

# Statistical inference for mean-field queueing systems

Ioannis Lambadaris\*, Ahmed Sid-Ali†, Wei Sun ‡, Yiqiang Q. Zhao§

## Abstract

Mean-field limits have been used now as a standard tool in approximations, including for networks with a large number of nodes. Statistical inference on mean-fielded models has attracted more attention recently mainly due to the rapid emergence of data-driven systems. However, studies reported in the literature have been mainly limited to continuous models. In this paper, we initiate a study of statistical inference on discrete mean-field models (or jump processes) in terms of a well-known and extensively studied model, known as the power-of- $L$  ( $L \geq 2$ ), or the supermarket model, to demonstrate how to deal with new challenges in discrete models. We focus on system parameter estimation based on the observations of system states at discrete time epochs over a finite period. We show that by harnessing the weak convergence results developed for the supermarket model in the literature, an asymptotic inference scheme based on an approximate least squares estimation can be obtained from the mean-field limiting equation. Also, by leveraging the law of large numbers alongside the central limit theorem, the consistency of the estimator and its asymptotic normality can be established when the number of servers and the number of observations go to infinity. Moreover, numerical results for the power-of-two model are provided to show the efficiency and accuracy of the proposed estimator.

*2020 Mathematics Subject Classification:* 60K25 60K35 62F10 62F12

*Keywords:* Mean-field; Queueing systems; Supermarket model; Least square estimation; Consistency; Asymptotic normality

## 1 Introduction

The origins of mean-field theory trace back to the pioneering works of Curie [8] and Weiss [36] in magnetism and phase transitions. Since then, this theory has expanded across a wide array of fields to study interacting particle systems, including statistical physics [9, 16, 23, 24], biological systems [10, 30], communication networks [4, 20, 21, 22], and mathematical finance [19, 27].

Moreover, the application of mean-field theory in queueing systems is traced back to the work of Dobrushin and Sukhov [14] and has since proliferated due to its many benefits, see, e.g., [11, 12, 31, 35] for further developments. Indeed, in stochastic service systems, particularly those involving multiple

---

\*Department of Systems and Computer Engineering, Carleton University, Ottawa, Ontario, Canada. ioannis@sce.carleton.ca

†School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada. ahmedsidali@cunet.carleton.ca

‡Department of Mathematics and Statistics, Concordia University, Montreal, Canada. wei.sun@concordia.ca

§School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada. zhao@math.carleton.ca

parallel queues, load balancing is commonly applied to enhance performance by shortening queues, reducing wait times, and increasing throughput. This balancing mechanism effectively modifies the input-output dynamics to improve the system's quality of service. When such systems are viewed as interacting systems, mean-field theory becomes a natural framework to study their behavior. By using mean-field analysis, the performance of large systems can be evaluated by examining their limiting behavior as the system size approaches infinity. In particular, the limit often reduces when solving a deterministic system known as the mean-field limit, which corresponds to a McKean-Vlasov-type stochastic differential equation (SDE) solution. In McKean-Vlasov SDEs, the coefficients depend on both the process itself and its distribution, forming a class of non-linear SDEs. The study of such equations was initiated by McKean [23], inspired by Kac's work in kinetic theory [25].

Although extensive literature exists on mean-field interacting systems, research on their statistical inference has only gained attention in recent years. The pioneering work in this area is Kasonga's seminal paper [26], which addressed parameter estimation for interacting particle systems modeled by Itô SDEs through a maximum likelihood approach. After Kasonga's work, interest in the topic waned for nearly two decades before reemerging with significant contributions, as seen in [1, 3, 5, 17, 29, 32] and references therein. Since then, the field has grown steadily, establishing itself as a crucial area of research. This renewed interest can be attributed to novel applications of mean-field theory and the rise of new technologies enabling access to massive datasets generated by systems of interacting agents.

To date, most statistical inference studies on mean-field models, such as those mentioned above, focus on interacting diffusion systems, with limited research on statistical inference for mean-field systems with jumps. Notable exceptions include [13] and [28], where the authors proposed asymptotic estimation for the Bernoulli interaction parameter in a system of interacting Hawkes processes as both the number of particles and time approach infinity. In particular, as discussed in [37], statistical inference in mean-field queueing models remains largely unexplored, despite statistical inference in queueing systems being an active area of research. The reader can consult for example [2] where a comprehensive survey on parameter and state estimation for queueing systems across various estimation paradigms was provided, yet it does not address statistical inference for mean-field queueing systems. To fill this gap, we propose in the current paper a statistical inference scheme for the parameters governing a specific mean-field queueing system, namely the supermarket model, also known as the power of  $L \geq 2$  choices. Thus, to the best of our knowledge, our current proposal represents a novel contribution in this area.

The supermarket model was independently introduced by Vvedenskaya et al. [35] and Mitzenmacher [31]. It represents a system of  $N$  parallel identical queues, each served by a single server with a service rate  $\nu$  and infinite buffer capacity. Tasks arrive at a rate of  $N\lambda$ ; each task is allocated  $L$  queues chosen uniformly at random among the  $N$  and joins the shortest one, with ties resolved uniformly. All events in this system are independent. In particular, [35] and [31] studied the asymptotic behavior of the system as the number of servers becomes large, showing that the process associated with queue lengths converges to a deterministic limit represented by an infinite system of ordinary differential equations (ODEs). This model and its extensions have since become widely studied due to its theoretical and practical importance; see, e.g., [20, 6, 7] and the references therein. However, as noted, the statistical inference for this model remains unexplored, which is the focus of this paper.

We propose a statistical inference scheme to estimate the arrival and service rates in a supermarket model based on aggregate data obtained from discrete observations of a moment of the system over a finite period. To this end, we propose to exploit the ODE obtained at the mean-field limit to construct

an approximate least square estimator (LSE). Then, using the law of large numbers and the central limit theorem established in the literature, we show that the proposed estimator is consistent and asymptotically normal as the number of servers and observations grows large. In addition, we test our estimator on synthetic data obtained by simulation which shows the accuracy of our approach.

It is worthwhile to provide the following remark: An intriguing general approach to statistical inference for mean-field systems was proposed in [18]. This approach leverages a “misspecified” or limiting model, created by approximating the system through large-systems asymptotics, incorporating the law of large numbers and central limit theorem. This enables constructing an approximate likelihood function, evaluated against the data generated by the true model. The estimator is then obtained by maximizing this approximate likelihood function. A key advantage of this method is that the approximate likelihood has a conditionally Gaussian structure, due to the central limit theorem, which allows for efficient numerical evaluation of the estimator. Although one might consider using a similar method to estimate parameters for the supermarket model, the complexity of the approximate likelihood for this model complicates the analysis of the estimator’s asymptotic properties. This difficulty motivates the adoption of an alternative approach, specifically, an approximate LSE scheme.

The rest of the paper is organized as follows: First, in Section 2, we recall the supermarket model and introduce the appropriate notations. We also review some well-known asymptotic results about the model, including a new technical result in Proposition 2.1 that will be used in the sequel. In Section 3, we introduce the inference scheme along with our LSE and prove both the consistency and asymptotic normality of the estimator. To facilitate the reading, we put all the long proofs in the appendices. Section 4 provides simulations demonstrating that our estimator accurately predicts the system parameters and validates the asymptotic normality result. Finally, in Section 5, we present conclusions and open questions, followed by the bibliography.

## 2 Queuing network with selection of the shortest queue among several servers

### 2.1 The setting

We start by recalling the supermarket model, first introduced separately in [31] and [35]. Consider a network with  $N$  identical queues, each with a single server of service rate  $\nu$  and an infinite buffer. Tasks arrive at rate  $N\lambda$ , and each task is allocated  $L$  queues with uniform probability among the  $N$  servers and elects to join the shortest one, ties being resolved uniformly. The  $L$  selected queues may coincide. All these random events are independent. Let  $X_i^N(t)$  denote the length of the  $i$ -th queue at time  $t$  and define the empirical measure process

$$\varrho_j^N(t) := \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{\{X_i^N(t)=j\}}, \quad j = 0, 1, \dots,$$

which takes values in the space  $\mathcal{P}(\mathbb{Z}_+)$  of probability measures on  $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$  identified with the infinite-dimensional simplex

$$\mathcal{S} := \left\{ s = \{s_j\}_{j \in \mathbb{Z}_+} \in \mathbb{R}_+^{\mathbb{Z}_+} : \sum_{j=0}^{\infty} s_j = 1 \right\},$$

where  $\mathbb{R}_+$  is the set of all non-negative real numbers. Define the subspace  $\mathcal{S}^N := \mathcal{S} \cap \frac{\mathbb{Z}_+^{\mathbb{Z}_+}}{N}$ . Thus,  $\varrho^N(t) \in \mathcal{S}^N$  for all  $t \geq 0$ . Throughout this paper, we fix a constant  $T > 0$ . Let  $\mathcal{D}([0, T], \mathcal{S})$  be the Skorokhod space of càdlàg functions defined on  $[0, T]$  with values in  $\mathcal{S}$ , equipped with the usual Skorokhod topology. Let  $\mathcal{C}([0, T], \mathcal{S})$  be the space of continuous functions defined on  $[0, T]$  with values on  $\mathcal{S}$ , equipped with the uniform topology.

## 2.2 Law of large numbers and central limit theorem for the empirical process

We recall now some results describing the asymptotic behavior of the supermarket model that will be used to build the statistical inference scheme and the related analysis. For  $p \geq 1$ , denote by  $\ell_p$  the space of  $p$ -th summable sequences, i.e.,

$$\ell_p = \left\{ x = \{x_j\}_{j \in \mathbb{Z}_+} \in \mathbb{R}^{\mathbb{Z}_+} : \sum_{j=0}^{\infty} |x_j|^p < \infty \right\},$$

and denote by  $\|\cdot\|_p$  the norm on it. In particular, let  $\ell_2$  be the space of square summable sequences equipped with the inner product

$$\langle x, y \rangle = \sum_{j=0}^{\infty} x_j y_j,$$

which makes it a Hilbert space. Moreover, define its subspace

$$\tilde{\ell}_2 := \left\{ s \in \ell_2 : \sum_{j=0}^{\infty} j^2 s_j^2 < \infty, \sum_{j=0}^{\infty} s_j = 0 \right\}.$$

Furthermore, for any  $j \in \mathbb{Z}_+$ , denote by  $e_j \in \ell_2$  the vector with 1 at the  $j$ -th coordinate and 0 elsewhere.

We first state the law of large numbers established in [20, Theorem 3.4] and reformulated in [6, Corollary 1].

**Theorem 2.1** *Suppose that  $\varrho^N(0) \rightarrow \varrho_0$  in  $\mathcal{S}$  as  $N \rightarrow \infty$ . Then  $\varrho^N \rightarrow \varrho$  in probability in  $\mathcal{D}([0, T], \mathcal{S})$ , where  $\varrho$  is the unique solution in  $\mathcal{C}([0, T], \mathcal{S})$  to the ODE:*

$$\dot{\varrho}(t) = F(\varrho(t)), \quad \varrho(0) = \varrho_0, \tag{2.1}$$

and, for any  $x \in \ell_1$ ,

$$\begin{aligned} F(x) &= \lambda \sum_{j=0}^{\infty} \left[ \sum_{i=1}^L \binom{L}{i} x_{j-1}^i \left( \sum_{m=j}^{\infty} x_m \right)^{L-i} - \sum_{i=1}^L \binom{L}{i} x_j^i \left( \sum_{m=j+1}^{\infty} x_m \right)^{L-i} \right] e_j \\ &\quad + \nu \sum_{j=0}^{\infty} [x_{j+1} - x_j] e_j. \end{aligned} \tag{2.2}$$

Next, we state results about the fluctuations of the empirical measure process  $\varrho^N$  from its law of large number limit  $\varrho$ . To this end, define the process

$$\mathcal{Z}^N(t) := \sqrt{N}(\varrho^N(t) - \varrho(t)), \quad t \in [0, T],$$

the operator

$$\begin{aligned} \Phi(t) &:= \lambda \sum_{j=0}^{\infty} (e_{j+1} - e_j)(e_{j+1} - e_j)^T \left( \sum_{i=1}^L \binom{L}{i} [\varrho_j(t)]^i \left( \sum_{m=j+1}^{\infty} \varrho_m(t) \right)^{L-i} \right) \\ &\quad + \nu \sum_{j=1}^{\infty} (e_{j-1} - e_j)(e_{j-1} - e_j)^T \varrho_j(t), \end{aligned} \quad (2.3)$$

and the map  $G : \tilde{\ell}_2 \times \mathcal{S} \rightarrow \tilde{\ell}_2$  by

$$G_j(x, s) := \left. \frac{\partial}{\partial u} F_j(s + ux) \right|_{u=0}, \quad j \in \mathbb{Z}_+, x \in \tilde{\ell}_2, s \in \mathcal{S}. \quad (2.4)$$

Finally, we recall the definition of cylindrical Brownian motion which is a generalization of the scalar Brownian motion to Hilbert spaces.

**Definition 2.1** *A collection of continuous real-valued stochastic processes  $\{(W_t(h))_{0 \leq t \leq T} : h \in \ell_2\}$  defined on a filtered probability space  $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$  is called an  $\ell_2$ -cylindrical Brownian motion if, for every  $h \in \ell_2$ ,  $(W_t(h))_{0 \leq t \leq T}$  is an  $\{\mathcal{F}_t\}$ -Brownian motion with variance  $t\|h\|_2^2$  and, for all  $h, k \in \ell_2$ ,*

$$\langle W(h), W(k) \rangle_t = t \langle h, k \rangle_2, \quad t \in [0, T].$$

We state now the central limit theorem introduced in [6, Theorem 2].

**Theorem 2.2** *Suppose that  $\sup_{N \in \mathbb{N}} \sum_{j=0}^{\infty} j^2 \varrho_j^N(0) < \infty$  and  $\varrho^N(0) \rightarrow \varrho_0$  in  $\mathcal{S}$  as  $N \rightarrow \infty$ . Also, suppose that  $\mathcal{Z}^N(0) \rightarrow z_0$  in  $\ell_2$  and that*

$$\sup_{N \in \mathbb{N}} \sum_{j=0}^{\infty} j^2 (\mathcal{Z}_j^N(0))^2 < \infty.$$

*Then  $\mathcal{Z}^N$  converges to  $\mathcal{Z}$  in distribution in  $\mathcal{D}([0, T], \ell_2)$  as  $N \rightarrow \infty$ , where  $\mathcal{Z}$  is the unique weak solution to the following SDE:*

$$d\mathcal{Z}(t) = G(\mathcal{Z}(t), \varrho(t))dt + a(t)dW(t), \quad \mathcal{Z}(0) = z_0 \in \tilde{\ell}_2, \quad (2.5)$$

*$G$  is defined by (2.4),  $a(t)$  is the symmetric square root of the operator  $\Phi(t)$  in (2.3), i.e.,  $a^2(t) = \Phi(t)$ , and  $W$  is an  $\ell_2$ -cylindrical Brownian motion.*

**Remark 1** *The stochastic integral  $\int_0^t a(s)dW(s)$  represents an  $\ell_2$ -valued Gaussian martingale  $M(t)$  given as*

$$M_i(t) = \sum_{j=0}^{\infty} \int_0^t A_{ij}(s)dB_j(s), \quad t \in [0, T], i \in \mathbb{Z}_+, \quad (2.6)$$

*with  $A_{ij}(s) = \langle e_i, a(s)e_j \rangle_2$  and  $\{B_j\}_{j \in \mathbb{Z}_+}$  is an independent sequence of standard Brownian motions. The well-posedness of the SDE (2.5) was established in [6, Proposition 2].*

### 2.3 The power of two choices model

For the sake of simplicity, we focus in this paper on a special case of the supermarket model obtained when  $L = 2$ . This model, known as the power of two choices, was first introduced and analyzed in [31]. In this case, the operator  $F$  in (2.2) takes the following explicit form:

$$F(x) = \lambda \sum_{j=0}^{\infty} \left[ 2x_{j-1} \sum_{m=j}^{\infty} x_m - 2x_j \sum_{m=j+1}^{\infty} x_m + x_{j-1}^2 - x_j^2 \right] e_j + \nu \sum_{j=0}^{\infty} [x_{j+1} - x_j] e_j, \quad x \in \ell_1, \quad (2.7)$$

the map  $G : \tilde{\ell}_2 \times \mathcal{S} \rightarrow \ell_2$  in (2.4) takes now the explicit form:

$$G_j(x, s) = 2\lambda \sum_{m=j}^{\infty} [x_{j-1}s_m + s_{j-1}x_m - x_j s_{m+1} - s_j x_{m+1}] + \nu(x_{j+1} - x_j), \quad j \in \mathbb{Z}_+, x \in \tilde{\ell}_2, s \in \mathcal{S},$$

and for any  $t \in [0, T]$ , the operator  $\Phi(t)$  in (2.3) becomes

$$\begin{aligned} \Phi(t) &= \lambda \sum_{j=0}^{\infty} (e_{j+1} - e_j)(e_{j+1} - e_j)^T \left( 2\varrho_j(t) \sum_{m=j+1}^{\infty} \varrho_m(t) + [\varrho_j(t)]^2 \right) \\ &\quad + \nu \sum_{j=1}^{\infty} (e_{j-1} - e_j)(e_{j-1} - e_j)^T \varrho_j(t). \end{aligned}$$

The following result shows that the solution to (2.5) is a Gaussian process. Although the proof is given for the power of two choices model, it can be adapted to cover the case with general  $L \geq 2$ .

**Proposition 2.1** *Suppose that  $L = 2$ . Then, the solution  $\mathcal{Z}(t)$  to the SDE (2.5) is a Gaussian process.*

**Proof** See Appendix A. □

## 3 Statistical inference of the supermarket model

Suppose the service and arrival rates,  $\nu$  and  $\lambda$ , that govern the system are unknown. Our goal is to estimate these parameters using observations collected over a specific time interval  $[0, T]$ . The complexity of the system's dynamics makes brute-force Monte Carlo estimation computationally intensive, particularly as the number of nodes,  $N$ , increases. Therefore, we propose developing an estimator that utilizes the weak convergence results outlined in Section 2, specifically the law of large numbers presented in Theorem 2.1 and the central limit theorem in Theorem 2.2.

In particular, we construct an approximate LSE based on the ODE given in (2.1). We subsequently demonstrate that this estimator is consistent and asymptotically normal as both the system size and the number of observations tend to infinity. For convenience, we denote the vector of unknown parameters as  $\theta = (\lambda, \nu)$ .

### 3.1 The data

Our objective is to estimate the parameter vector  $\theta$  that governs the dynamics of the power-of-two model based on observations collected over a finite time interval. Specifically, we assume that observations are not available for every server in the network; instead, they are gathered as an aggregate measure of the system. Collecting individual data for each server can be prohibitively costly in practice, particularly for large networks, which justifies our approach to data collection.

In this context, we assume that the available data for inference includes observations of the empirical measure of the system,  $\varrho^N(t)$ , over the finite time interval  $[0, T]$  at  $m$  discrete points, defined as  $t_k = \frac{kT}{m}$ :

$$\varrho_j^N(t_k) := \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{X_i^N(t_k)=j\}}, \quad j \in \mathbb{Z}_+, \quad k = 1, 2, \dots, m.$$

The observed data is then represented as

$$D^{N,m} := \{\varrho^N(t_k) : 1 \leq k \leq m\}.$$

Thus, this dataset reflects a realization of the system governed by the true parameters, say  $\theta^* = (\lambda^*, \nu^*)$ .

### 3.2 Least square estimator (LSE)

Recall from Theorem 2.1 that the empirical measure  $\varrho^N$  converges in probability towards  $\varrho$  the unique solution to the ODE (2.1) as the number of servers  $N \rightarrow \infty$ . We propose to take advantage of this result to build an approximate LSE for the parameters  $\theta^* = (\lambda^*, \nu^*)$  based on the dataset  $D^{N,m}$ .

**The least square function:** Let us first introduce the following functions defined for all  $j \in \mathbb{Z}_+$  and  $x \in \ell_1$ :

$$U_j(x) := 2x_{j-1} \sum_{i=j}^{\infty} x_i - 2x_j \sum_{i=j+1}^{\infty} x_i + x_{j-1}^2 - x_j^2, \quad (3.1)$$

and

$$V_j(x) := x_{j+1} - x_j. \quad (3.2)$$

Therefore, by (2.7), one can observe that

$$F(x) = \lambda \sum_{j=0}^{\infty} U_j(x) e_j + \nu \sum_{j=0}^{\infty} V_j(x) e_j. \quad (3.3)$$

Moreover, let  $\varrho(t)$  be the unique solution to the ODE (2.1). To highlight the dependence on the parameter  $\theta = (\lambda, \nu)$ , we will write  $\varrho(t, \theta)$  in the sequel. Furthermore, let us introduce the quadratic function defined for any  $\lambda, \nu \in \mathbb{R}$  by

$$\mathcal{G}(\lambda, \nu) := \sum_{j=0}^{\infty} \left[ \varrho_j(T, \theta^*) - \varrho_j(0, \theta^*) - \lambda \int_0^T U_j(\varrho(s, \theta^*)) ds - \nu \int_0^T V_j(\varrho(s, \theta^*)) ds \right]^2.$$

Notice that, by (2.1), one immediately observes that  $\mathcal{G}(\lambda^*, \nu^*) = 0$ . Thus, the function  $\mathcal{G}(\lambda, \nu)$  attains its minimum at  $\theta^* = (\lambda^*, \nu^*)$ , namely

$$\begin{aligned} 0 &= \frac{\partial \mathcal{G}}{\partial \lambda} \Big|_{(\lambda, \nu) = (\lambda^*, \nu^*)} \\ &= -2 \sum_{j=0}^{\infty} \left[ \varrho_j(T, \theta^*) - \varrho_j(0, \theta^*) - \lambda^* \int_0^T U_j(\varrho(s, \theta^*)) ds - \nu^* \int_0^T V_j(\varrho(s, \theta^*)) ds \right] \int_0^T U_j(\varrho(s, \theta^*)) ds, \end{aligned} \quad (3.4)$$

and

$$\begin{aligned} 0 &= \frac{\partial \mathcal{G}}{\partial \nu} \Big|_{(\lambda, \nu) = (\lambda^*, \nu^*)} \\ &= -2 \sum_{j=0}^{\infty} \left[ \varrho_j(T, \theta^*) - \varrho_j(0, \theta^*) - \lambda^* \int_0^T U_j(\varrho(s, \theta^*)) ds - \nu^* \int_0^T V_j(\varrho(s, \theta^*)) ds \right] \int_0^T V_j(\varrho(s, \theta^*)) ds. \end{aligned} \quad (3.5)$$

Solving the equations (3.4) and (3.5) leads to

$$\begin{pmatrix} \lambda^* \\ \nu^* \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \frac{1}{a_{11}a_{22} - (a_{12})^2} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad (3.6)$$

where

$$\begin{aligned} a_{11} &:= \sum_{j=0}^{\infty} \left[ \int_0^T U_j(\varrho(s, \theta^*)) ds \right]^2, \\ a_{12} = a_{21} &:= \sum_{j=0}^{\infty} \int_0^T U_j(\varrho(s, \theta^*)) ds \int_0^T V_j(\varrho(s, \theta^*)) ds, \\ a_{22} &:= \sum_{j=0}^{\infty} \left[ \int_0^T V_j(\varrho(s, \theta^*)) ds \right]^2, \\ b_1 &:= \sum_{j=0}^{\infty} [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \int_0^T U_j(\varrho(s, \theta^*)) ds, \end{aligned}$$

and

$$b_2 := \sum_{j=0}^{\infty} [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \int_0^T V_j(\varrho(s, \theta^*)) ds.$$

Note that the right-hand side in (3.6) is well posed only if  $a_{11}a_{22} \neq (a_{12})^2$ . Nevertheless, a simple application of Hölder's inequality tells us that  $a_{11}a_{22} \geq (a_{12})^2$ . Therefore, one needs to investigate the conditions under which the inequality

$$a_{11}a_{22} > (a_{12})^2 \quad (3.7)$$

holds. Below we give as examples two sufficient conditions that ensure the validity of (3.7).



**Lemma 3.1** *Suppose that one of the following conditions holds:*

$$\int_0^T \varrho_1(s) ds > \int_0^T \varrho_0(s) ds > 0, \quad (3.8)$$

or

$$\int_0^T \varrho_0(s) ds > \int_0^T \varrho_1(s) ds > \int_0^T \varrho_2(s) ds > 0. \quad (3.9)$$

Then, the inequality (3.7) holds.

**Proof** See Appendix B. □

**The approximate LSE:** Recall that the data  $D^{N,m}$  are observed on the prelimiting finite  $N$ -system. Hence, using (3.6) we construct the following approximate LSE defined for any  $m \in \mathbb{N}$  and  $N \geq 2$  as

$$\begin{pmatrix} \lambda^{N,m} \\ \nu^{N,m} \end{pmatrix} := \frac{1}{a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2} \begin{pmatrix} a_{22}^{N,m} & -a_{12}^{N,m} \\ -a_{21}^{N,m} & a_{11}^{N,m} \end{pmatrix} \begin{pmatrix} b_1^{N,m} \\ b_2^{N,m} \end{pmatrix}, \quad (3.10)$$

where

$$\begin{aligned} a_{11}^{N,m} &:= \sum_{j=0}^{\infty} \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) \right]^2, \\ a_{12}^{N,m} = a_{21}^{N,m} &:= \sum_{j=0}^{\infty} \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) \right] \left[ \frac{T}{m} \sum_{k=1}^m V_j(\varrho^N(t_k, \theta^*)) \right], \\ a_{22}^{N,m} &:= \sum_{j=0}^{\infty} \left[ \frac{T}{m} \sum_{k=1}^m V_j(\varrho^N(t_k, \theta^*)) \right]^2, \end{aligned}$$

$$b_1^{N,m} := \sum_{j=0}^{\infty} [\varrho_j^N(T, \theta^*) - \varrho_j^N(0, \theta^*)] \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) \right],$$

and

$$b_2^{N,m} := \sum_{j=0}^{\infty} [\varrho_j^N(T, \theta^*) - \varrho_j^N(0, \theta^*)] \left[ \frac{T}{m} \sum_{k=1}^m V_j(\varrho^N(t_k, \theta^*)) \right],$$

with  $t_k = \frac{kT}{m}$  for  $k = 1, 2, \dots, m$ . We will show next that the estimator in (3.10) is consistent and approximately normal.

### 3.3 Consistency of the LSE

We show in this section the consistency of the approximate estimator given in (3.10). To this end, let us first prove the following technical lemma:

**Lemma 3.2** *Let  $\{\mu_n, n \in \mathbb{Z}_+\}$  be a sequence of probability measures on  $\mathbb{Z}_+$ . Then, the following statements are equivalent:*

(i)  $\mu_n$  converges weakly to  $\mu_0$  as  $n \rightarrow \infty$ .

(ii)  $\lim_{n \rightarrow \infty} \mu_n(j) = \mu_0(j)$  for all  $j \in \mathbb{Z}_+$ .

(iii)  $\lim_{n \rightarrow \infty} \sum_{j=0}^{\infty} |\mu_n(j) - \mu_0(j)| = 0$ .

**Proof.** Obviously, (iii) $\Rightarrow$ (i) $\Rightarrow$ (ii). (ii) $\Rightarrow$ (iii) is complete by the observation that, for any  $m \in \mathbb{N}$ ,

$$\begin{aligned} \sum_{j=0}^{\infty} |\mu_n(j) - \mu_0(j)| &\leq \sum_{j=0}^m |\mu_n(j) - \mu_0(j)| + \sum_{j=m+1}^{\infty} \mu_n(j) + \sum_{j=m+1}^{\infty} \mu_0(j) \\ &= \sum_{j=0}^m |\mu_n(j) - \mu_0(j)| + \sum_{j=0}^m [\mu_0(j) - \mu_n(j)] + 2 \sum_{j=m+1}^{\infty} \mu_0(j) \\ &\leq 2 \sum_{j=0}^m |\mu_n(j) - \mu_0(j)| + 2 \sum_{j=m+1}^{\infty} \mu_0(j). \end{aligned}$$

□

We are now ready to prove the consistency of the estimator in (3.10).

**Theorem 3.1** *Suppose that  $\varrho^N(0) \rightarrow \varrho_0$  in  $\mathcal{S}$  as  $N \rightarrow \infty$  and (3.7) holds. Then,*

$$\begin{pmatrix} \lambda^{N,m} \\ \nu^{N,m} \end{pmatrix} \longrightarrow \begin{pmatrix} \lambda^* \\ \nu^* \end{pmatrix} \text{ in probability as } N \rightarrow \infty \text{ and } m \rightarrow \infty.$$

**Proof** See Appendix C.

□

### 3.4 Asymptotic normality of the LSE

We now analyze the asymptotic distribution of the approximate LSE given in (3.10). In particular, one aims to prove the asymptotic normality of

$$\sqrt{N} \left( \begin{pmatrix} \lambda^{N,m} \\ \nu^{N,m} \end{pmatrix} - \begin{pmatrix} \lambda^* \\ \nu^* \end{pmatrix} \right), \quad (3.11)$$

as the network size  $N$  and the number of observations  $m$  go to infinity. To do so, we will prove that (3.11) converges towards a linear combination of the process  $\mathcal{Z}$  which is itself a Gaussian process by Proposition 2.1.

To ease of notations, define the following quantities:

$$\begin{aligned}
\mathcal{I} = \sum_{j=0}^{\infty} & \left\{ a_{22} \left( [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right. \right. \\
& \left. \left. + [\mathcal{Z}_j(T, \theta^*) - \mathcal{Z}_j(0, \theta^*)] \int_0^T U_j(\varrho(s)) ds \right) \right. \\
& + 2b_1 \int_0^T V_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i V_j(s, \theta^*) \mathcal{Z}_i(s, \theta^*) ds \right] \\
& - b_2 \left( \int_0^T V_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right. \\
& \left. + \int_0^T U_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i V_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right) \\
& - a_{12} \left( [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i V_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right. \\
& \left. + [\mathcal{Z}_j(T, \theta^*) - \mathcal{Z}_j(0, \theta^*)] \int_0^T V_j(\varrho(s)) ds \right) \left. \right\},
\end{aligned}$$

$$\begin{aligned}
\mathcal{J} = \sum_{j=0}^{\infty} & \left\{ a_{11} \left( [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i V_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right. \right. \\
& \left. \left. + [\mathcal{Z}_j(T, \theta^*) - \mathcal{Z}_j(0, \theta^*)] \int_0^T V_j(\varrho(s)) ds \right) \right. \\
& + 2b_2 \int_0^T U_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i U_j(s, \theta^*) \mathcal{Z}_i(s, \theta^*) ds \right] \\
& - b_1 \left( \int_0^T V_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right. \\
& \left. + \int_0^T U_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i V_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right) \\
& - a_{12} \left( [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right. \\
& \left. + [\mathcal{Z}_j(T, \theta^*) - \mathcal{Z}_j(0, \theta^*)] \int_0^T U_j(\varrho(s)) ds \right) \left. \right\},
\end{aligned}$$

and

$$\begin{aligned} \mathcal{K} = 2 \sum_{j=0}^{\infty} \left\{ & -a_{11} \int_0^T V_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i V_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right. \\ & - a_{22} \int_0^T U_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \\ & + a_{12} \left( \int_0^T V_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right. \\ & \left. \left. + \int_0^T U_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i V_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right) \right\}. \end{aligned}$$

One can observe that  $\mathcal{I}$ ,  $\mathcal{J}$ ,  $\mathcal{K}$  are linear combinations of the Gaussian process  $\mathcal{Z}(t)$  solution to the SDE (2.5) (see Proposition 2.1).

The following theorem, states the asymptotic normality of the approximate LSE (3.11) under the assumption that  $\frac{m}{\sqrt{N}} \rightarrow \infty$ .

**Theorem 3.2** *Suppose that (3.7) holds. Then, under the assumption of Theorem 2.2, as  $N, m, \frac{m}{\sqrt{N}} \rightarrow \infty$ ,*

$$\sqrt{N} \left( \begin{pmatrix} \lambda^{N,m} \\ \nu^{N,m} \end{pmatrix} - \begin{pmatrix} \lambda^* \\ \nu^* \end{pmatrix} \right) \xrightarrow{d} \frac{1}{[a_{11}a_{22} - (a_{12})^2]^2} \begin{pmatrix} [a_{11}a_{22} - (a_{12})^2]\mathcal{I} + (a_{22}b_1 - a_{12}b_2)\mathcal{K} \\ [a_{11}a_{22} - (a_{12})^2]\mathcal{J} + (a_{11}b_2 - a_{12}b_1)\mathcal{K} \end{pmatrix}.$$

**Proof** See Appendix D. □

**Remark 2** *By Proposition 2.1, we know that the limiting distribution given by Theorem 3.2 is normal. Moreover, explicit (but tedious) expressions of the mean and the covariance matrix of the limiting normal distribution can be obtained by virtue of (2.6) and (A.5).*

## 4 Numerical experiments

In this section, we evaluate the consistency and asymptotic normality of the approximate LSE  $\hat{\theta} = (\lambda^{N,m}, \nu^{N,m})$  defined in (3.10) using simulated data. Specifically, we aim to validate the assertions made in Theorem 3.1 and Theorem 3.2. To this end, we propose to generate the datasets  $D^{N,m} = \{\varrho^N(t_k) : 1 \leq k \leq m\}$  by simulating the power of two choices model, with true parameters  $(\theta = (\lambda, \nu) = (0.5, 1))$ , for various network sizes  $N \in \mathbb{N}$  and different numbers of observations  $m \in \mathbb{N}$ . For each combination of  $(N, m)$ , we simulate a total of 100 samples which are then utilized to estimate the arrival and service rates  $\theta = (\lambda, \nu)$  using the approximate LSE  $\hat{\theta} = (\lambda^{N,m}, \nu^{N,m})$ . The estimated values are plotted in Figure 1 and Figure 2.

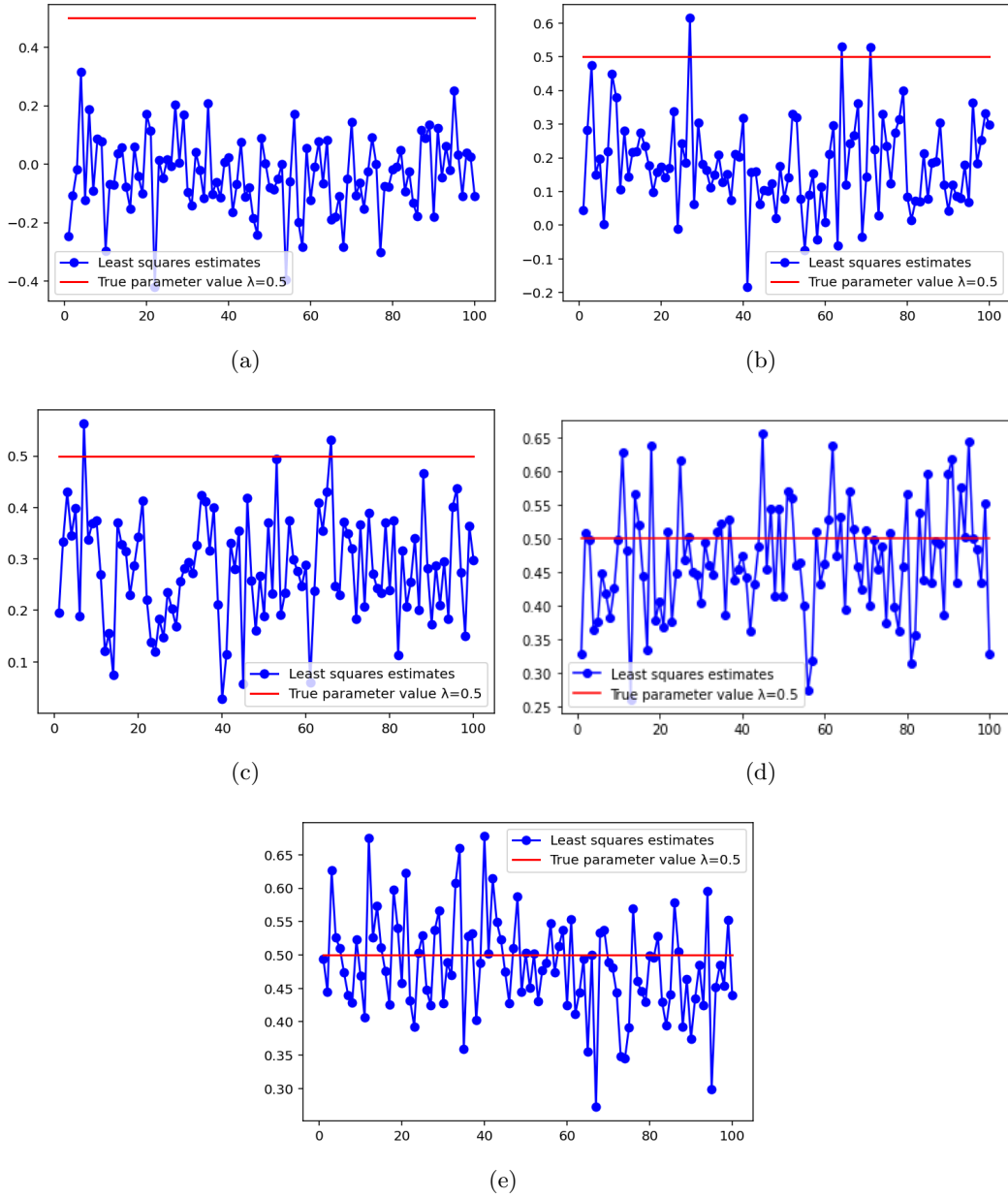


Figure 1: Values of 100 Least squares estimates of the parameter  $\lambda = 0.5$  with different network sizes  $N$  and number of observations  $m$ : (a)  $N=100$ ,  $m=1000$ ; (b)  $N=500$ ,  $m=10000$ ; (c)  $N=1000$ ,  $m=10000$ ; (d)  $N=2000$ ,  $m=20000$ ; (e)  $N=3000$ ,  $m=30000$

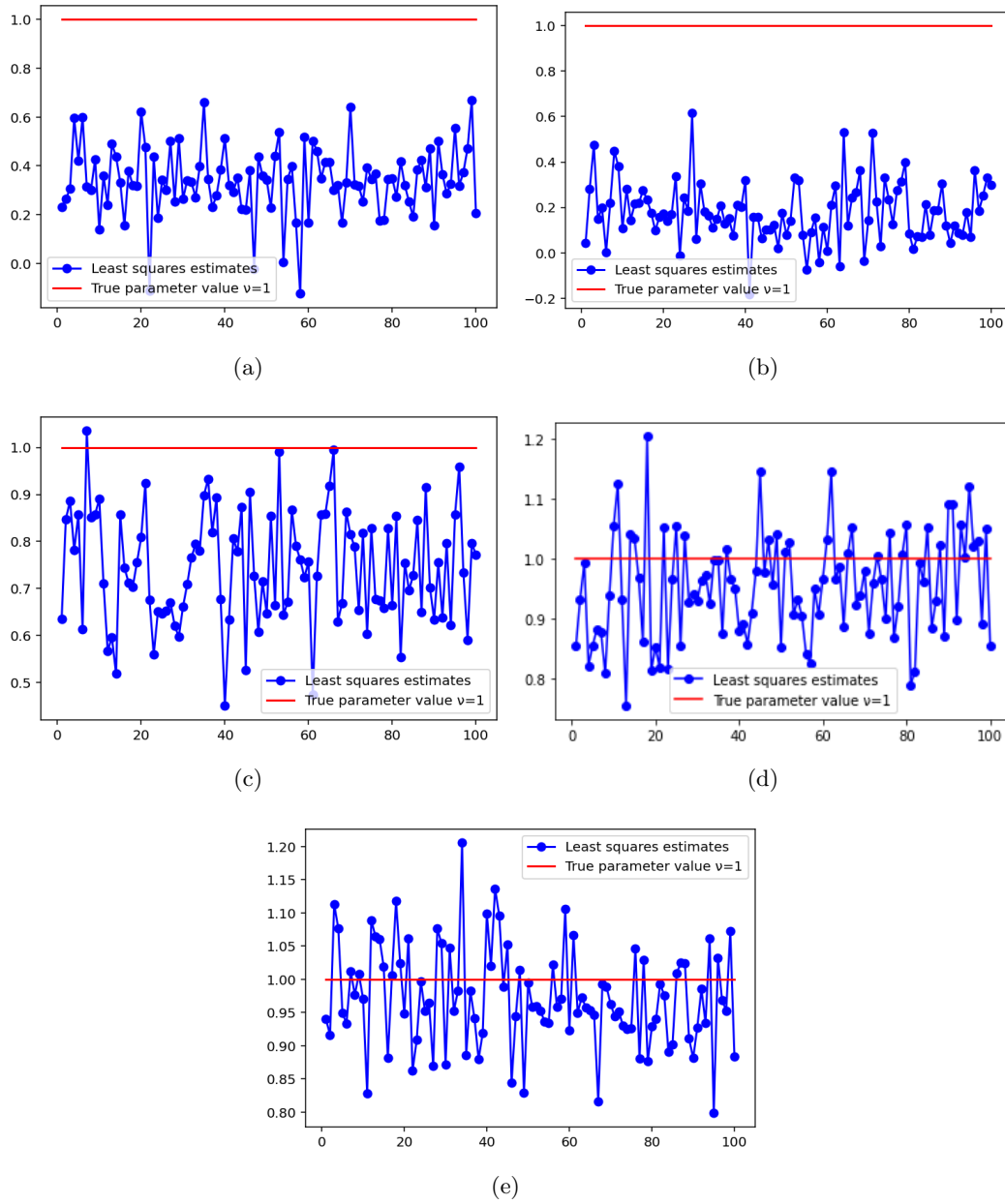


Figure 2: Values of 100 Least squares estimates of the parameter  $\nu = 1$  with different network sizes  $N$  and number of observations  $m$ : (a)  $N=100, m=1000$ ; (b)  $N=500, m=10000$ ; (c)  $N=1000, m=10000$ ; (d)  $N=2000, m=20000$ ; (e)  $N=3000, m=30000$

## 4.1 Consistency of the estimator

We aim to numerically assess the consistency of the estimator  $\hat{\theta} = (\lambda^{N,m}, \nu^{N,m})$ . To achieve this, we utilize the estimated values of  $\hat{\theta} = (\lambda^{N,m}, \nu^{N,m})$  obtained from simulated datasets for various values of  $N$  and  $m$  to calculate the following empirical moments:

- The empirical mean

$$\bar{\hat{\theta}} = (\overline{\lambda^{N,m}}, \overline{\nu^{N,m}}) = \frac{1}{100} \left( \sum_{i=1}^{100} \lambda_i^{N,m}, \sum_{i=1}^{100} \nu_i^{N,m} \right) \approx \mathbb{E}(\lambda^{N,m}, \nu^{N,m}).$$

- The empirical standard deviation

$$s_{\hat{\theta}} = (s_{\lambda^{N,m}}, s_{\nu^{N,m}}) = \sqrt{\frac{1}{99} \left( \sum_{i=1}^{100} (\lambda_i^{N,m} - \overline{\lambda^{N,m}})^2, \sum_{i=1}^{100} (\nu_i^{N,m} - \overline{\nu^{N,m}})^2 \right)} \approx \sqrt{\mathbb{V}(\lambda^{N,m}, \nu^{N,m})}.$$

- The empirical mean square error

$$MSE = \frac{1}{100} \left( \sum_{i=1}^{100} (\lambda_i^{N,m} - \lambda)^2, \sum_{i=1}^{100} (\nu_i^{N,m} - \nu)^2 \right) \approx \mathbb{E}((\lambda^{N,m} - \lambda)^2, (\nu^{N,m} - \nu)^2).$$

- The empirical mean error

$$\text{Mean - Error} = \frac{1}{100} \left( \sum_{i=1}^{100} (\lambda_i^{N,m} - \lambda), \sum_{i=1}^{100} (\nu_i^{N,m} - \nu) \right) \approx \mathbb{E}((\lambda^{N,m} - \lambda), (\nu^{N,m} - \nu)).$$

The results are presented in Table 1. As observed, as  $N$  and  $m$  increase, the empirical mean of the estimator converges to the true parameter values, while the corresponding standard deviations decrease. Furthermore, both the mean squared error and the absolute mean error diminish as  $N$  and  $m$  grow larger. These findings validate the consistency of the estimator, as stated in Theorem 3.1.

| Parameters moments    | $\hat{\theta} = (\overline{\lambda^{N,m}}, \overline{\nu^{N,m}})$ | $s_{\hat{\theta}} = (s_{\lambda^{N,m}}, s_{\nu^{N,m}})$ | MSE            | Mean Error     |
|-----------------------|---|---|----------------|----------------|
| $N = 100, m = 1000$   | (-0.03, 0.33)   | (0.13, 0.14)  | (0.30, 0.45)   | (-0.53, -0.66) |
| $N = 500, m = 10000$  | (0.18, 0.61)  | (0.13, 0.15)  | (0.12, 0.17)   | (-0.31, -0.38) |
| $N = 1000, m = 10000$ | 0.28, 0.74)   | (0.10, 0.11)  | (0.05, 0.07)   | (-0.21, -0.25) |
| $N = 2000, m = 20000$ | (0.46, 0.95)  | (0.08, 0.08)  | (0.008, 0.009) | (-0.03, -0.04) |
| $N = 3000, m = 30000$ | (0.48, 0.97)  | (0.07, 0.07)  | (0.005, 0.006) | (-0.01, -0.02) |

Table 1: Empirical moments of the approximate LSE  $\theta^{N,m} = (\lambda^{N,m}, \nu^{N,m})$

### 4.1.1 Asymptotic normality

In this section, we assess the asymptotic normality of the approximate LSE  $\hat{\theta} = (\lambda^{N,m}, \nu^{N,m})$  as established in Theorem 3.2. Specifically, we numerically verify that the normalized error term  $\sqrt{N}((\lambda^{N,m} - \lambda), (\nu^{N,m} - \nu))$  converges to a Gaussian distribution as  $N$  and  $m$  approach infinity.

We utilize the simulated 100 samples from the power of two choices model with the true parameter  $\theta = (\lambda, \nu) = (0.5, 1)$ , varying the network size  $N$  and the number of observations  $m$ . For each combination of  $(N, m)$ , we compute the first four empirical moments of the normalized error terms  $\sqrt{N}((\lambda^{N,m} - \lambda), (\nu^{N,m} - \nu))$ : the mean, variance, skewness, and kurtosis. The results are summarized in Table 2. Notably, we observe that the skewness and kurtosis values tend to approximate those of a normal distribution (0 for skewness and 3 for kurtosis), even for relatively small values of network size  $N$  and number of observations  $m$ .

| Normalized error moments | Mean           | Variance       | Skewness       | Kurtosis     |
|--------------------------|----------------|----------------|----------------|--------------|
| N=100,m=1000             | (-5.37, -6.60) | (1.75, 2.07)   | (-0.16, -0.41) | (3.44, 4.31) |
| N=500,m=10000            | (-7.15, -8.65) | (8.99, 11.32)  | (0.51, 0.60)   | (3.90, 4.21) |
| N=1000,m=10000           | (-6.89, -8.08) | (11.08, 14.38) | (0.01, 0.02)   | (2.72, 2.51) |
| N=2000, m=20000          | (-1.47, -2.01) | (13.96, 15.59) | (0.07, 0.20)   | (2.80, 2.68) |
| N=3000, m=30000          | (-0.86, -1.48) | (17.00, 17.15) | (0.10, 0.26)   | (3.40, 3.07) |

Table 2: Empirical moments of the normalized errors  $\sqrt{N}((\lambda^{N,m} - \lambda), (\nu^{N,m} - \nu))$

To further substantiate our findings, we test the normality of the normalized error terms using a Kolmogorov-Smirnov test. We conduct this test on the 100 simulated datasets across the different values of network size  $N$  and number of observations  $m$ . The resulting  $p$ -values are presented in Table 3. As shown, the  $p$ -values are sufficiently large, suggesting that the null hypothesis asserting that the error terms  $\sqrt{N}((\lambda^{N,m} - \lambda), (\nu^{N,m} - \nu))$  follow a normal distribution is not rejected, even for smaller values of  $N$  and  $m$ . This result indicates that the error terms tend to the normal distribution quickly. Nevertheless, it is also noted that the  $p$ -values are high for all  $(N, m)$  combinations, and they do not necessarily increase with larger network sizes  $N$  and higher numbers of observations  $m$ . This effect may be attributed to the fact that the true mean and variance of the error terms are unknown, requiring the use of empirical values, which may account for the observed outcomes.

| Network and data sizes $(N, m)$ | P-value for $\sqrt{N}(\lambda^{N,m} - \lambda)$ | P-value for $\sqrt{N}(\nu^{N,m} - \nu)$ |
|---------------------------------|---|---|
| $(N, m) = (100, 1000)$          | 0.71  | 0.48                                    |
| $(N, m) = (500, 10000)$         | 0.55  | 0.59                                    |
| $(N, m) = (1000, 10000)$        | 0.99  | 0.33                                    |
| $(N, m) = (2000, 20000)$        | 0.97  | 0.94                                    |
| $(N, m) = (3000, 30000)$        | 0.65  | 0.66                                    |

Table 3: Kolmogorov-Smirnov tests for the normalized parameter estimator errors  $\sqrt{N}((\lambda^{N,m} - \lambda), (\nu^{N,m} - \nu))$



Finally, we plot the histograms of the normalized error terms  $\sqrt{N}((\lambda^{N,m} - \lambda))$  and  $\sqrt{N}((\nu^{N,m} - \nu))$ , along with a kernel density estimator. The results are shown in Figure 3 and Figure 4. Once again, we observe that the assertion of normality for the error terms aligns well with the empirical data.

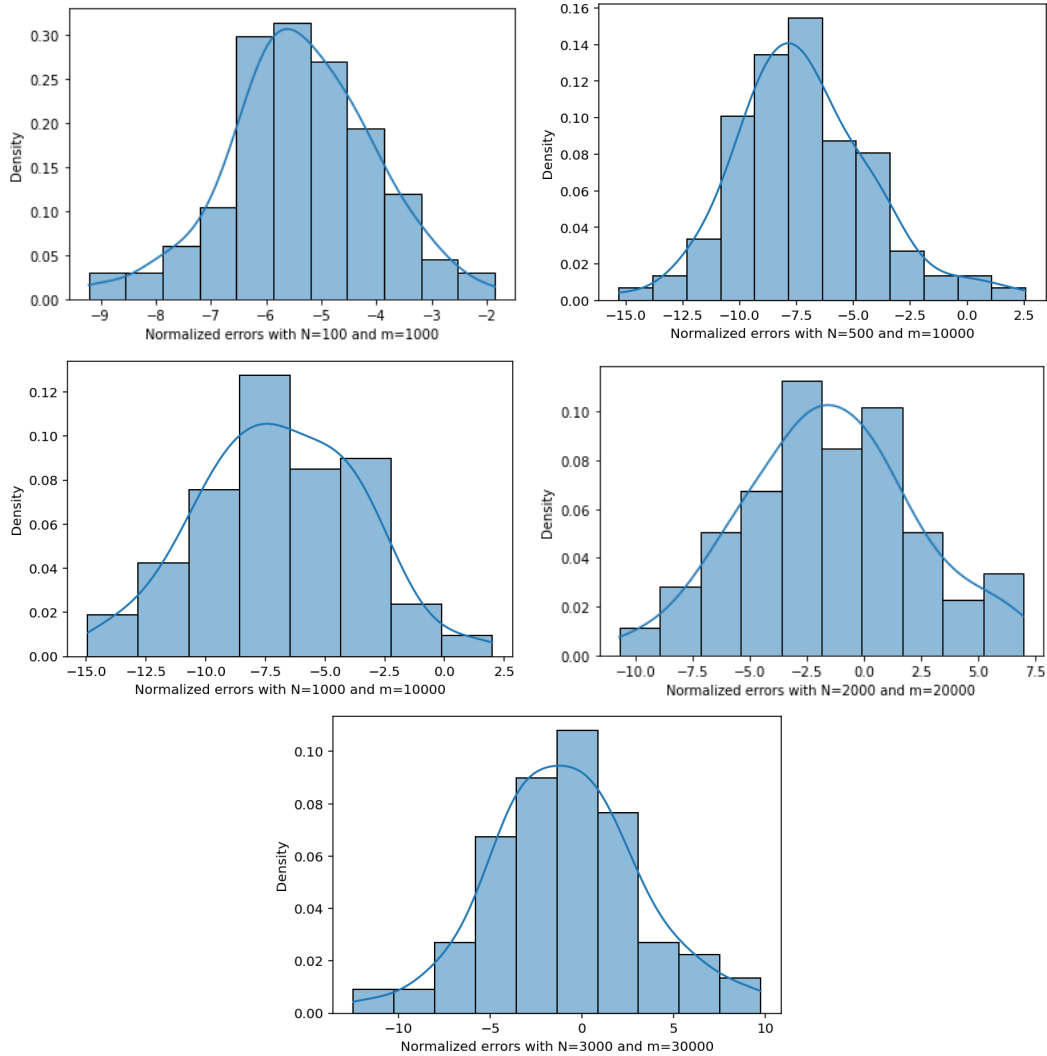


Figure 3: Histograms of the normalized errors  $\sqrt{N}(\lambda^{N,m} - \lambda)$  with the associated kernel density estimate plots for different values of the network size  $N$  and the number of observations  $m$  with the true parameters  $\theta = (\lambda, \nu) = (0.5, 1)$

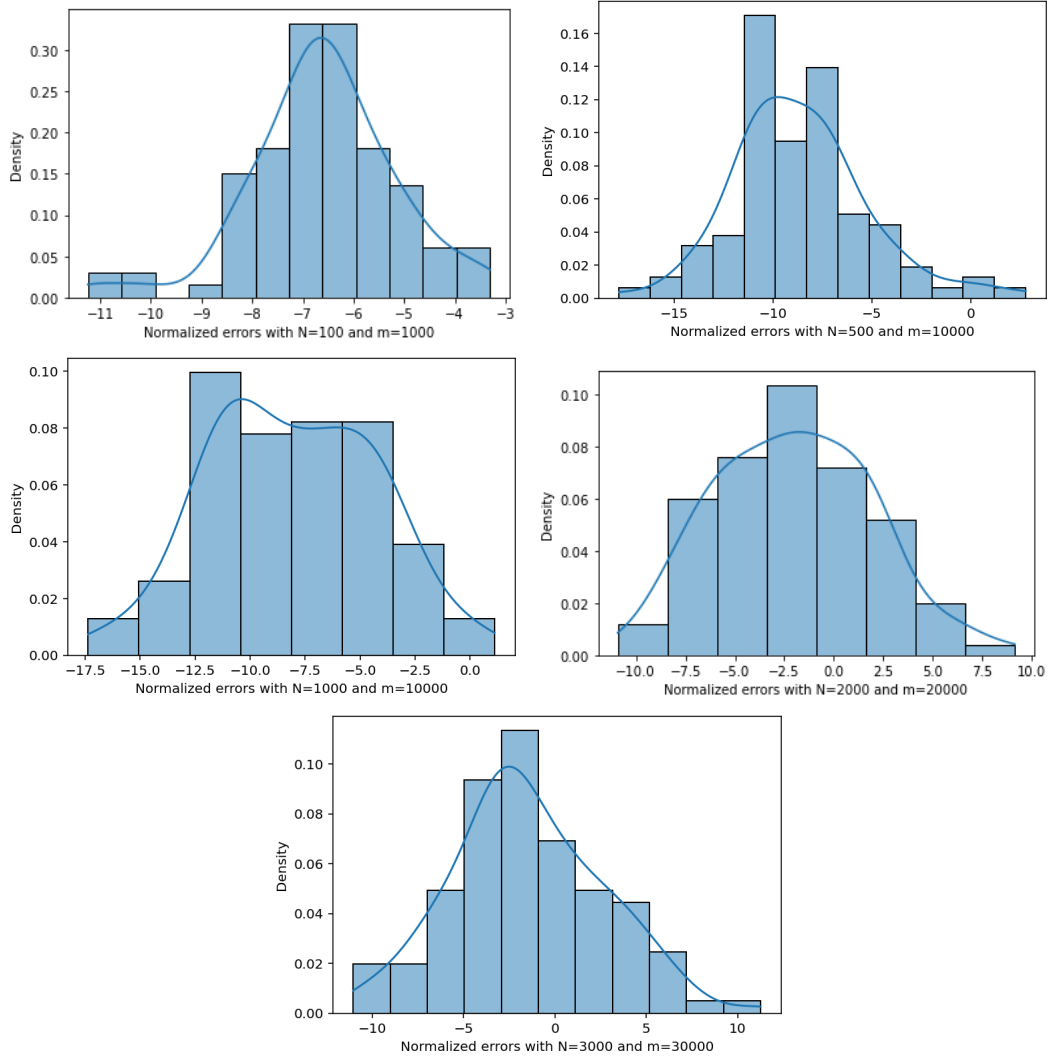


Figure 4: Histograms of the normalized errors  $\sqrt{N}(\nu^{N,m} - \nu)$  with the associated kernel density estimate plots for different values of the network size  $N$  and the number of observations  $m$ , corresponding to true parameters  $\theta = (\lambda, \nu) = (0.5, 1)$

## 5 Conclusions and perspectives

In this paper, we considered the parameter estimation problem of the supermarket model. Based on an aggregate dataset, we constructed an approximate LSE by exploiting the law of large numbers together with the central limit theorem established for the model in the literature. Moreover, we proved the consistency together with the asymptotic normality of the estimator as both the size of the network and the number of observations go to infinity. Finally, we presented a numerical study where we tested

our estimator against synthetic data obtained by simulating the power-of-two model highlighting our theoretical results.

The current work is the first statistical scheme for mean-field queuing systems and opens a new perspective. One naturally aims to investigate the statistical inference problem for other models. For instance, one can investigate the approximate LSE approach to the model proposed in [6] for load balancing mechanisms in cloud storage systems which is a generalization of the supermarket model. The established law of large numbers together with the central limit theorem make the approximate LSE approach used in the current work conceivable, provided that one can overcome the technical difficulty arising from the more complicated mean-field limiting equation. Another variation of the supermarket model for which one can study the parameter estimation problem is the one introduced in [7] in which the servers can communicate with their neighbors and where the neighborhood relationships are described in terms of a suitable graph. Again, the limit as the number of servers goes to infinity was identified, which can be exploited to build a statistical scheme, however, no central limit theorem nor a stationary distribution was established, therefore, the asymptotic normality of the estimator cannot be obtained by the similar scheme used in the current paper. Another interesting open problem is the nonparametric estimation of the interaction kernel in general mean-field queuing systems studied in [11]. Indeed, one can consider the exploitation of the limiting mean-field equation to build an estimator. However, contrary to our current proposal where the unknown parameters enter linearly in the mean-field limiting equation, in the nonparametric estimation one needs to deal with an optimization problem in function space. A potential avenue is to exploit the stationary distribution to build an estimator in the stationary regime and then investigate a justification for the interchange of limits  $N \rightarrow \infty$  and  $t \rightarrow \infty$ .

Finally, the problem of statistical inference for general mean-field models on discrete space remains open and, as mentioned in the introduction, very few references exist.

## A Proof of Proposition 2.1

Let  $\xi(t) = (\xi_j(t))^T$  be the solution to the following infinite-dimensional ODE:

$$d\xi(t) = G(M(t) + \xi(t), \varrho(t)) dt, \quad \xi(0) = z_0 \in \tilde{\ell}_2.$$

Define  $\mathcal{Z}(t) := M(t) + \xi(t)$ . Then,  $\mathcal{Z}(t)$  satisfies the SDE:

$$d\mathcal{Z}(t) = G(\mathcal{Z}(t), \varrho(t))dt + a(t)dW(t), \quad \mathcal{Z}(0) = z_0.$$

By [6, Proposition 2], we know that  $\mathcal{Z}(t) \in \tilde{\ell}_2$  for all  $t \in [0, T]$  almost surely. By the estimate [6, Page 69, line -1] and Fatou's lemma (cf. [6, Page 78, Proof of Theorem 2]), we get

$$E \left[ \sup_{t \in [0, T]} \sum_{m=0}^{\infty} (m+1)^2 M_m^2(t) \right] < \infty.$$

Then,  $M(t), \xi(t) \in \ell_1$  for all  $t \in [0, T]$  almost surely. Moreover, for any  $j \in \mathbb{Z}_+$ , we have  $\xi_j(0) = (z_0)_j$ , and

$$\begin{aligned}
\frac{d\xi_j(t)}{dt} &= G_j(\mathcal{Z}(t), \varrho(t)) \\
&= 2\lambda \sum_{m=j}^{\infty} \{ [M_{j-1}(t) + \xi_{j-1}(t)]\varrho_m(t) + [M_m(t) + \xi_m(t)]\varrho_{j-1}(t) \\
&\quad - [M_j(t) + \xi_j(t)]\varrho_{m+1}(t) - [M_{m+1}(t) + \xi_{m+1}(t)]\varrho_j(t) \} \\
&\quad + \nu \{ [M_{j+1}(t) + \xi_{j+1}(t)] - [M_j(t) + \xi_j(t)] \} \\
&= 2\lambda [M_{j-1}(t) - M_j(t)] \sum_{m=j}^{\infty} \varrho_m(t) + 2\lambda [\xi_{j-1}(t) - \xi_j(t)] \sum_{m=j}^{\infty} \varrho_m(t) \\
&\quad + 2\lambda [\varrho_{j-1}(t) - \varrho_j(t)] \sum_{m=j}^{\infty} [M_m(t) + \xi_m(t)] + 2\lambda M_j(t)\varrho_j(t) + 2\lambda \xi_j(t)\varrho_j(t) + 2\lambda \varrho_j(t)[M_j(t) + \xi_j(t)] \\
&\quad + \nu \{ [M_{j+1}(t) - M_j(t)] + [\xi_{j+1}(t) - \xi_j(t)] \} \\
&= 2\lambda [M_{j-1}(t) - M_j(t)] \sum_{m=j}^{\infty} \varrho_m(t) + 2\lambda [\varrho_{j-1}(t) - \varrho_j(t)] \sum_{m=j}^{\infty} M_m(t) + \nu [M_{j+1}(t) - M_j(t)] \\
&\quad + 4\lambda M_j(t)\varrho_j(t) + 4\lambda \xi_j(t)\varrho_j(t) \\
&\quad + 2\lambda [\varrho_{j-1}(t) - \varrho_j(t)] \sum_{m=j}^{\infty} \xi_m(t) + 2\lambda [\xi_{j-1}(t) - \xi_j(t)] \sum_{m=j}^{\infty} \varrho_m(t) + \nu [\xi_{j+1}(t) - \xi_j(t)] \\
&= 2\lambda \left( \sum_{m=j}^{\infty} \varrho_m(t) \right) [M_{j-1}(t) - M_j(t)] + 2\lambda [\varrho_{j-1}(t) - \varrho_j(t)] \sum_{m=j}^{\infty} M_m(t) + 4\lambda \varrho_j(t)M_j(t) + \nu [M_{j+1}(t) - M_j(t)] \\
&\quad + 2\lambda [\varrho_{j-1}(t) - \varrho_j(t)] \sum_{m=j+2}^{\infty} \xi_m(t) + \{ 2\lambda [\varrho_{j-1}(t) - \varrho_j(t)] + \nu \} \xi_{j+1}(t) \\
&\quad + \left\{ 2\lambda \left[ \varrho_{j-1}(t) + \varrho_j(t) - \sum_{m=j}^{\infty} \varrho_m(t) \right] - \nu \right\} \xi_j(t) + 2\lambda \left( \sum_{m=j}^{\infty} \varrho_m(t) \right) \xi_{j-1}(t). \tag{A.1}
\end{aligned}$$

Since  $\sum_{m=0}^{\infty} \mathcal{Z}_m(t) = 0$ , we get

$$\sum_{m=j+2}^{\infty} \xi_m(t) = - \sum_{m=0}^{\infty} M_m(t) - \sum_{m=0}^{j+1} \xi_m(t).$$

Similarly, by  $\sum_{m=0}^{\infty} \varrho_m(t) = 1$ , we get

$$\sum_{m=j}^{\infty} \varrho_m(t) = 1 - \sum_{m=0}^{j-1} \varrho_m(t).$$

Then, we can rewrite (A.1) as follows:

$$\frac{d\xi_j(t)}{dt} = 2\lambda \left[ 1 - \sum_{m=0}^{j-1} \varrho_m(t) \right] [M_{j-1}(t) - M_j(t)] - 2\lambda [\varrho_{j-1}(t) - \varrho_j(t)] \sum_{m=0}^{j-1} M_m(t) + 4\lambda \varrho_j(t)M_j(t) + \nu [M_{j+1}(t) - M_j(t)]$$

$$\begin{aligned}
& -2\lambda[\varrho_{j-1}(t) - \varrho_j(t)] \sum_{m=0}^{j-2} \xi_m(t) + 2\lambda \left[ 1 - \sum_{m=0}^{j-2} \varrho_m(t) - 2\varrho_{j-1}(t) + \varrho_j(t) \right] \xi_{j-1}(t) \\
& + \left\{ 2\lambda \left[ -1 + \sum_{m=0}^{j-1} \varrho_m(t) + 2\varrho_j(t) \right] - \nu \right\} \xi_j(t) + \nu \xi_{j+1}(t).
\end{aligned} \tag{A.2}$$

Note that (A.2) can be regarded as an infinite-dimensional non-autonomous linear system of ODEs with random coefficients. Define  $g(t) = (g_j(t))_{j=0}^\infty$  and  $C(t) = (C_{jl}(t))_{j,l=0}^\infty$  by

$$\begin{aligned}
g_j(t) & := 2\lambda \left[ 1 - \sum_{m=0}^{j-1} \varrho_m(t) \right] [M_{j-1}(t) - M_j(t)] - 2\lambda[\varrho_{j-1}(t) - \varrho_j(t)] \sum_{m=0}^{j-1} M_m(t) \\
& \quad + 4\lambda\varrho_j(t)M_j(t) + \nu[M_{j+1}(t) - M_j(t)] \\
& = -2\lambda[\varrho_{j-1}(t) - \varrho_j(t)] \sum_{m=0}^{j-2} M_m(t) + 2\lambda \left[ 1 - \sum_{m=0}^{j-2} \varrho_m(t) - 2\varrho_{j-1}(t) + \varrho_j(t) \right] M_{j-1}(t) \\
& \quad + \left\{ 2\lambda \left[ -1 + \sum_{m=0}^{j-1} \varrho_m(t) + 2\varrho_j(t) \right] - \nu \right\} M_j(t) + \nu M_{j+1}(t),
\end{aligned} \tag{A.3}$$

and

$$C_{jl}(t) := \begin{cases} -2\lambda[\varrho_{j-1}(t) - \varrho_j(t)], & \text{if } 0 \leq l \leq j-2, \\ 2\lambda \left[ 1 - \sum_{m=0}^{j-2} \varrho_m(t) - 2\varrho_{j-1}(t) + \varrho_j(t) \right], & \text{if } l = j-1, \\ 2\lambda \left[ -1 + \sum_{m=0}^{j-1} \varrho_m(t) + 2\varrho_j(t) \right] - \nu, & \text{if } l = j, \\ \nu, & \text{if } l = j+1, \\ 0, & \text{if } l \geq j+2. \end{cases} \tag{A.4}$$

Then, (A.2) becomes

$$\frac{d\xi(t)}{dt} = C(t)\xi(t) + g(t), \quad \xi(0) = z_0.$$

Denote  $\mathcal{C} = 6\lambda + \nu$ . Therefore, for any  $j \in \mathbb{Z}_+$  and  $t \in [0, T]$ , by (A.3), we get

$$|g_j(t)| \leq \mathcal{C} \left\{ |M_{j-1}(t)| + |M_j(t)| + |M_{j+1}(t)| + [\varrho_{j-1}(t) + \varrho_j(t)] \left[ \frac{\pi^2}{6} \sup_{t \in [0, T]} \sum_{m=0}^{\infty} (m+1)^2 M_m^2(t) \right]^{\frac{1}{2}} \right\}.$$

Then,

$$\sup_{t \in [0, T]} \|g(t)\|_1 \leq 5\mathcal{C} \left[ \frac{\pi^2}{6} \sup_{t \in [0, T]} \sum_{m=0}^{\infty} (m+1)^2 M_m^2(t) \right]^{\frac{1}{2}}.$$

For  $x \in \ell_1$ ,  $j \in \mathbb{Z}_+$  and  $t \in [0, T]$ , by (A.4), we get

$$|(C(t)x)_j| \leq \mathcal{C} \left\{ [\varrho_{j-1}(t) + \varrho_j(t)] \sum_{l=0}^{j-2} |x_l| + |x_{j-1}| + |x_j| + |x_{j+1}| \right\}.$$

Then,

$$\|C(t)x\|_1 \leq 5\mathcal{C}\|x\|_1.$$

Hence, by induction, we obtain that

$$\|[C(t)]^n x\|_1 \leq (5\mathcal{C})^n \|x\|_1, \quad n \in \mathbb{N}.$$

Thus, we have the following explicit expressions:

$$\xi(t) = e^{\int_0^t C(s)ds} z_0 + \int_0^t e^{\int_s^t C(u)du} g(s) ds, \quad \mathcal{Z}(t) = M(t) + \xi(t). \quad (\text{A.5})$$

Since  $M(t)$  is a Gaussian martingale, by (A.3), we deduce that the distributions of  $\xi(t)$  and  $\mathcal{Z}(t)$  are both Gaussian. The proof is complete.

## B Proof of Lemma 3.1

By Hölder's inequality, we find that  $a_{11}a_{22} \geq (a_{12})^2$  and the equality sign holds if and only if  $a_{11} = 0$ , or  $a_{22} = 0$ , or

$$a_{11}, a_{22} > 0 \quad \text{and}$$

for all  $j' \in \mathbb{Z}_+$  (see, e.g. [33])

$$\frac{\int_0^T U_{j'}(\varrho(s, \theta^*)) ds}{\left( \sum_{j=0}^{\infty} \left[ \int_0^T U_j(\varrho(s, \theta^*)) ds \right]^2 \right)^{\frac{1}{2}}} = \frac{\int_0^T V_{j'}(\varrho(s, \theta^*)) ds}{\left( \sum_{j=0}^{\infty} \left[ \int_0^T V_j(\varrho(s, \theta^*)) ds \right]^2 \right)^{\frac{1}{2}}}. \quad (\text{B.1})$$

Note that for  $s \in [0, T]$ ,

$$U_0(\varrho(s, \theta^*)) = -2\varrho_0(s, \theta^*) \sum_{i=1}^{\infty} \varrho_i(s, \theta^*) - (\varrho_0(s, \theta^*))^2,$$

and

$$U_1(\varrho(s, \theta^*)) = 2\varrho_0(s, \theta^*) \sum_{i=1}^{\infty} \varrho_i(s, \theta^*) - 2\varrho_1(s, \theta^*) \sum_{i=2}^{\infty} \varrho_i(s, \theta^*) + (\varrho_0(s, \theta^*))^2 - (\varrho_1(s, \theta^*))^2,$$

Then, by the fact that  $\sum_{i=0}^{\infty} \varrho_i(s, \theta^*) = 1$ , we get

$$U_0(\varrho(s, \theta^*)) = -2\varrho_0(s, \theta^*)(1 - \varrho_0(s, \theta^*)) - (\varrho_0(s, \theta^*))^2 = \varrho_0(s, \theta^*)(\varrho_0(s, \theta^*) - 2),$$

and

$$\begin{aligned} U_1(\varrho(s, \theta^*)) &= 2\varrho_0(s, \theta^*)(1 - \varrho_0(s, \theta^*)) - 2\varrho_1(s, \theta^*)(1 - \varrho_0(s, \theta^*) - \varrho_1(s, \theta^*)) + (\varrho_0(s, \theta^*))^2 - (\varrho_1(s, \theta^*))^2 \\ &= (\varrho_0(s, \theta^*) - \varrho_1(s, \theta^*))(2 - \varrho_0(s, \theta^*)) + (\varrho_1(s, \theta^*))(\varrho_0(s, \theta^*) + \varrho_1(s, \theta^*)). \end{aligned}$$

We have

$$V_0(\varrho(s, \theta^*)) = \varrho_1(s, \theta^*) - \varrho_0(s, \theta^*),$$

and

$$V_1(\varrho(s, \theta^*)) = \varrho_2(s, \theta^*) - \varrho_1(s, \theta^*).$$

(a) Suppose that (3.8) holds. Then,  $\int_0^T U_0(\varrho(s, \theta^*))ds < 0$  and  $\int_0^T V_0(\varrho(s, \theta^*))ds > 0$ . Thus, (B.1) cannot hold and hence (3.7) holds.

(b) Suppose that (3.9) holds. Then,  $\int_0^T U_1(\varrho(s, \theta^*))ds > 0$  and  $\int_0^T V_1(\varrho(s, \theta^*))ds < 0$ . Thus, (B.1) cannot hold and hence (3.7) holds.

## C Proof of Theorem 3.1

By Theorem 2.1, [15, (5.7), page 117 and Proposition 5.3, page 119] and Lemma 3.2, we obtain that

$$\sup_{t \in [0, T]} \sum_{j=0}^{\infty} |\varrho_j^N(t, \theta^*) - \varrho_j(t, \theta^*)| \rightarrow 0 \quad (\text{C.1})$$

in probability as  $N \rightarrow \infty$ . Moreover,

$$\begin{aligned} |a_{11}^{N,m} - a_{11}| &\leq \sum_{j=0}^{\infty} \left| \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) \right]^2 - \left[ \int_0^T U_j(\varrho(s))ds \right]^2 \right| \\ &\leq \sup_{j \in \mathbb{Z}_+} \left\{ \left| \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) + \int_0^T U_j(\varrho(s))ds \right| \right\} \\ &\quad \times \sum_{j=0}^{\infty} \left| \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) - \int_0^T U_j(\varrho(s))ds \right| \\ &\leq T \sup_{j \in \mathbb{Z}_+} \left\{ \sup_{t \in [0, T]} |U_j(\varrho^N(t, \theta^*))| + \sup_{t \in [0, T]} |U_j(\varrho(t, \theta^*))| \right\} \\ &\quad \times \sum_{j=0}^{\infty} \left\{ \left| \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) - \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) \right| \right. \\ &\quad \left. + \left| \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) - \int_0^T U_j(\varrho(s))ds \right| \right\}. \end{aligned}$$

By (3.1), we get  $|U_j(x)| \leq 6\|x\|_1$  for all  $x \in \ell_1$  and  $j \in \mathbb{Z}_+$ . Then

$$\begin{aligned} |a_{11}^{N,m} - a_{11}| &\leq 12T \left\{ T \sup_{t \in [0, T]} \sum_{j=0}^{\infty} |U_j(\varrho^N(t, \theta^*)) - U_j(\varrho(t, \theta^*))| \right. \\ &\quad \left. + \sum_{j=0}^{\infty} \left| \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) - \int_0^T U_j(\varrho(s))ds \right| \right\}. \end{aligned}$$

Therefore, by (C.1), we get that the right hand side of the last inequality goes to 0 in probability as  $N, m \rightarrow \infty$ . Similarly, we can show that

$$|a_{12}^{N,m} - a_{12}|, |a_{22}^{N,m} - a_{22}|, |b_1^{N,m} - b_1|, |b_2^{N,m} - b_2| \rightarrow 0 \text{ in probability as } N, m \rightarrow \infty.$$

Therefore, the proof is complete.

## D Proof of Theorem 3.2

By (3.6) and (3.10), we get

$$\sqrt{N} \left( \begin{pmatrix} \lambda^{N,m} \\ \nu^{N,m} \end{pmatrix} - \begin{pmatrix} \lambda^* \\ \nu^* \end{pmatrix} \right) = \sqrt{N} \begin{pmatrix} \frac{\frac{a_{22}^{N,m} b_1^{N,m} - a_{12}^{N,m} b_2^{N,m}}{a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2} - \frac{a_{22} b_1 - a_{12} b_2}{a_{11} a_{22} - (a_{12})^2}}{\frac{a_{22}^{N,m} b_1^{N,m} - a_{12}^{N,m} b_2^{N,m}}{a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2} - \frac{a_{22} b_1 - a_{12} b_2}{a_{11} a_{22} - (a_{12})^2}} \\ \frac{-a_{21}^{N,m} b_1^{N,m} + a_{11}^{N,m} b_2^{N,m}}{a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2} - \frac{-a_{21} b_1 + a_{11} b_2}{a_{11} a_{22} - (a_{12})^2} \end{pmatrix}.$$

Moreover, simple calculations lead to

$$\begin{aligned} \frac{\frac{a_{22}^{N,m} b_1^{N,m} - a_{12}^{N,m} b_2^{N,m}}{a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2} - \frac{a_{22} b_1 - a_{12} b_2}{a_{11} a_{22} - (a_{12})^2}}{\frac{a_{22}^{N,m} b_1^{N,m} - a_{12}^{N,m} b_2^{N,m}}{a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2} - \frac{a_{22} b_1 - a_{12} b_2}{a_{11} a_{22} - (a_{12})^2}} &= \frac{(a_{22}^{N,m} b_1^{N,m} - a_{12}^{N,m} b_2^{N,m}) - (a_{22} b_1 - a_{12} b_2)}{a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2} \\ &+ \frac{(a_{22} b_1 - a_{12} b_2)[(a_{11} a_{22} - (a_{12})^2) - (a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2)]}{(a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2)(a_{11} a_{22} - (a_{12})^2)}, \end{aligned}$$

and

$$\begin{aligned} \frac{\frac{-a_{21}^{N,m} b_1^{N,m} + a_{11}^{N,m} b_2^{N,m}}{a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2} - \frac{-a_{21} b_1 + a_{11} b_2}{a_{11} a_{22} - (a_{12})^2}}{\frac{-a_{21}^{N,m} b_1^{N,m} + a_{11}^{N,m} b_2^{N,m}}{a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2} - \frac{-a_{21} b_1 + a_{11} b_2}{a_{11} a_{22} - (a_{12})^2}} &= \frac{(-a_{21}^{N,m} b_1^{N,m} + a_{11}^{N,m} b_2^{N,m}) - (-a_{21} b_1 + a_{11} b_2)}{a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2} \\ &+ \frac{(-a_{21} b_1 + a_{11} b_2)[(a_{11} a_{22} - (a_{12})^2) - (a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2)]}{(a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2)(a_{11} a_{22} - (a_{12})^2)}. \end{aligned}$$

To simplify notation, let us define

$$\mathcal{I}^{N,m} := \sqrt{N} \left[ (a_{22}^{N,m} b_1^{N,m} - a_{12}^{N,m} b_2^{N,m}) - (a_{22} b_1 - a_{12} b_2) \right],$$

$$\mathcal{J}^{N,m} := \sqrt{N} \left[ (-a_{21}^{N,m} b_1^{N,m} + a_{11}^{N,m} b_2^{N,m}) - (-a_{21} b_1 + a_{11} b_2) \right],$$

$$\mathcal{H}^{N,m} := a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2,$$

and

$$\mathcal{K}^{N,m} := \sqrt{N} \left[ (a_{11} a_{22} - (a_{12})^2) - (a_{11}^{N,m} a_{22}^{N,m} - (a_{12}^{N,m})^2) \right].$$



We will analyze the convergence of  $\mathcal{I}^{N,m}$ ,  $\mathcal{J}^{N,m}$  and  $\mathcal{K}^{N,m}$  as  $N, m, \frac{m}{\sqrt{N}} \rightarrow \infty$ . By the Skorohod representation theorem (cf. [15, Page 102]), we can and do assume without loss of generality that  $\mathcal{Z}^N$  converges to  $\mathcal{Z}$  in probability in  $\mathcal{D}([0, T], \ell_2)$ . To save space, we will prove the convergence of  $\mathcal{I}^{N,m}$ , the convergence of the other terms follows by similar arguments.

First, we have

$$\mathcal{I}^{N,m} = \mathcal{I}_1^{N,m} + \mathcal{I}_2^{N,m}$$

with

$$\mathcal{I}_1^{N,m} = \sqrt{N}(a_{22}^{N,m}b_1^{N,m} - a_{22}b_1) \quad \text{and} \quad \mathcal{I}_2^{N,m} = \sqrt{N}(a_{12}b_2 - a_{12}^{N,m}b_2^{N,m}).$$

Moreover,

$$\mathcal{I}_1^{N,m} = \sqrt{N}a_{22}(b_1^{N,m} - b_1) + \sqrt{N}b_1^{N,m}(a_{22}^{N,m} - a_{22}),$$

and

$$\mathcal{I}_2^{N,m} = -\sqrt{N}(a_{12}(b_2^{N,m} - b_2) + b_2^{N,m}(a_{12}^{N,m} - a_{12})).$$

Then, by adding and subtracting terms, we get

$$\begin{aligned} & b_1^{N,m} - b_1 \\ &= \sum_{j=0}^{\infty} [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \left\{ \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) \right] - \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) \right] \right\} \\ &+ \sum_{j=0}^{\infty} \left\{ [\varrho_j^N(T, \theta^*) - \varrho_j^N(0, \theta^*)] - [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \right\} \left\{ \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) \right] - \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) \right] \right\} \\ &+ \sum_{j=0}^{\infty} [\varrho_j^N(T, \theta^*) - \varrho_j^N(0, \theta^*)] \left\{ \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) \right] - \int_0^T U_j(\varrho(s)) ds \right\} \\ &+ \sum_{j=0}^{\infty} \left\{ [\varrho_j^N(T, \theta^*) - \varrho_j^N(0, \theta^*)] - [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \right\} \int_0^T U_j(\varrho(s)) ds. \end{aligned} \tag{D.1}$$

Below we consider the convergence of each summation term in (D.1).

(a) By Theorem 2.2, as  $N \rightarrow \infty$ , the term

$$\sqrt{N} \sum_{j=0}^{\infty} \left\{ [\varrho_j^N(T, \theta^*) - \varrho_j^N(0, \theta^*)] - [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \right\} \int_0^T U_j(\varrho(s)) ds$$

converges in probability to

$$\sum_{j=0}^{\infty} [\mathcal{Z}_j(T, \theta^*) - \mathcal{Z}_j(0, \theta^*)] \int_0^T U_j(\varrho(s)) ds.$$

(b) By (2.1) and (3.1)–(3.3), we get

$$\begin{aligned}
|U_j(\varrho(t_k, \theta^*)) - U_j(\varrho(s))| &\leq 2\varrho_{j-1}(t_k, \theta^*) \sum_{i=0}^{\infty} |\varrho_i(t_k, \theta^*) - \varrho_i(s, \theta^*)| + 2|\varrho_{j-1}(t_k, \theta^*) - \varrho_{j-1}(s, \theta^*)| \\
&\quad + 2\varrho_j(t_k, \theta^*) \sum_{i=0}^{\infty} |\varrho_i(t_k, \theta^*) - \varrho_i(s, \theta^*)| + 2|\varrho_j(t_k, \theta^*) - \varrho_j(s, \theta^*)| \\
&\quad + 2|\varrho_{j-1}(t_k, \theta^*) - \varrho_{j-1}(s, \theta^*)| + 2|\varrho_j(t_k, \theta^*) - \varrho_j(s, \theta^*)| \\
&\leq 6 \sum_{i=0}^{\infty} |\varrho_i(t_k, \theta^*) - \varrho_i(s, \theta^*)| \\
&\leq 6 \sum_{i=0}^{\infty} \int_s^{t_k} |F_i(\varrho(u, \theta^*))| du \\
&\leq \frac{6(6\lambda + 2\nu)T}{m}.
\end{aligned}$$

Then,

$$\begin{aligned}
&\left| \sqrt{N} \sum_{j=0}^{\infty} [\varrho_j^N(T, \theta^*) - \varrho_j^N(0, \theta^*)] \left\{ \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) \right] - \int_0^T U_j(\varrho(s)) ds \right\} \right| \\
&= \left| \sqrt{N} \sum_{j=0}^{\infty} [\varrho_j^N(T, \theta^*) - \varrho_j^N(0, \theta^*)] \left\{ \sum_{k=1}^m \int_{t_{k-1}}^{t_k} [U_j(\varrho(t_k, \theta^*)) - U_j(\varrho(s))] ds \right\} \right| \\
&\leq \sqrt{N} \sum_{j=0}^{\infty} [\varrho_j^N(T, \theta^*) + \varrho_j^N(0, \theta^*)] \left\{ \sum_{k=1}^m \int_{t_{k-1}}^{t_k} \frac{6(6\lambda + 2\nu)T}{m} ds \right\} \\
&= \frac{24(3\lambda + \nu)T^2 \sqrt{N}}{m} \\
&\rightarrow 0 \quad \text{as } \frac{m}{\sqrt{N}} \rightarrow \infty.
\end{aligned}$$

(c) By the estimate [6, Page 70, line 4], we have that

$$\sup_{N \in \mathbb{N}} E \left[ \sup_{t \in [0, T]} \sum_{j=0}^{\infty} (j+1)^2 (\mathcal{Z}_j^N(t, \theta^*))^2 \right] < \infty. \tag{D.2}$$

Then, by Fatou's lemma, we get

$$E \left[ \sup_{t \in [0, T]} \sum_{j=0}^{\infty} (j+1)^2 (\mathcal{Z}_j(t, \theta^*))^2 \right] < \infty. \tag{D.3}$$

By (2.1) and (3.1)–(3.3), we get

$$|U_j(\varrho^N(t_k, \theta^*)) - U_j(\varrho(t_k, \theta^*))| \leq 6 \sum_{i=0}^{\infty} |\varrho_i^N(t_k, \theta^*) - \varrho_i(t_k, \theta^*)|,$$

which together with (C.1) and (D.2) implies that

$$\begin{aligned}
& \left| \sqrt{N} \sum_{j=0}^{\infty} \left\{ [\varrho_j^N(T, \theta^*) - \varrho_j^N(0, \theta^*)] - [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \right\} \right. \\
& \quad \cdot \left. \left\{ \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) \right] - \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) \right] \right\} \right| \\
&= \left| \sum_{j=0}^{\infty} [\mathcal{Z}_j^N(T, \theta^*) - \mathcal{Z}_j^N(0, \theta^*)] \left\{ \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) \right] - \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) \right] \right\} \right| \\
&\leq 6 \sum_{j=0}^{\infty} [|\mathcal{Z}_j^N(T, \theta^*)| + |\mathcal{Z}_j^N(0, \theta^*)|] \left\{ \frac{T}{m} \sum_{k=1}^m \sum_{i=0}^{\infty} |\varrho_i^N(t_k, \theta^*) - \varrho_i(t_k, \theta^*)| \right\} \\
&\leq 6T \sum_{j=0}^{\infty} [|\mathcal{Z}_j^N(T, \theta^*)| + |\mathcal{Z}_j^N(0, \theta^*)|] \cdot \sup_{t \in [0, T]} \sum_{i=0}^{\infty} |\varrho_i^N(t, \theta^*) - \varrho_i(t, \theta^*)| \\
&\rightarrow 0 \text{ in probability as } N \rightarrow \infty.
\end{aligned}$$

(d) For  $i \in \mathbb{Z}_+$  and  $k \in \{1, 2, \dots, m\}$ , define a non-negative measure  $\tau^{i,k}$  on  $\mathbb{Z}_+$  by

$$\tau_l^{i,k} := \begin{cases} \varrho_l(t_k, \theta^*), & l < i, \\ \varrho_l^N(t_k, \theta^*), & l \geq i. \end{cases}$$

Note that

$$\partial_l U_j(x) = \begin{cases} 0, & l < j-1, \\ 2 \sum_{p=j-1}^{\infty} x_p, & l = j-1, \\ 2x_{j-1} - 2 \sum_{p=j}^{\infty} x_p, & l = j, \\ 2x_{j-1} - 2x_j, & l \geq j+1. \end{cases}$$

Then, for  $s \in [t_{k-1}, t_k]$ , we have

$$\begin{aligned}
& \left| \sqrt{N} [U_j(\varrho^N(t_k, \theta^*)) - U_j(\varrho(t_k, \theta^*))] - \sum_{i=0}^{\infty} \partial_i U_j(\varrho(s)) \mathcal{Z}_i^N(t_k, \theta^*) \right| \\
&= \left| \sum_{i=0}^{\infty} \left\{ \sqrt{N} [U_j(\tau^{i,k}) - U_j(\tau^{i+1,k})] - \partial_i U_j(\varrho(s)) \mathcal{Z}_i^N(t_k, \theta^*) \right\} \right| \\
&\leq \sum_{i=0}^{\infty} \left| \sqrt{N} [U_j(\tau^{i,k}) - U_j(\tau^{i+1,k})] - \partial_i U_j(\varrho(s)) \mathcal{Z}_i^N(t_k, \theta^*) \right| \\
&= \sum_{i=0}^{\infty} \left| \sqrt{N} \int_{\varrho_i(t_k, \theta^*)}^{\varrho_i^N(t_k, \theta^*)} \partial_i U_j(\varrho_0(t_k, \theta^*), \dots, \varrho_{i-1}(t_k, \theta^*), u, \varrho_{i+1}^N(t_k, \theta^*), \varrho_{i+2}^N(t_k, \theta^*), \dots) du \right|
\end{aligned}$$

$$\begin{aligned}
& \left| -\sqrt{N} \int_{\varrho_i(t_k, \theta^*)}^{\varrho_i^N(t_k, \theta^*)} \partial_i U_j(\varrho(s)) du \right| \\
\leq & 2 \left\{ \sup_{t \in [0, T]} \sum_{i=0}^{\infty} |\mathcal{Z}_i^N(t, \theta^*)| \right\} \left\{ \sup_{1 \leq k \leq m} \sum_{i=0}^{\infty} \sup_{s, t \in [t_{k-1}, t_k]} |\varrho_i(s, \theta^*) - \varrho_i(t, \theta^*)| + \sup_{t \in [0, T]} \sum_{i=0}^{\infty} |\varrho_i^N(t, \theta^*) - \varrho_i(t, \theta^*)| \right\}.
\end{aligned} \tag{D.4}$$

By [15, (5.7), page 117 and Proposition 5.3, page 119], we have that

$$\sup_{t \in [0, T]} \|\mathcal{Z}^N(t, \theta^*) - \mathcal{Z}(t, \theta^*)\|_2 \rightarrow 0 \tag{D.5}$$

in probability as  $N \rightarrow \infty$ . Then, by (C.1) and (D.2)–(D.5), we get

$$\begin{aligned}
& \left| \sqrt{N} \sum_{j=0}^{\infty} [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \left\{ \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) \right] - \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) \right] \right\} \right. \\
& \left. - \sum_{j=0}^{\infty} [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \right| \\
\leq & \sum_{j=0}^{\infty} [\varrho_j(T, \theta^*) + \varrho_j(0, \theta^*)] \left\{ \sum_{k=1}^m \int_{t_{k-1}}^{t_k} \left| \sqrt{N} [U_j(\varrho^N(t_k, \theta^*)) - U_j(\varrho(t_k, \theta^*))] - \sum_{i=0}^{\infty} \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) \right| ds \right\} \\
\leq & 2 \sup_{j \in \mathbb{Z}_+} \left\{ \sum_{k=1}^m \int_{t_{k-1}}^{t_k} \left| \sqrt{N} [U_j(\varrho^N(t_k, \theta^*)) - U_j(\varrho(t_k, \theta^*))] - \sum_{i=0}^{\infty} \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) \right| ds \right\} \\
\leq & 2 \sup_{j \in \mathbb{Z}_+} \left\{ \sum_{k=1}^m \int_{t_{k-1}}^{t_k} \left| \sqrt{N} [U_j(\varrho^N(t_k, \theta^*)) - U_j(\varrho(t_k, \theta^*))] - \sum_{i=0}^{\infty} \partial_i U_j(\varrho(s)) \mathcal{Z}_i^N(t_k, \theta^*) \right| ds \right\} \\
& + 2 \sup_{j \in \mathbb{Z}_+} \left\{ \sum_{k=1}^m \int_{t_{k-1}}^{t_k} \left| \sum_{i=0}^{\infty} \partial_i U_j(\varrho(s)) \mathcal{Z}_i^N(t_k, \theta^*) - \sum_{i=0}^{\infty} \partial_i U_j(\varrho(s)) \mathcal{Z}_i(t_k, \theta^*) \right| ds \right\} \\
& + 2 \sup_{j \in \mathbb{Z}_+} \left\{ \sum_{k=1}^m \int_{t_{k-1}}^{t_k} \left| \sum_{i=0}^{\infty} \partial_i U_j(\varrho(s)) \mathcal{Z}_i(t_k, \theta^*) - \sum_{i=0}^{\infty} \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) \right| ds \right\} \\
\leq & 4T \left\{ \sup_{t \in [0, T]} \sum_{i=0}^{\infty} |\mathcal{Z}_i^N(t, \theta^*)| \right\} \left\{ \sup_{1 \leq k \leq m} \sum_{i=0}^{\infty} \sup_{s, t \in [t_{k-1}, t_k]} |\varrho_i(s, \theta^*) - \varrho_i(t, \theta^*)| + \sup_{t \in [0, T]} \sum_{i=0}^{\infty} |\varrho_i^N(t, \theta^*) - \varrho_i(t, \theta^*)| \right\} \\
& + 12T \sup_{t \in [0, T]} \sum_{i=0}^{\infty} |\mathcal{Z}_i^N(t, \theta^*) - \mathcal{Z}_i(t, \theta^*)| + 12 \sum_{k=1}^m \int_{t_{k-1}}^{t_k} \sum_{i=0}^{\infty} |\mathcal{Z}_i(t_k, \theta^*) - \mathcal{Z}_i(s, \theta^*)| ds \\
\rightarrow & 0 \text{ in probability as } N, m \rightarrow \infty.
\end{aligned}$$

Thus, by (a)–(d), we deduce that as  $N, m, \frac{m}{\sqrt{N}} \rightarrow \infty$ ,  $\sqrt{N}(b_1^{N, m} - b_1)$  converges in probability to

$$\sum_{j=0}^{\infty} [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right]$$

$$+ \sum_{j=0}^{\infty} [\mathcal{Z}_j(T, \theta^*) - \mathcal{Z}_j(0, \theta^*)] \int_0^T U_j(\varrho(s)) ds. \quad (D.6)$$

By adding and subtracting terms, we get

$$\begin{aligned} a_{12}^{N,m} - a_{12} &= \sum_{j=0}^{\infty} \left\{ \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho^N(t_k, \theta^*)) \right] - \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) \right] \right\} \left[ \frac{T}{m} \sum_{k=1}^m V_j(\varrho^N(t_k, \theta^*)) \right] \\ &+ \sum_{j=0}^{\infty} \left\{ \left[ \frac{T}{m} \sum_{k=1}^m U_j(\varrho(t_k, \theta^*)) \right] - \int_0^T U_j(\varrho(s)) ds \right\} \left[ \frac{T}{m} \sum_{k=1}^m V_j(\varrho^N(t_k, \theta^*)) \right] \\ &+ \sum_{j=0}^{\infty} \int_0^T U_j(\varrho(s)) ds \left\{ \left[ \frac{T}{m} \sum_{k=1}^m V_j(\varrho^N(t_k, \theta^*)) \right] - \left[ \frac{T}{m} \sum_{k=1}^m V_j(\varrho(t_k, \theta^*)) \right] \right\} \\ &+ \sum_{j=0}^{\infty} \int_0^T U_j(\varrho(s)) ds \left\{ \left[ \frac{T}{m} \sum_{k=1}^m V_j(\varrho(t_k, \theta^*)) \right] - \int_0^T V_j(\varrho(s)) ds \right\}. \end{aligned}$$

Similar to the above argument, we can show that as  $N, m, \frac{m}{\sqrt{N}} \rightarrow \infty$ ,  $\sqrt{N}(a_{12}^{N,m} - a_{12})$  converges in probability to

$$\begin{aligned} &\sum_{j=0}^{\infty} \int_0^T V_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i U_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \\ &+ \sum_{j=0}^{\infty} \int_0^T U_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i V_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right]. \end{aligned} \quad (D.7)$$

Similarly, as  $N, m, \frac{m}{\sqrt{N}} \rightarrow \infty$ ,  $\sqrt{N}(b_2^{N,m} - b_2)$  converges in probability to

$$\begin{aligned} &\sum_{j=0}^{\infty} [\varrho_j(T, \theta^*) - \varrho_j(0, \theta^*)] \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i V_j(\varrho(s)) \mathcal{Z}_i(s, \theta^*) ds \right] \\ &+ \sum_{j=0}^{\infty} [\mathcal{Z}_j(T, \theta^*) - \mathcal{Z}_j(0, \theta^*)] \int_0^T V_j(\varrho(s)) ds, \end{aligned} \quad (D.8)$$

and  $\sqrt{N}(a_{22}^{N,m} - a_{22})$  converges in probability to

$$2 \int_0^T V_j(\varrho(s)) ds \left[ \sum_{i=0}^{\infty} \int_0^T \partial_i V_j(s, \theta^*) \mathcal{Z}_i(s, \theta^*) ds \right]. \quad (D.9)$$

Therefore, using (D.6)–(D.9), we deduce that

$$\mathcal{I}^{N,m} \xrightarrow{P} \mathcal{I},$$

Similarly, we obtain

$$\mathcal{J}^{N,m} \xrightarrow{P} \mathcal{J}, \quad \mathcal{K}^{N,m} \xrightarrow{P} \mathcal{K}.$$

Finally, by Theorem 2.1, we deduce that  $\mathcal{H}^{N,m} \rightarrow a_{11}a_{22} - (a_{12})^2$  in probability as  $N, m \rightarrow \infty$ . Therefore, the proof is complete by the continuous mapping theorem (cf. [34, Theorem 2.3]).

## References

- [1] C. Amorino, A. Heidari, V. Pilipauskaite, and M. Podolskij. Parameter estimation of discretely observed interacting particle systems. *Stochastic Processes and their Applications*, 163:350–386, 2023.
- [2] A. Asanjarani, Y. Nazarathy, and P. Taylor. A survey of parameter and state estimation in queues. *Queueing Syst*, 97:39–80, 2021.
- [3] D. Belomestny, V. Pilipauskaite, and M. Podolskij. Semiparametric estimation of mckean-vlasov sdes. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 59(1):79–96, 2023.
- [4] M. Benaïm and J.Y. Le Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65(11–12):823–838, 2008.
- [5] P.N. Bishwal. Estimation in interacting diffusions: Continuous and discrete sampling. *Applied Mathematics-a Journal of Chinese Universities Series B*, 02:1154–1158, 2011.
- [6] A. Budhiraja and E. Friedlander. Diffusion approximations for load balancing mechanisms in cloud storage systems. *Advances in Applied Probability*, 51(1):41–86, 2019.
- [7] A. Budhiraja, D. Mukherjee, and R. Wu. Supermarket model on graphs. *The Annals of Applied Probability*, 29(3):1740–1777, 2019.
- [8] P. Curie. Magnetic properties of materials at various temperatures. *Ann. Chem. Phys*, 5(289), 1895.
- [9] D.A. Dawson. Critical dynamics and fluctuations for a mean field model of cooperative behaviour. *J. Statist. Phys.*, 41:29–85, 1983.
- [10] D.A. Dawson. Introductory lectures on stochastic population systems. *arXiv:1705.03781 [math.PR]*, 2017.
- [11] D.A. Dawson, J. Tang, and Y.Q. Zhao. Balancing queues by mean field interactions. *Queueing Syst*, 49:335–361, 2005.
- [12] D.A. Dawson, J. Tang, and Y.Q. Zhao. Performance analysis of joining the shortest queue model among a large number of queues. *Asia-Pacific Journal of Operational Research*, 36(4), 2019.
- [13] S. Delattre and N. Fournier. Statistical inference versus mean field limit for hawkes processes. *Electronic Journal of Statistics*, 10(1):1223–1295, 2016.
- [14] R.L. Dobrushin and Y.M. Sukhov. Asymptotic investigation of star-shaped message switching networks with a large number of radial rays. *Probl. Inf. Trans.*, 12(1):49–66, 1976.
- [15] S.N. Ethier and T.G. Kurtz. *The infinitely-many-alleles model with selection as a measure-valued diffusion. Stochastic Methods in Biology*, volume 70. Springer-Verlag, Berlin-Heidelberg-N.Y., 1987.
- [16] J. Gärtner. On the mckean-vlasov limit for interacting diffusions. *Math. Nachr.*, 137:197–248, 1988.

- [17] V. Genon-Catalot and C. Larédo. Parametric inference for small variance and long time horizon mckean-vlasov diffusion models. *Electronic Journal of Statistics*, 15(2):5811–5854.
- [18] K. Giesecke, G. Schwenkler, and J.A. Sirignano. Inference for large financial systems. *Math. Fin.*, 30:823–838, 2019.
- [19] K. Giesecke, K. Spiliopoulos, R.B. Sowers, and J.A. Sirignano. Large portfolio asymptotics for loss from default. *Mathematical Finance*, 25(1):77–114, 2015.
- [20] C. Graham. Chaoticity on path space for a queueing network with selection of the shortest queue amongst several. *J. Appl. Prob.*, 37(1):198–211, 2000.
- [21] C. Graham and S. Méléard. Propagation of chaos for a fully connected loss network with alternate routing. *Stoch. Processes Appl.*, 44:159–180, 1993.
- [22] C. Graham and S. Méléard. Dynamic asymptotic results for a generalized star-shaped loss network. *Ann. Appl. Prob.*, 5, 1995.
- [23] H.P. McKean Jr. A class of markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA*, 56(6):1907–1911, 1966.
- [24] H.P. McKean Jr. Speed of approach to equilibrium for kac’s caricature of a maxwellian gas. *Arch. Ration. Mech. Anal.*, 21(5):343–367, 1966.
- [25] M. Kac. Foundations of kinetic theory. In Calif University of California Press, Berkeley, editor, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Contributions to Astronomy and Physics*, pages 171–197, 1956.
- [26] R.A. Kasonga. Maximum likelihood theory for large interacting systems. *SIAM Journal on Applied Mathematics*, 50(3):865–875, 1990.
- [27] O. Kley, C. Klüppelberg, and L. Reichel. Systemic risk through contagion in a core-periphery structured banking network. *Banach Center publications*, 104:133–149, 2015.
- [28] C. Liu. Statistical inference for a partially observed interacting system of hawkes processes. *Stochastic Processes and their Applications*, 130(9):5636–5694, 2020.
- [29] L. Della Maestra and M. Hoffmann. Nonparametric estimation for interacting particle systems: Mckean-vlasov models. *Probab. Theory Relat. Fields*, 182:551–613, 2022.
- [30] S. Méléard and V. Bansaye. *Some Stochastic Models for Structured Populations: Scaling Limits and Long Time Behavior*. Springer, 2015.
- [31] M. Mitzenmacher. The power of two choices in randomized load balancing. *PhD thesis, University of California at Berkeley*, 1996.
- [32] L. Sharrock, N. Kantas, P. Parpas, and G.A. Pavliotis. Online parameter estimation for the mckean-vlasov stochastic differential equation. *Stochastic Processes and their Applications*, 162:481–546, 2023.

- [33] J.M. Steele. *The Cauchy-Schwarz Master Class*. Cambridge University Press, 2004.
- [34] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.
- [35] N.D. Vvedenskaya, R.L. Dobrushin, and F.I. Karpelevich. Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problems Inform. Transmission*, 32(1):15–27, 1996.
- [36] P. Weiss. L’hypothèse du champ moléculaire et la propriété ferromagnétique. *J. Phys. Theor. Appl.*, 6(1):661–690, 1907.
- [37] Y.Q. Zhao. Statistical inference for mean-field queueing models. *Queueing Syst*, 100:569–571, 2022.