+p_{12}}}\right)$$

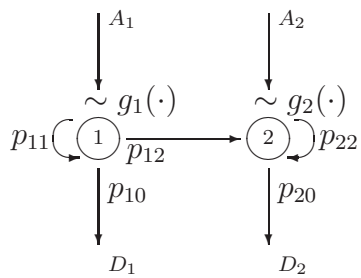

FIG 3. *Tandem system of two nodes with feedback.*

$$+\sum_{j\in\mathbb{Z}} \mathbf{1}_B \left(\sigma_j^{A_2} + \gamma_j^{(2)} + \sum_{s'=1}^{\widetilde{W}_j^{p_{22}}} \gamma_{j,s'}^{(22)} \right),$$

with Bernoulli variables $Z_j^{\frac{p_{10}}{p_{10}+p_{12}}} \sim \text{Bern}(\frac{p_{10}}{p_{10}+p_{12}})$ as well as random variables $\gamma_j^{(1)}, (\gamma_{j,l}^{(11)})_{l=1,2,\ldots,W_j^{p_{11}}} \sim g_1(\cdot)$, and $\gamma_j^{(12)}, \gamma_j^{(2)}, (\gamma_{j,s}^{(122)})_{s=1,2,\ldots,W_j^{p_{22}}}$, $(\gamma_{j,s'}^{(22)})_{s'=1,2,\ldots,\widetilde{W}_j^{p_{22}}} \sim g_2(\cdot)$ for all $j \in \mathbb{Z}$. Furthermore, $W_j^{p_{11}}$ are geometric random variables with success probability $(1-p_{11})$, i.e. $P(W_j^{p_{11}} = N) = p_{11}^N(1-p_{11})$ for $N \geq 0$, for all $j \in \mathbb{Z}$. Analogously, $W_j^{p_{22}}$ and $\widetilde{W}_j^{p_{22}}$ are geometric with success probability $(1-p_{22})$ for all $j \in \mathbb{Z}$. All random variables are independent from each other. In a matrix notation this yields

$$E\left[\begin{pmatrix} D_1(B) \\ D_2(B) \end{pmatrix} \middle| \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \right] = \int\int \mathbf{1}_{B-u}(s)\mathbf{h}(s)ds\, d\begin{pmatrix} A_1 \\ A_2 \end{pmatrix}(u)$$

with matrix

$$\mathbf{h}(s) = \begin{pmatrix} g_1 * \sum_{k=0}^{\infty} g_1^{*k} p_{11}^k p_{10} & 0 \\ g_1 * \sum_{k=0}^{\infty} g_1^{*k} p_{11}^k * g_2 * \sum_{s=0}^{\infty} g_2^{*s} p_{22}^s (1-p_{22}) p_{12} & g_2 * \sum_{s=0}^{\infty} g_2^{*s} p_{22}^s (1-p_{22}) \end{pmatrix}(s)$$

for $s \in \mathbb{R}$.

The corresponding transfer function $\mathbf{H}(\cdot)$ is then (note that $|G_1(\lambda)|p_{11} < 1$):

$$\mathbf{H}(\lambda) = \begin{pmatrix} \frac{p_{10}G_1(\lambda)}{1-p_{11}G_1(\lambda)} & 0 \\ \frac{p_{12}G_1(\lambda)}{1-p_{11}G_1(\lambda)} \frac{p_{20}G_2(\lambda)}{1-p_{22}G_2(\lambda)} & \frac{p_{20}G_2(\lambda)}{1-p_{22}G_2(\lambda)} \end{pmatrix} \text{ for } \lambda \in \mathbb{R}.$$

This yields that for identification of the five unknown parameters p_{10}, p_{12}, p_{20}, $G_1(\lambda)$, $G_2(\lambda), \lambda \in \mathbb{R}$, there are just three equations at hand. With respect to the routing probabilities we first observe that $\frac{p_{12}}{p_{10}} = \frac{H_{21}(\lambda)}{H_{11}(\lambda)H_{22}(\lambda)}$, thus, if the

feedback probability p_{11} is unknown we can just determine the ratio between p_{12} and p_{10}, not their absolute values. Furthermore, it is technically and intuitively clear that we cannot separate p_{11} and $G_1(\lambda)$ (and p_{22} and $G_2(\lambda)$ respectively) here without any further information. Therefore, for tandem systems with positive feedback probabilities according to Figure 3 with our method we may just identify the total sojourn time distributions of customers at the nodes, given by

$$\breve{G}_1(\lambda) := \sum_{k=0}^{\infty}(1-p_{11})p_{11}^k G_1^{k+1}(\lambda) = \frac{(1-p_{11})G_1(\lambda)}{1-p_{11}G_1(\lambda)} \text{ and } \breve{G}_2(\lambda) := \frac{(1-p_{22})G_2(\lambda)}{1-p_{22}G_2(\lambda)}.$$

Asymptotically normal estimators for the unknown parameters may then be constructed according to the relations

$$\breve{G}_2(\lambda) = H_{22}(\lambda),\ \breve{G}_1(\lambda) = H_{11}(\lambda) + \frac{H_{21}(\lambda)}{H_{22}(\lambda)}, \text{ and}$$
$$\frac{p_{12}}{p_{10}+p_{12}} = \frac{H_{21}(\lambda)}{H_{21}(\lambda)+H_{11}(\lambda)H_{22}(\lambda)}.$$

The case of queueing systems of two nodes with positive routing probability p_{21} for jumping back from node 2 to node 1 is much more involved. The statistical analysis of such stochastic networks of general topology is postponed to a following paper.

7. Simulations

We illustrate in this section the practicality of the estimators derived with a simulation study performed in R. We simulated the tandem network of two nodes given in Figure 1 under the following conditions. The external input processes A_1 and A_2 are assumed to be self-exciting processes with exponential decay (see [20] and [27]), i.e., they are characterized by their conditional intensities (with $\nu_i \geq 0, \tau_i < \beta_i$), for $t \in \mathbb{R},\ i = 1, 2$,

$$E\left[A_i((0,t]) - \lim_{\substack{s\uparrow t\\ s\neq t}} A_i((0,s])\middle|\{\sigma_j^{A_i} : \sigma_j^{A_i} < t\}\right] = \nu_i + \int_{-\infty}^{t} \tau_i \exp(-\beta_i(t-u))\, dA_i(u).$$

These processes belong to the class of Hawkes' processes which constitute the binary analogues of autoregressive time series. They have been extensively applied in various fields in science and engineering such as e.g. statistical seismology, auditory physiology, and electrical engineering, see [24]. With the chosen exponential decay the corresponding intensity is a Markov process and the covariances of the sequence of jump intervals are positive and decreasing, see [27]. In [28] an algorithm to simulate the process is proposed which we have followed. The advantage of this choice of arrival processes lies in the fact that we can easily show the sensitivity of our estimators with respect to the inner dependence of the arrival processes. If we increase β_i, then the inner dependence of

the process A_i decreases (compare with [28], Figures 7 and 9) and intuitively the estimators should then behave in a better way.

The service times are chosen to be exponentially distributed with parameter $\theta_1 = 0.8$ (node 1) and $\theta_2 = 3$ (node 2) respectively. The routing probability is set equal to $p_1 = 0.6$. The weight function W used in the calculations of the smoothed periodograms as estimators for the spectra according to (18) is assumed to be the triangular kernel given by

$$W(x) := (2 - 4|x|) \cdot \mathbf{1}_{[-0.5,0.5]}(x) \quad \text{for } x \in \mathbb{R}.$$

T is set to 20000. The bandwidth B_T is taken of order $O(T^{-0.25})$ fulfilling the conditions of the theorems. The characteristic function of the exponential distribution with parameter θ is explicitly computable. For the Fourier transform of the service time distribution at node j follows for $j = 1, 2$: $G_j(\lambda) = \frac{\theta_j}{\theta_j + i\lambda}$ which yields for its real part $\mathrm{Re}(G_j(\lambda)) = \frac{\theta_j^2}{\theta_j^2 + \lambda^2}$ and for its imaginary part $\mathrm{Im}(G_j(\lambda)) = \frac{-\theta_j \lambda}{\theta_j^2 + \lambda^2}$.

In the plots for the estimation of the real and imaginary parts of the Fourier transforms $G_j(\cdot)$ the black solid lines give the true functions and the dashed lines display the estimated functions. In addition we have plotted confidence intervals as grey solid lines. We have to stress here that these are just pointwise confidence intervals given for each λ which we have connected (compare with Remark 5.3). Up to now we have no uniform asymptotic normality results for our estimators at hand and thus no uniform confidence bands (over all λ) are available. Furthermore, we have estimated the entries of Σ which occur in the calculation of the confidence intervals which is the reason for these intervals sometimes being not sufficiently accurate. Nevertheless, the simulation results illustrate the good quality of our estimators.

We present here results for two different parameter settings for the arrival processes A_1 and A_2. In the first case called "β large" the processes are chosen to show little inner dependence. Here $\tau_1 = 0.4, \beta_1 = 18, \nu_1 = 1.6$ and $\tau_2 = 0.2, \beta_2 = 15, \nu_2 = 1.5$. On the contrary in the second case called "β small" the processes show more inner dependence, the parameters are set to $\tau_1 = 1, \beta_1 = 1.2, \nu_1 = 0.29$ and $\tau_2 = 0.9, \beta_2 = 1, \nu_2 = 0.16$. In both cases in our simulated time interval approximately 30000 customers entered the network via A_1 and A_2 respectively so that the situations can be well compared.

Figure 4 illustrates the different inner dependence structure of the processes in the cases "β large" (above) and "β small" (below). Here a cross indicates a customer having entered the system via the process A_1 and a circle indicates a customer having entered the system via A_2. In case of "β small" the greater inner dependence is shown by obvious clusters in the set of arrival points of A_1 and A_2, whereas in case of "β large" the occurrence of arrival points is more regular.

Figures 5 and 7 present the estimators given in (27) and (28) for the real and imaginary parts of $G_1(\cdot)$ (above) and $G_2(\cdot)$ (below) in case of "β large" and "β small". We observe that as expected the behaviour of the estimators is better in

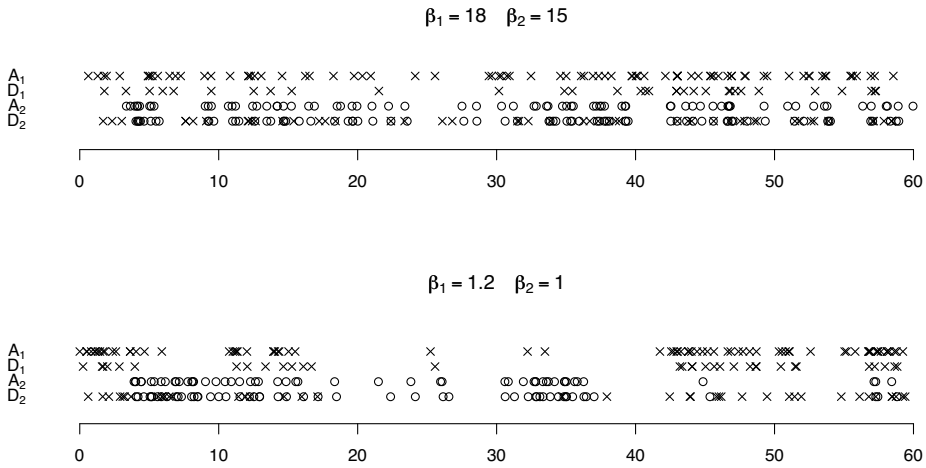

FIG 4. *Illustration of the inner dependence of the arrival processes A_1 and A_2 in the cases "β large" (above) and "β small" (below).*

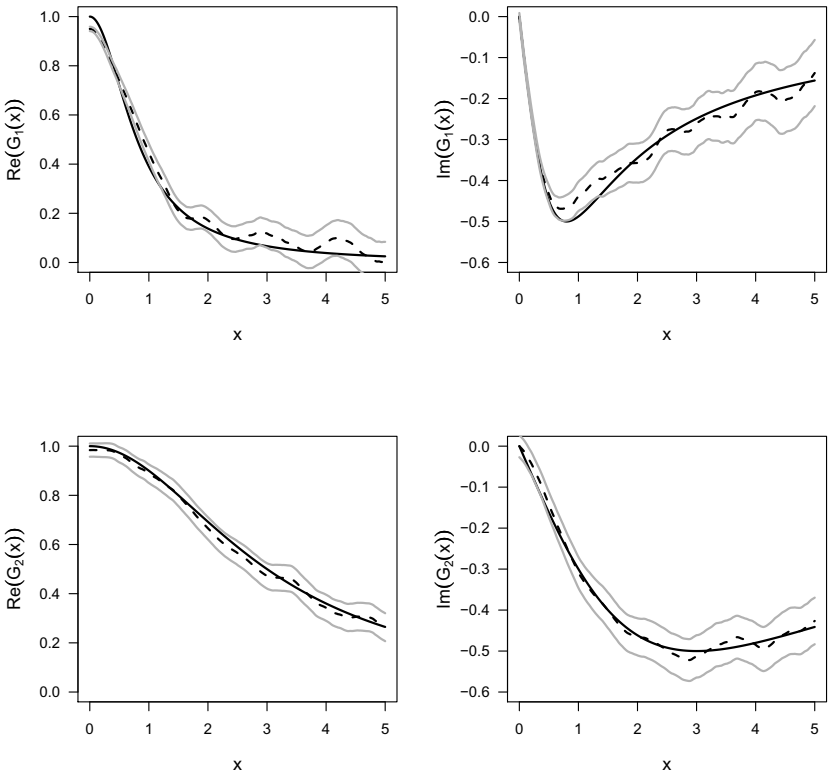

FIG 5. *Estimation of $G_1(\cdot)$ and $G_2(\cdot)$ for "β large".*

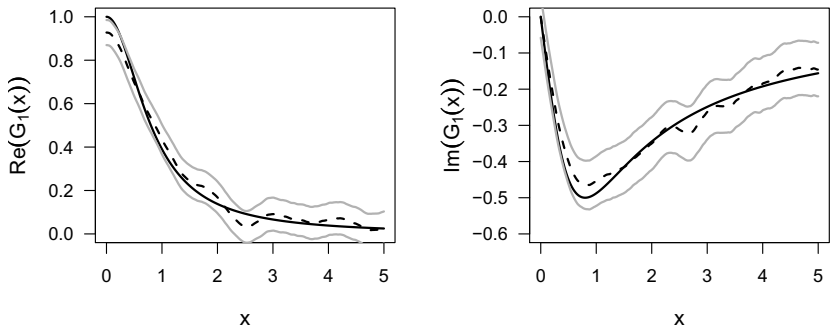

FIG 6. *Estimation of* $G_1(\cdot)$ *based on* $\widetilde{p_1}^{(T)}$ *for "*β *large".*

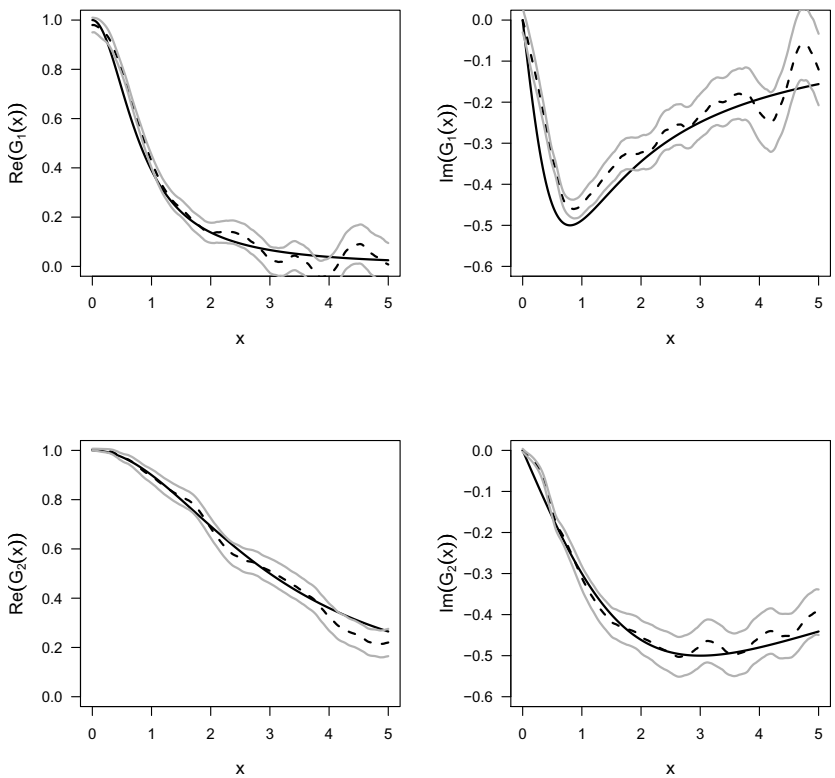

FIG 7. *Estimation of* $G_1(\cdot)$ *and* $G_2(\cdot)$ *for "*β *small".*

case of "β large". We notice that for "β small" the estimators are in particular inaccurate in the frequency interval $(0, 1)$, and since the asymptotic variance is also estimated the confidence intervals are here also not satisfactory. Furthermore, note that in the definition of the estimator for $G_1(\lambda)$ the estimator for the

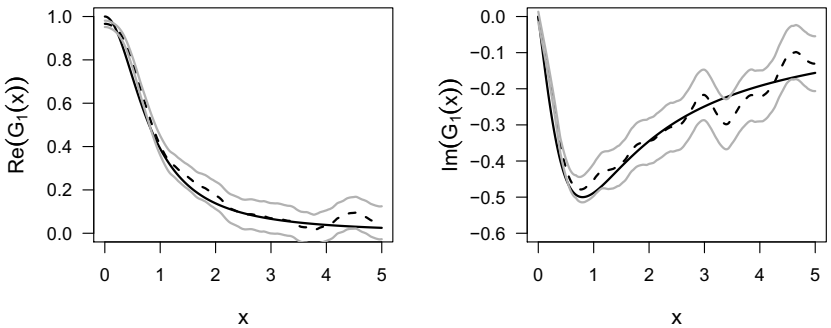

FIG 8. *Estimation of $G_1(\cdot)$ based on $\widetilde{p_1}^{(T)}$ for "β small".*

cross-spectrum $f_{D_2A_2}(\lambda)$ appears in the denominator whose absolute value can become very small. Simulations have shown that every time the absolute value of $f_{D_2A_2}(\lambda)$ is near to zero the estimator for $G_1(\lambda)$ behaves badly, which is no surprise. Thus, we have stabilized the estimator $G_1^{(T)}(\cdot)$ in our simulations by adding the small term 10^{-3} in the denominator of the second summand. Then the behaviour of $G_1^{(T)}(\cdot)$ is very good as Figures 5 and 7 show. An even better estimation result for $G_1(\cdot)$ can be achieved by computing the estimator $\widetilde{p_1}^{(T)}$ from (33) and plugging it into the term

$$\widetilde{G_1}^{(T)}(\lambda) = \frac{f_{D1A1}^{(T)}(\lambda)}{f_{A1A1}^{(T)}(\lambda)} \cdot \frac{1}{1-\widetilde{p_1}^{(T)}}, \tag{40}$$

suggested by the relation $H_{11}(\lambda) = G_1(\lambda)(1-p_1)$, compare with (7). By standard computations the following asymptotic result can be proved for the estimator $\widetilde{G_1}^{(T)}(\lambda)$:

$$\sqrt{B_T T}\left(\widetilde{G_1}^{(T)}(\lambda) - \frac{f_{D1A1}(\lambda)}{f_{A1A1}(\lambda)} \cdot \frac{1}{1-p_1}\right)$$
$$\xrightarrow{\mathcal{D}} \mathcal{N}_1^{\mathbb{C}}\left(0, 2\pi \int W(\alpha)^2 d\alpha \cdot \frac{f_{\varepsilon_1\varepsilon_1}(\lambda)}{f_{A_1A_1}(\lambda)} \cdot \frac{1}{(1-p_1)^2}\right).$$

The corresponding plots are given in Figure 6 ("β large") and 8 ("β small").

Figure 9 is based on 200 Monte Carlo simulations for two different estimators of the routing probability p_1. The parameters for the arrival processes are chosen as in case of "β small", but with $T = 2000$. Instead of using the optimal estimator $p_1^{(T)}$ defined in (31) we averaged the values of the estimator $\widehat{p_1}^{(T)}$ given in (29) over five different frequencies λ. This considerably reduced the computational complexity. The result is shown in Figure 9 (above). The plots below correspond to the intuitive estimator $\widetilde{p}_1^{(T)}$ given in (33). The histograms and the q-q-plots show a very good performance for both estimators; as expected from the theoretical results the intuitive estimator has a smaller variance.

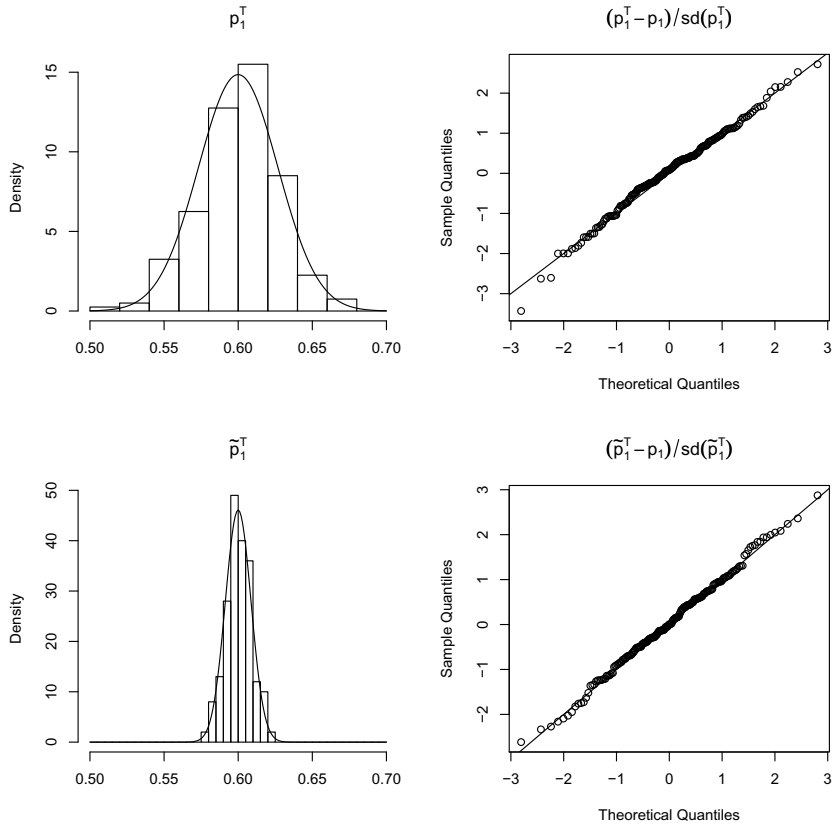

FIG 9. *Estimation of the routing probability p_1 based on 200 MC simulations.*

Acknowledgement

We thank the two referees and the associate editor for their valuable comments. The work of the first author has been supported by the Deutsche Forschungsgemeinschaft under DA 187/15-1.

References

[1] ASMUSSEN, S. (2003). *Applied Probability and Queues.* Second edition, Springer-Verlag, New York. MR1978607

[2] BINGHAM, N.H. AND S.M. PITTS (1999). Non-parametric estimation for the $M|G|\infty$ queue. *Ann. Inst. Statist. Math.*, 51(1), 71–97. MR1704647

[3] BINGHAM, N.H. AND S.M. PITTS (1999). Nonparametric inference from $M|G|1$ busy periods. *Comm. in Statistics - Stochastic Models*, 15, 247–272. MR1683605

[4] Böhm, H. and R. von Sachs (2009). Shrinkage estimation in the frequency domain of multivariate time series. *Journal of Multivariate Analysis*, 100(5), 913–935. MR2498723

[5] Bohman, H. (1960). Approximate Fourier analysis of distribution functions. *Arkiv för Matematik*, band 4, nr 10, 99–157. MR0126668

[6] Brillinger, D.R. (1972). The spectral analysis of stationary interval functions. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I: Theory of statistics, Univ. California Press, Berkeley, Calif., 483–513. MR0407972

[7] Brillinger, D.R. (1974). Cross-spectral analysis of processes with stationary increments including the stationary $G/G/\infty$ queue. *Annals of Probability*, 2, 815–827. MR0359221

[8] Brillinger, D.R. (1975). *Statistical inference for stationary point processes.* In: Stochastic processes and related topics, Proc. Summer Res. Inst., Statist. Inference for Stochastic Processes, Indiana Univ., Bloomington, Ind., Vol. 1, 55–99. Academic Press, New York. MR0381201

[9] Brillinger, D.R. (1981). *Time series. Data analysis and theory.* International Series in Decision Processes. Holt, Rinehart and Winston, Inc., New York-Montreal, Que.-London. MR0595684

[10] Brockwell, P.J. and R.A. Davis (1991). *Time Series: Theory and Methods.* Second edition, Springer-Verlag, New York. MR1093459

[11] Daley, D.J. and D. Vere-Jones (2003). *An introduction to the theory of point processes. Vol. 1.* Elementary theory and methods. Second edition. Probability and its Applications (New York). Springer-Verlag, New York. MR1950431

[12] Daley, D.J. and D. Vere-Jones (2008). *An introduction to the theory of point processes. Vol. 2.* General Theory and Structure. Second edition. Probability and its Applications (New York). Springer-Verlag, New York. MR2371524

[13] D'Auria, B. and S.I. Resnick (2008). The influence of dependence on data network models. *Advances in Applied Probability*, 40, 60–94. MR2411815

[14] Chu, Y-K. and J-C. Ke (2006). Confidence intervals of mean response time for an M/G/1 queueing system: Bootstrap simulation. *Applied Mathematics and Computation*, 180, 255–263. MR2263386

[15] Chu, Y-K. and J-C. Ke (2007). Mean response time for a G/G/1 queueing system: Simulated computation. *Applied Mathematics and Computation*, 186, 772–779. MR2314537

[16] Hall, P. (1985). On inference in a one-dimensional mosaic and an M/G/∞ queue. *Advances in Applied Probability*, 17, 210–229. MR0778601

[17] Hall, P. (1988) *Introduction to the theory of coverage processes.* John Wiley & Sons, New York. MR0973404

[18] Hall, P. and J. Park (2004). Nonparametric inference about service time distribution from indirect measurements. *J. R. Statist. Soc. B*, 66(4), 861–875. MR2102469

[19] Hansen, M.B. and S.M. Pitts (2006). Nonparametric inference from the $M/G/1$ workload. *Bernoulli*, 12(4), 737–759. MR2248235

[20] Hawkes, A.G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58, 83–90. MR0278410

[21] Jolai, F., S.M. Asadzadeh and M.R. Taghizadeh (2008). Performance estimation of an email contact center by a finite source discrete time Geo/Geo/1 queue with disasters. *Computers and Industrial Engineering*, 55, 543–556.

[22] Ke, J.-C. and Y.-K. Chu (2006). Nonparametric and simulated analysis of intensity for a queueing system. *Applied Mathematics and Computation*, 183, 1280–1291. MR2294087

[23] Liu, Z., L. Winter, C. Xia and F. Zhang (2006). Parameter inference of queueing models for IT systems using end-to-end measurements. *Performance Evaluation*, 63, pp. 36–60.

[24] Mino, H. (2001). Parameter estimation of the intensity process of self-exciting point processes using the EM algorithm. *IEEE Transactions on Instrumentation & Measurement*, 50, 658–664.

[25] Narayan Bhat, U. and S. Subba Rao (1987). Statistical analysis of queueing systems. *Queueing Systems*, 1, 217–247. MR0896227

[26] Neudecker, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products. *Journal of the American Statistical Association*, 64, 953–963.

[27] Oakes, D. (1975). The Markovian self-exciting process. *J. of Applied Probability*, 12, 69–77. MR0362522

[28] Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point process. *Annals of the Institute of Statistical Mathematics*, 31(B), 145–155. MR0541960

[29] Pitts, S. (1994). Nonparametric estimation of the stationary waiting time distribution function for the $GI|G|1$ queue. *The Annals of Statistics*, 3, 1428–1446. MR1311982

[30] Sharma, V. and R. Mazumdar (1998). Estimating traffic parameters in queueing systems with local information. *Performance Evaluation*, 32, 217–230.

[31] Stewart, G.W. (1990). Stochastic perturbation theory. *SIAM Review*, 32(4), pp. 579–610. MR1084571

[32] Van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge. MR1652247

[33] Xi, B., H. Chen, W.S. Cleveland and Th. Telkamp (2010). Statistical analysis and modeling of Internet VoIP traffic for network engineering. *Electronic Journal of Statistics*, 4, pp. 58–116. MR2645478