Nonlinear Dyn (2018) 91:2411–2427
https://doi.org/10.1007/s11071-017-4021-0

ORIGINAL PAPER

An analysis of queues with delayed information and time-varying arrival rates

Jamol Pender · Richard H. Rand · Elizabeth Wesson

Received: 14 January 2017 / Accepted: 17 December 2017 / Published online: 3 February 2018

Abstract Understanding how delayed information impacts queueing systems is an important area of research. However, much of the current literature neglects one important feature of many queueing systems, namely non-stationary arrivals. Non-stationary arrivals model the fact that customers tend to access services during certain times of the day and not at a constant rate. In this paper, we analyze two two-dimensional deterministic fluid models that incorporate customer choice behavior based on delayed queue length information with time-varying arrivals. In the first model, customers receive queue length information that is delayed by a constant Δ. In the second model, customers receive information about the queue length through a moving average of the queue length where the moving average window is Δ. We analyze the impact of a time-varying arrival rate and show using asymptotic analysis that the time-varying arrival rate does not impact the critical delay unless the frequency of the time-varying arrival rate is twice that of the critical delay. When the frequency of the arrival rate is twice that of the critical delay, then the stability is enlarged by a wedge that is determined by the model parameters. As a result, this problem allows us to combine the theory of nonlinear dynamics, parametric excitation, delays, and time-varying queues together to provide insight into the impact of information on queueing systems.

Keywords Queues · Delay differential equation · Time-varying rates · Asymptotics · Two-variable expansions · Slow flows · Fluid limits

J. Pender (✉)
School of Operations Research and Information Engineering, Cornell University, 228 Rhodes Hall, Ithaca, NY 14853, USA
e-mail: jjp274@cornell.edu

R. H. Rand
Department of Mathematics, Sibley School of Mechanical and Aerospace Engineering, Cornell University, 535 Malott Hall, Ithaca, NY 14853, USA
e-mail: rand@math.cornell.edu

E. Wesson
Department of Mathematics, Cornell University, 582 Malott Hall, Ithaca, NY 14853, USA
e-mail: enw27@cornell.edu

1 Introduction

Understanding the impact of providing delayed information to customers in queueing systems is a very important problem in the queueing and engineering literature. Many companies where customers are forced to wait in line often choose to provide their customers with waiting time or queue length information. Consequently, the information that is provided can affect a customer's choice of using the service and joining the queue. One common example of this communication between the service and customer is delay announcements. Delay announcements commonly inform cus-

tomers about the average waiting time to start service. These announcements are not only important because they give the customer information about their wait, but also the announcements have the possibility of influencing the possibility that a customer will return to use the service again or remain in. As a consequence, understanding the impact of providing queue length information to customers on customer choices and system operations, as well as the development of methods to support such announcements, has attracted the attention of the queueing systems community recently.

Much of the research on providing queue length or waiting time information to customers focuses on the impact of delay announcements in call centers. There is a vast literature on this subject, but our focus is quite different. Work by Ibrahim and Whitt [15–18] develops new real-time estimators for estimating delays in various queueing systems. The work of Armony and Maglaras [3], Guo and Zipkin [10], Hassin [12], Armony et al. [4], Guo and Zipkin [11], Jouini et al. [21,22], Allon and Bassamboo [1], Allon et al. [2], Ibrahim et al. [19], Whitt [37] and references therein analyzes the impact of delay announcements on the queueing process and the abandonment process of the system. Lastly, the work of Hui and Tse [13], Hul et al. [14], Pruyn and Smidts [32], Munichor and Rafaeli [26], Sarel and Marmorstein [34], Taylor [35] explores the behavioral aspect of customer waiting and how delays affect customer decisions. This paper is concerned about the impact of time-varying arrival rates and delayed information on the queue length process. Thus, it is mostly related to the work by Armony and Maglaras [3], Guo and Zipkin [10], Hassin [12], Armony et al. [4], Guo and Zipkin [11], Jouini et al. [21,22], Allon and Bassamboo [1], Allon et al. [2], Ibrahim et al. [19], Whitt [37], Armony et al. [5], Dong et al. [7].

More recently, there is also research that considers how information can impact the dynamics of queueing systems. Work by Jennings and Pender [20] compares ticket queues with standard queues. In a ticket queue, the manager of the queue is unaware of when a customer abandons and is only notified of the abandonment when the customer would have entered service. This artificially inflates the queue length process, and Jennings and Pender [20] determines how much the queue length is inflated because of this loss of information. However, this work does not consider the aspect of choice and delays in publishing the information to customers, which is the case in many healthcare and transportation settings.

This paper analyzes two deterministic queueing models, which describe the dynamics of customer choice and delayed queue length information. In the first model, the customer receives information about the queue length which is delayed by a parameter Δ. In the second model, we use a moving average of the queue length over the time interval Δ to represent the queue length information given to the customer. The models that we analyze are identical to the models that were analyzed in Pender et al. [31]; however, in this paper we add a time-varying arrival rate, which is a significant generalization. This is because queues with time varying arrival rates are notoriously difficult to analyze since many of the standard techniques do not apply.

However, in this paper, we apply asymptotic analysis techniques like matched asymptotic expansions and the two-variable expansion method to analyze our new time-varying queueing systems with delayed information. We show in both models that when the time-varying arrival is sinusoidal and the sinusoidal part is small, then the time-varying part of the arrival rate does not affect the stability of the queueing dynamics unless the frequency of the arrival rate is twice that of the oscillation frequency.

The main results in this work represent a novel contribution to the literature in queueing theory and dynamical systems because many real-world queueing systems have time-varying rates. Moreover, it is important to understand when the time-varying arrival rate will have an affect on the stability dynamics of the system. We will show that when the time-varying amplitude is small relative to the base arrival rate and the frequency of the arrival rate is not twice that of the critical delay, then the stability dynamics are roughly identical to the non-time-varying case. This is a significant result since the amplitude of the time-varying arrival rate in many queueing systems are on the order of 20% of the average arrival rate and are not large relative to the average arrival rate. Furthermore, we precisely show in the small amplitude setting when the time variation has an impact on the critical delay and when it does not. This will give queueing managers more insight into how to operate their queues when in

the presence of delayed information and time-varying arrival rates.

1.1 Main contributions of paper

The contributions of this work can be summarized as follows.

- We analyze two two-dimensional fluid models with time-varying arrival rates that incorporate customer choice based on delayed queue length information. In the first model, the information provided to the customer is the queue length delayed by a constant Δ, and in the second model, the information provided to the customer is a moving average of the queue length over a time window of size Δ. We show that the impact of the time-varying arrival does not shift the value of the critical delay unless the frequency of the arrival rate is twice that of the critical frequency.
- We show in both models using the method of multiple time scales that the critical delay, which determines the stability of the delay differential equations, can be shifted by the incorporation of time-varying arrival rates. We also determine the size and impact of this shift.

1.2 Organization of paper

The remainder of this paper is organized as follows. Section 2 gives a brief overview of the infinite server queue with time-varying arrival rates and describes the constant delay fluid model. Using asymptotic expansions, we derive the critical delay threshold under which the queues are balanced if the delay is below the threshold and the queues are asynchronized if the delay is above the threshold. We also show that the increased or decreased stability because of the time-varying arrival rates depends on the sign of the amplitude. Section 3 describes a constant moving average delay fluid model. Using similar asymptotic expansions, we derive the critical delay threshold under which the queues are balanced if the delay is below the threshold and the queues are asynchronized if the delay is above the threshold in the case of time-varying arrival rates. Finally, in Sect. 4, we conclude with directions for future research related to this work.

2 Constant delay fluid model

In this section, we present a fluid model with customer choice based on the queue length with a constant delay. Thus, we begin with two infinite server queues operating in parallel, where customers choose which queue to join by taking the size of the queue length into account. However, we add the twist that the queue length information that is reported to the customer is delayed by a constant Δ. Therefore, the queue length that the customer receives is actually the queue length Δ time units in the past. Our choice model is identical to that of a Multinomial Logit Model (MNL) Ben-Akiva and Bierlaire [6], Train [36] where the utility for being served in the ith queue with delayed queue length $Q_i(t-\Delta)$ is $u_i(Q_i(t-\Delta)) = Q_i(t-\Delta)$. Thus, our deterministic queueing model with customer choice, delayed information, and time-varying arrival rates can be represented by the two-dimensional system of delay differential equations

$$\dot{q}_1(t) = \lambda(t) \cdot \frac{\exp(-q_1(t-\Delta))}{\exp(-q_1(t-\Delta)) + \exp(-q_2(t-\Delta))} - \mu q_1(t) \tag{2.1}$$

$$\dot{q}_2(t) = \lambda(t) \cdot \frac{\exp(-q_2(t-\Delta))}{\exp(-q_1(t-\Delta)) + \exp(-q_2(t-\Delta))} - \mu q_2(t) \tag{2.2}$$

where we assume that $q_1(t)$ and $q_2(t)$ start with different initial functions $\varphi_1(t)$ and $\varphi_2(t)$ on the interval $[-\Delta, 0]$, $\lambda(t)$ is the total arrival rate to both queues, and μ is the service rate of each queue.

With respect to Eqs. (2.1)–(2.2), the arrival rate is given by

$$\text{Arrival Rate to } i\text{th Station} = \lambda(t) \cdot \frac{\exp(-q_i(t-\Delta))}{\exp(-q_1(t-\Delta)) + \exp(-q_2(t-\Delta))} \tag{2.3}$$

is state dependent and delayed. One reason that it has this exponential form given by a MNL model is that as more customers that are present in a queue, the less likely a customer will be willing to join that line. If the two queues are equal, then a customer is ambivalent toward joining a line. On the other hand, the departure rate is given by

$$\text{Departure Rate from } i\text{th Station} = \mu q_i(t) \tag{2.4}$$

and describes the rate at which customers are leaving the queue. Since we are modeling an infinite server queue, it follows that the more customers that are present in that queue, the larger the rate of departure.

Remark When the two delay differential equations are started with the same initial functions, they are identical for all time because of the symmetry of the problem. Therefore, we will start the system with non-identical initial conditions, so the problem is no longer trivial and the dynamics are not identical.

In the constant delay model, it is critical to understand the case when the arrival rate is constant and does not depend on time. In Pender et al. [31], the authors show that the critical delay can be determined from the model parameters and the following theorem is from Pender et al. [31].

Proposition 2.1 *For the constant delay choice model, the critical delay parameter is given by the following expression*

$$\Delta_{\text{cr}}(\lambda, \mu) = \frac{2\arccos(-2\mu/\lambda)}{\sqrt{\lambda^2 - 4\mu^2}}. \tag{2.5}$$

Proof See Pender et al. [31]. □

However, the model of Pender et al. [31] neglects to consider the impact of time-varying arrival rates. Time-varying arrival rates are important to incorporate into one's model of queues since real customer behavior is dynamic and is not constant over time. To this end, we will exploit asymptotic analysis and perturbation methods to obtain some insight into the impact of time-varying arrival rates.

2.1 Understanding the $M_t/M/\infty$ queue

Before we analyze the queueing model with customer choice, it is important to understand the dynamics of the infinite server queue with a time-varying arrival rate since it will be essential to our future analysis. We know from the work of Eick et al. [8,9] that the infinite server queue or the M_t/G/∞ queue has a Poisson distribution when initialized at zero or with a Poisson distribution with mean rate $q_\infty(t)$ where

$$\begin{aligned} q_\infty(t) &= E[Q_\infty(t)] && (2.6)\\ &= \int_{-\infty}^{t} \overline{G}(t-u)\lambda(u)\mathrm{d}u && (2.7)\\ &= E\left[\int_{t-S}^{t} \lambda(u)\mathrm{d}u\right] && (2.8)\\ &= E[\lambda(t-S_e)]\cdot E[S] && (2.9) \end{aligned}$$

where S represents a service time with distribution G, $\overline{G} = 1 - G(t) = \mathbb{P}(S > t)$, and S_e is a random variable with distribution that follows the stationary excess of residual-lifetime cdf G_e, defined by

$$G_e(t) \equiv \mathbb{P}(S_e < t) = \frac{1}{E[S]}\int_0^t \overline{G}(u)\mathrm{d}u, \quad t \geq 0. \tag{2.10}$$

The exact analysis of the infinite server queue is often useful since it represents the dynamics of the queueing process if there were an unlimited amount of resources to satisfy the demand process. Moreover, as observed in Pender [28], when the service time distribution is exponential, the mean of the queue length process $q^\infty(t)$ is the solution to the following ordinary linear differential equation

$$\dot{q}_\infty(t) = \lambda(t) - \mu \cdot q_\infty(t). \tag{2.11}$$

Proposition 2.2 *The solution to the mean of the $M_t/M_t/\infty$ queue with initial value q_0 is given by*

$$\begin{aligned} &E[Q_\infty(t)] = q_\infty(t) && (2.12)\\ &= q_0 \cdot \exp\left\{-\int_0^t \mu(s)\mathrm{d}s\right\} && (2.13)\\ &\quad + \left(\exp\left\{-\int_0^t \mu(s)\mathrm{d}s\right\}\right.\\ &\quad \left.\cdot\left(\int_0^t \lambda(s)\exp\left\{\int_0^s \mu(r)\mathrm{d}r\right\}\mathrm{d}s\right)\right). && (2.14) \end{aligned}$$

Proof We can exploit the fact that the mean of the Markovian time-varying infinite server queue solves a linear ordinary differential equation. Therefore, we can use standard ode theory to find the mean of the infinite server queue. For more details, see for example, Pender [28]. □

Corollary 2.3 *In the special case where $q_0 = 0$, μ is constant, and $\lambda(t) = \lambda + \lambda\cdot\alpha\sin(\gamma t)$, the mean queue has the following representation*

$$E[Q_\infty(t)] = \frac{\lambda}{\mu}\cdot(1-\exp(-\mu t)) + \frac{\lambda\cdot\alpha}{\mu^2+\gamma^2} \cdot\left[(\mu\cdot\sin(\gamma t)-\gamma\cdot\cos(\gamma t)) + \exp(-\mu t)\cdot\gamma\right].$$

Moreover, when t is very large, then we have that

$$E[Q_\infty(t)] \approx \frac{\lambda}{\mu} + \frac{\lambda\cdot\alpha}{\mu^2+\gamma^2} \cdot\left[\mu\cdot\sin(\gamma t)-\gamma\cdot\cos(\gamma t)\right].$$

2.2 Constant delay model with time-varying arrivals

Although the case where the constant delay queueing model has a constant arrival rate λ, the extension to more complicated arrival functions such as $\lambda(t) = \lambda + \lambda\cdot\alpha\sin(\gamma t)$ is quite difficult to analyze. However, we can analyze the system when the time-varying arrival rate is close to the constant rate case using perturbation theory. Thus, we assume that the queue length equations for the constant delay model satisfy the following delay differential equations

$$\dot{q}_1(t) = (\lambda+\lambda\cdot\alpha\cdot\epsilon\sin(\gamma t)) \cdot\frac{\exp(-q_1(t-\Delta))}{\exp(-q_1(t-\Delta))+\exp(-q_2(t-\Delta))} - \mu q_1(t) \tag{2.15}$$

$$\dot{q}_2(t) = (\lambda+\lambda\cdot\alpha\cdot\epsilon\sin(\gamma t)) \cdot\frac{\exp(-q_2(t-\Delta))}{\exp(-q_1(t-\Delta))+\exp(-q_2(t-\Delta))} - \mu q_2(t) \tag{2.16}$$

where we assume that $q_1(t)$ and $q_2(t)$ start with different initial functions $\varphi_1(t)$ and $\varphi_2(t)$ on the interval $[-\Delta, 0]$ and we assume that $0 \le \alpha \le 1$ and $\epsilon \ll 1$.

In order to begin our analysis of the delay differential equations, we need to understand the case where $\epsilon = 0$. Fortunately, this analysis has been carried out in Pender et al. [31] and we give a brief outline of the analysis for the reader's convenience. The first step to understanding the case when $\epsilon = 0$ to compute the equilibrium in this case.

In our case, the delay differential equations given in Eqs. (2.15)–(2.16) are symmetric. Moreover, in the case where the delay $\Delta = 0$, the two equations converge to the same point since in equilibrium each queue will receive exactly one half of the arrivals and the two service rates are identical. This is also true in the case where the arrival process contains delays in the queue length since in equilibrium, the delayed queue length is equal to the non-delayed queue length. Thus, we have in equilibrium that

$$q_1(t) = q_2(t) = \frac{q_\infty(t)}{2} \quad \text{as } t\to\infty. \tag{2.17}$$

and

$$q_1(t-\Delta) = q_2(t-\Delta) = \frac{q_\infty(t-\Delta)}{2} \quad \text{as } t\to\infty. \tag{2.18}$$

Now that we know the equilibrium for Eqs. (2.1)–(2.2), we need to understand the stability of the delay differential equations around the equilibrium. The first step in doing this is to linearize the nonlinear delay differential equations around the equilibrium point. This can be achieved by setting the queue lengths to

$$q_1(t) = \frac{q_\infty(t)}{2} + u(t) \tag{2.19}$$

$$q_2(t) = \frac{q_\infty(t)}{2} - u(t) \tag{2.20}$$

where $u(t)$ is a perturbation function about the equilibrium point $\frac{q_\infty(t)}{2}$. By substituting Eqs. (2.19)–(2.20) into Eqs. (2.1)–(2.2), respectively, and linearizing around the point $u(t) = 0$, we have that the perturbation function solves the following delay differential equation

$$\dot{u}(t) = -\frac{\lambda}{2}\cdot u(t-\Delta) - \mu\cdot u(t). \tag{2.21}$$

Therefore, it only remains for us to analyze Eq. (2.21) to understand the stability of the constant delay queueing system.

Now, by substituting $u(t) = \exp(i\omega t)$ in Eq. (2.21) and applying the techniques from Pender et al. [31], we obtain the following values for ω_{cr} and Δ_{cr}:

$$\omega_{\text{cr}} = \frac{1}{2}\sqrt{\lambda^2-4\mu^2} \tag{2.22}$$

$$\Delta_{\text{cr}} = \frac{2\arccos(-2\mu/\lambda)}{\sqrt{\lambda^2-4\mu^2}}. \tag{2.23}$$

Note that Eq. (2.21) possesses a special solution for $\Delta = \Delta_{\text{cr}}$ that is given by:

$$u(t) = A\cos\omega_{\text{cr}}t + B\sin\omega_{\text{cr}}t \tag{2.24}$$

where A and B are arbitrary constants.

2.3 Asymptotic expansions for constant delay model

Now that we understand the case where $\epsilon = 0$, it remains for us to understand the general case. One important observation to make is that in the previous subsection, we did not use the arrival rate in any way. Therefore, the same analysis can be repeated with the time-varying arrival rate with no changes. Following the same steps as in the case $\epsilon = 0$, we arrive at the case where we need to analyze the following delay differential equation

$$\dot{z}(t) = -\frac{\lambda}{2}(1 + \alpha \cdot \epsilon \cdot \sin \gamma t) \cdot z(t - \Delta) - \mu z(t), \quad \epsilon \ll 1. \tag{2.25}$$

However, since the arrival rate is not constant this time, we do not have a simple way to find the stability of the equation. Therefore, we will exploit the fact that the time-varying arrival rate is near the constant arrival rate and use the two-variable expansion method or the method of multiple time scales developed by Kevorkian and Cole [23].

Theorem 2.4 *The only resonant frequency γ of the time-varying arrival rate function for the first-order two-variable expansion is $\gamma = 2\omega_{\text{cr}}$. For this value of γ, the change in stability occurs at the value Δ_{mod} where*

$$\Delta_{\text{mod}} = \Delta_{\text{cr}} - \epsilon \sqrt{\frac{\alpha^2}{\lambda^2 - 4\mu^2}}. \tag{2.26}$$

Proof We expand time into two variables ξ and η that represent regular and slow time, respectively, i.e.,

$$\xi = t \quad \text{(regular time)} \quad \text{and} \quad \eta = \epsilon t \quad \text{(slow time)}. \tag{2.27}$$

Therefore, $z(t)$ now becomes $z(\xi, \eta)$, and

$$\dot{z}(t) = \frac{\mathrm{d}z}{\mathrm{d}t} = \frac{\partial z}{\partial \xi}\frac{\mathrm{d}\xi}{\mathrm{d}t} + \frac{\partial z}{\partial \eta}\frac{\mathrm{d}\eta}{\mathrm{d}t} = \frac{\partial z}{\partial \xi} + \epsilon \frac{\partial z}{\partial \eta} \tag{2.28}$$

Moreover, we have that

$$z(t - \Delta) = z(\xi - \Delta, \eta - \epsilon\Delta) \tag{2.29}$$

In discussing the dynamics of 2.25, we will detune the delay Δ off of its critical value:

$$\Delta = \Delta_{\text{cr}} + \epsilon\Delta_1 + O(\epsilon^2) \tag{2.30}$$

Substituting Eq. (2.30) into Eq. (2.29) and expanding term by term, we get

$$z(t - \Delta) = \bar{z} - \epsilon\Delta_1 \frac{\partial \bar{z}}{\partial \xi} - \epsilon\Delta_{\text{cr}} \frac{\partial \bar{z}}{\partial \eta} + O(\epsilon^2) \tag{2.31}$$

where

$$\bar{z} = z(\xi - \Delta_{\text{cr}}, \eta).$$

Equation 2.25 becomes, neglecting terms of $O(\epsilon^2)$,

$$\frac{\partial z}{\partial \xi} + \epsilon \frac{\partial z}{\partial \eta} = -\mu z - \frac{\lambda}{2}(1 + \alpha \cdot \epsilon \cdot \sin \gamma t) \times \left(\bar{z} - \epsilon\Delta_1 \frac{\partial \bar{z}}{\partial \xi} - \epsilon\Delta_{\text{cr}} \frac{\partial \bar{z}}{\partial \eta}\right) \tag{2.32}$$

Now, we expand z in a power series in ϵ:

$$z = z_0 + \epsilon z_1 + O(\epsilon^2) \tag{2.33}$$

Substituting (2.33) into (2.32), collecting terms, and equating similar powers of ϵ, we get

$$\frac{\partial z_0}{\partial \xi} + \mu z_0 + \frac{\lambda}{2}\bar{z}_0 = 0 \tag{2.34}$$

$$\frac{\partial z_1}{\partial \xi} + \mu z_1 + \frac{\lambda}{2}\bar{z}_1 = -\frac{\partial z_0}{\partial \eta} + \frac{\lambda}{2}\left(\Delta_1 \frac{\partial \bar{z}_0}{\partial \xi} + \Delta_{\text{cr}} \frac{\partial \bar{z}_0}{\partial \eta} - \alpha \bar{z}_0 \sin \gamma \xi\right) \tag{2.35}$$

Equation 2.34 has the solution given in Eq. (2.24):

$$z_0 = A(\eta) \cos \omega_{\text{cr}} \xi + B(\eta) \sin \omega_{\text{cr}} \xi. \tag{2.36}$$

The functions $A(\eta)$ and $B(\eta)$ give the slow flow of the system. We find differential equations on $A(\eta)$ and $B(\eta)$ by substituting Eqs. (2.36) into (2.35) and eliminating the resonant terms.

The next step is to substitute (2.36) into (2.35). The quantity $\bar{z}_0$ in (2.35) may be conveniently computed from the following expression, obtained from (2.34):

$$\begin{aligned}\bar{z}_0 &= \frac{2}{\lambda} \cdot \left(-\mu z_0 - \frac{\partial z_0}{\partial \xi}\right) \\ &= \frac{2}{\lambda} \cdot [-(\mu A + \omega_{\text{cr}} B) \cos \omega_{\text{cr}} \xi \\ &\quad + (\omega_{\text{cr}} A - \mu B) \sin \omega_{\text{cr}} \xi]\end{aligned} \tag{2.37}$$

Therefore, we have the following expressions for the terms in Eq. (2.35)

$$\frac{\partial z_0}{\partial \eta} = A' \cdot \cos(\omega_{\rm cr}\xi) + B' \cdot \sin(\omega_{\rm cr}\xi) \tag{2.38}$$

$$\frac{\partial \bar{z}_0}{\partial \xi} = \frac{2 \cdot \omega_{\rm cr}}{\lambda}\left[(A\omega_{\rm cr} - \mu B) \cdot \cos(\omega_{\rm cr}\xi) + (\mu A + B\omega_{\rm cr}) \cdot \sin(\omega_{\rm cr}\xi)\right] \tag{2.39}$$

$$\frac{\partial \bar{z}_0}{\partial \eta} = -\frac{2}{\lambda}\left[(\mu A' + B'\omega_{\rm cr}) \cdot \cos(\omega_{\rm cr}\xi) + (\mu B' - A'\omega_{\rm cr}) \cdot \sin(\omega_{\rm cr}\xi)\right] \tag{2.40}$$

$$\begin{aligned}\alpha \bar{z}_0 \sin\gamma\xi &= -\alpha \cdot [(\mu A + B\omega_{\rm cr}) \cdot \cos(\omega_{\rm cr}\xi) \\ &\quad + (\mu B - A\omega_{\rm cr}) \cdot \sin(\omega_{\rm cr}\xi)] \cdot \sin(\gamma\xi) \\ &= \frac{\alpha}{2}(A\omega_{\rm cr} - B\mu)\left[\cos((\gamma - \omega_{\rm cr})\xi) - \cos((\gamma + \omega_{\rm cr})\xi)\right] \\ &\quad - \frac{\alpha}{2}(A\mu + B\omega_{\rm cr})\left[\sin((\gamma - \omega_{\rm cr})\xi) + \sin((\gamma + \omega_{\rm cr})\xi)\right]\end{aligned} \tag{2.41}$$

Thus, after substituting Eqs. (2.36) into (2.35) and applying angle-sum identities, the only terms involving γ are of the form

$$\cos((\gamma \pm \omega_{\rm cr})\xi), \quad \sin((\gamma \pm \omega_{\rm cr})\xi) \tag{2.42}$$

Notice that $\gamma = 2\omega_{\rm cr}$ is the only resonant frequency for the arrival function. For any other value of γ, the terms involving γ at $O(\epsilon)$ are non-resonant, and the first-order two-variable expansion method does not capture any effect from the time-varying arrival function. This 2 to 1 resonance is a similar phenomenon to that arising from ordinary differential equations involving parametric excitation, see for example Ng and Rand [27], Ruelas et al. [33]. Therefore, we set $\gamma = 2\omega_{\rm cr}$, and Eq. (2.35) becomes

$$\begin{aligned}&\frac{\partial z_1}{\partial \xi} + \mu z_1 + \frac{\lambda}{2}\bar{z}_1 \\ &= \left[c_1 A'(\eta) + c_2 B'(\eta) + c_3 A(\eta) + c_4 B(\eta)\right] \\ &\quad \times \cos(\omega_{\rm cr}\xi) \\ &\quad + \left[c_5 A'(\eta) + c_6 B'(\eta) + c_7 A(\eta) + c_8 B(\eta)\right] \\ &\quad \times \sin(\omega_{\rm cr}\xi) \\ &\quad + \text{non-resonant terms}\end{aligned} \tag{2.43}$$

where

$$c_1 = 1 + \mu\Delta_{\rm cr}, \quad c_2 = \Delta_{\rm cr}\omega_{\rm cr}, \quad c_3 = \frac{\alpha\omega_{\rm cr}}{2} - \Delta_1\omega_{\rm cr}^2, \quad c_4 = -\frac{\alpha\mu}{2} + \Delta_1\mu\omega_{\rm cr} \tag{2.44}$$

$$c_5 = -\Delta_{\rm cr}\omega_{\rm cr}, \quad c_6 = 1 + \mu\Delta_{\rm cr}, \quad c_7 = -\frac{\alpha\mu}{2} - \Delta_1\mu\omega_{\rm cr}, \quad c_8 = -\frac{\alpha\omega_{\rm cr}}{2} + \Delta_1\omega_{\rm cr}^2 \tag{2.45}$$

Elimination of secular terms gives the slow flow:

$$\frac{{\rm d}A}{{\rm d}\eta} = K_1 A(\eta) + K_2 B(\eta) \tag{2.46}$$

$$\frac{{\rm d}B}{{\rm d}\eta} = K_3 A(\eta) + K_4 B(\eta) \tag{2.47}$$

where

$$K_1 = -\frac{\omega_{\rm cr}(2\alpha\Delta_{\rm cr}\mu + \alpha - 2\Delta_1\omega_{\rm cr})}{2\left(\Delta_{\rm cr}^2\omega_{\rm cr}^2 + (\Delta_{\rm cr}\mu + 1)^2\right)} \tag{2.48}$$

$$K_2 = \frac{\alpha\left(\Delta_{\rm cr}\mu^2 - \Delta_{\rm cr}\omega_{\rm cr}^2 + \mu\right) - 2\Delta_1\omega_{\rm cr}\left(\Delta_{\rm cr}\mu^2 + \Delta_{\rm cr}\omega_{\rm cr}^2 + \mu\right)}{2\left(\Delta_{\rm cr}^2\omega_{\rm cr}^2 + (\Delta_{\rm cr}\mu + 1)^2\right)} \tag{2.49}$$

$$K_3 = \frac{\alpha\left(\Delta_{\rm cr}\mu^2 - \Delta_{\rm cr}\omega_{\rm cr}^2 + \mu\right) + 2\Delta_1\omega_{\rm cr}\left(\Delta_{\rm cr}\mu^2 + \Delta_{\rm cr}\omega_{\rm cr}^2 + \mu\right)}{2\left(\Delta_{\rm cr}^2\omega_{\rm cr}^2 + (\Delta_{\rm cr}\mu + 1)^2\right)} \tag{2.50}$$

$$K_4 = \frac{\omega_{\rm cr}(2\alpha\Delta_{\rm cr}\mu + \alpha + 2\Delta_1\omega_{\rm cr})}{2\left(\Delta_{\rm cr}^2\omega_{\rm cr}^2 + (\Delta_{\rm cr}\mu + 1)^2\right)} \tag{2.51}$$

The equilibrium point $A(\eta) = B(\eta) = 0$ of the slow flow corresponds to a periodic solution for z_0, and the stability of the equilibrium corresponds to the stability of that periodic solution. The stability is determined by the eigenvalues of the matrix

$$K = \begin{bmatrix} K_1 & K_2 \\ K_3 & K_4 \end{bmatrix} \tag{2.52}$$

If both eigenvalues have negative real part, the equilibrium is stable. Since the eigenvalues are cumbersome to work with directly, we use the Routh–Hurwitz stability criterion:

Denote the characteristic polynomial of K by

$$\det(K - rI) = a_0 + a_1 r + a_2 r^2 = 0 \tag{2.53}$$

Then both eigenvalues have negative real part if and only if all the coefficients satisfy $a_i > 0$. From Eqs.

(2.48)–(2.51) and 2.53, and using the expression for ω_{cr} from Eq. (2.22), we have

$$a_0 = \frac{\left(\mu^2+\omega_{cr}^2\right)\left(4\Delta_1^2\omega_{cr}^2-\alpha^2\right)}{4\left(\Delta_{cr}^2\omega_{cr}^2+(\Delta_{cr}\mu+1)^2\right)} = \frac{\lambda^2\left(\Delta_1^2\left(\lambda^2-4\mu^2\right)-\alpha^2\right)}{4\left(\Delta_{cr}^2\lambda^2+8\Delta_{cr}\mu+4\right)} \tag{2.54}$$

$$a_1 = -\frac{2\Delta_1\omega_{cr}^2}{\Delta_{cr}^2\omega_{cr}^2+(\Delta_{cr}\mu+1)^2} = -\frac{2\Delta_1\left(\lambda^2-4\mu^2\right)}{\Delta_{cr}^2\lambda^2+8\Delta_{cr}\mu+4} \tag{2.55}$$

$$a_2 = 1 \tag{2.56}$$

Recall that ω_{cr} is real and positive only if $\lambda > 2\mu$. So, using this restriction, we find that all of the a_i are positive if and only if

$$\Delta_1 < -\frac{|\alpha|}{2\omega_{cr}} = -\sqrt{\frac{\alpha^2}{\lambda^2-4\mu^2}}. \tag{2.57}$$

Note that we can recover the case with no resonant forcing by setting the forcing amplitude $\alpha = 0$. With no forcing, the periodic solution for z_0 becomes unstable at $\Delta = \Delta_{cr}$, but with resonant forcing, the change of stability occurs when

$$\Delta_{mod} = \Delta_{cr} - \epsilon\sqrt{\frac{\alpha^2}{\lambda^2-4\mu^2}}. \tag{2.58}$$

□

2.4 Numerics for constant delay queueing model

In this section, we numerically integrate the delay two examples of delay differential equations with constant delays and compare the asymptotic results for determining the Hopf bifurcation that occurs. On the left of Fig. 1, we numerically integrate the two queues and plot the queue length as a function of time. In this example our lag in information is given by $\Delta = 1.947$. We see that the two equations eventually converge to the same limit as time is increased toward infinity. This implies that the system is stable and no oscillations or asynchronous dynamics will occur due to instability in this case. On the right of Fig. 1 is a zoomed-in version of the figure on the left. It is clear that the two delay equations

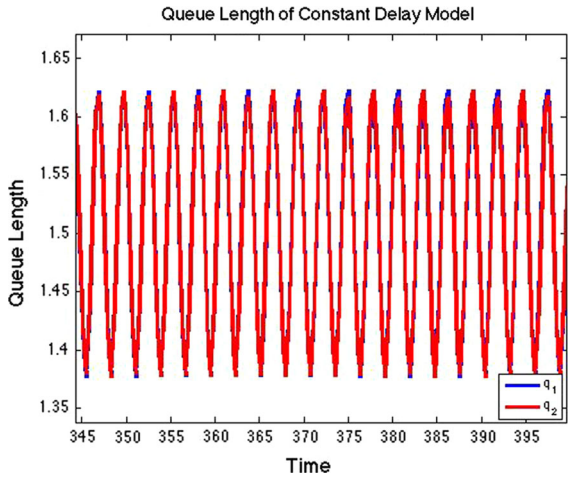

Fig. 1 $\Delta_{cr} = 2.0577$, $\Delta_{mod} = 1.9682$. $\lambda = 3$, $\mu = 1$, $\alpha = 1$, $\epsilon = .2$, $\gamma = \sqrt{5}$, $\Delta = 1.947$, $\varphi_1([-\Delta, 0]) = 1$,$\varphi_2([-\Delta, 0]) = 2$

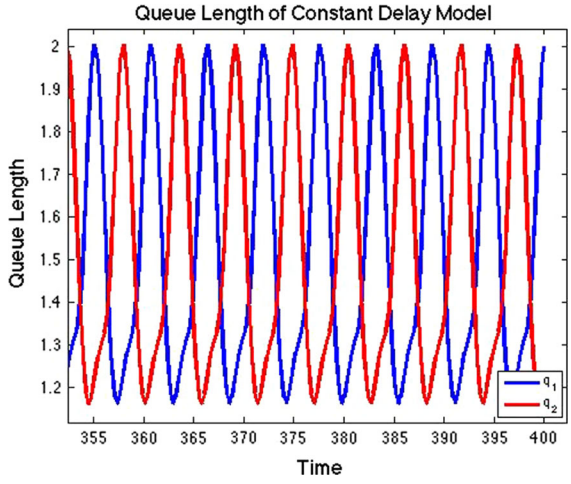

Fig. 2 $\Delta_{cr} = 2.0577$, $\Delta_{mod} = 1.9682$. $\lambda = 3$, $\mu = 1$, $\alpha = 1$, $\epsilon = .2$, $\gamma = \sqrt{5}$, $\Delta = 1.977$, $\varphi_1([-\Delta, 0]) = 1$,$\varphi_2([-\Delta, 0]) = 2$

are converging toward one another and this system is stable. However, in Fig. 2 we use the same parameters, but we make the lag in information $\Delta = 1.977$. This is below the critical delay in the constant case and above the modified critical delay when the time-varying arrival rate is taken into account. On the right of Fig. 2, we display a zoomed-in version of the figure on the left. We see that in this case the two queues oscillate and asynchronous behavior is observed. Thus, the asymptotic analysis performed works well at predicting the change in stability.

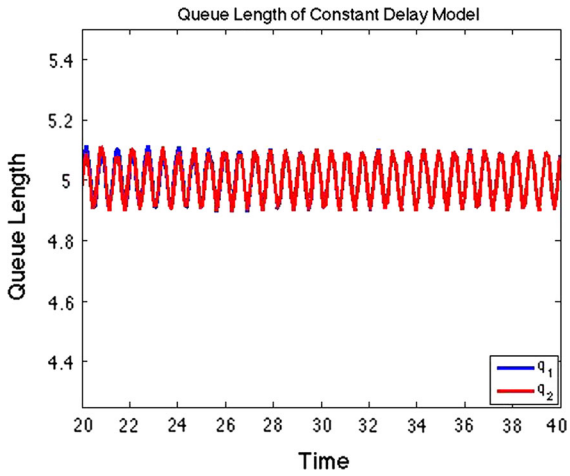

Fig. 3 $\Delta_{cr} = .3617$, $\Delta_{mod} = .3413$. $\lambda = 10$, $\mu = 1$, $\alpha = 1$, $\epsilon = .2$, $\gamma = \sqrt{96}$, $\Delta = .33$, $\varphi_1([-\Delta, 0]) = 3, \varphi_2([-\Delta, 0]) = 4$

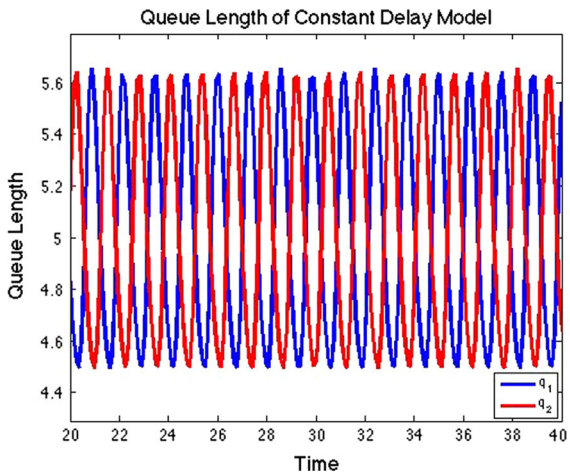

Fig. 4 $\Delta_{cr} = .3617$, $\Delta_{mod} = .3413$. $\lambda = 10$, $\mu = 1$, $\alpha = 1$, $\epsilon = .2$, $\gamma = \sqrt{96}$, $\Delta = .35$, $\varphi_1([-\Delta, 0]) = 3, \varphi_2([-\Delta, 0]) = 4$

As an additional numerical example, on the left of Fig. 3, we numerically integrate the two queues and plot the queue length as a function of time. In this example, our lag in information is given by $\Delta = .33$. We see that the two equations eventually converge to the same limit as time is increased toward infinity. This implies that the system is stable and no oscillations or asynchronous dynamics will occur due to instability in this case. On the right of Fig. 3 is a zoomed-in version of the figure on the left. It is clear that the two delay equations are converging toward one another and this system is stable. However, in Fig. 4, we use the same parameters, but we make the lag in information $\Delta = .35$. This is below the critical delay in the constant case and above the modified critical delay when the time-varying arrival rate is taken into account. On the right of Fig. 4, we display a zoomed-in version of the figure on the left. We see that in this case the two queues oscillate and asynchronous behavior is observed. Thus, the asymptotic analysis performed works well at predicting the change in stability.

3 Moving average delay fluid model

In this section, we present another fluid model with customer choice and where the delay information presented to the customer is a moving average. This model assumes that customers are informed about the queue length, but in the form of a moving average of the queue length between the current time and Δ time units in the past. These types of moving average models are currently used in many healthcare settings such as the one in Fig. 5. In Fig. 5, it is clear that the time information is given in past and is averages over a 4-h window. This is partially because patients in healthcare are quite heterogeneous and require different services and attention. Moreover, the system is not necessary FIFO or FCFS since patients have different priority levels. Thus, the moving average waiting time indicator might be attractive for these reasons. Like in the previous model with constant delays, customers in the moving average model also have the choice to join two parallel infinite server queues and they join according to the same multinomial logit model. Once again, the extension to more complicated arrival functions such as $\lambda(t) = \lambda + \lambda \cdot \alpha \sin(\gamma t)$ is quite difficult. However, like in the constant delay setting, we can analyze the system when the time-varying arrival rate is close to the

Fig. 5 Emergency room wait times via moving averages

constant rate case using perturbation theory and asymptotics. Thus, we assume that the queue length equations for the constant delay model satisfy the following delay differential equations

$$\lambda(t) \cdot \frac{\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)\mathrm{d}s\right)}{\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)\mathrm{d}s\right)+\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)\mathrm{d}s\right)} \tag{3.59}$$

and join the second queue at rate

$$\lambda(t) \cdot \frac{\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)\mathrm{d}s\right)}{\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)\mathrm{d}s\right)+\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)\mathrm{d}s\right)}. \tag{3.60}$$

Thus, our model for customer choice with delayed information in the form of a moving average can be represented by a two-dimensional system of functional differential equations

$$\begin{aligned}\dot{q}_1(t) &= (\lambda + \lambda\alpha\epsilon\sin(\gamma t)) \\ &\quad \times \frac{\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)\mathrm{d}s\right)}{\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)\mathrm{d}s\right)+\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)\mathrm{d}s\right)} \\ &\quad - \mu q_1(t)\end{aligned} \tag{3.61}$$

$$\begin{aligned}\dot{q}_2(t) &= (\lambda + \lambda\alpha\epsilon\sin(\gamma t)) \\ &\quad \times \frac{\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)\mathrm{d}s\right)}{\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_1(s)\mathrm{d}s\right)+\exp\left(-\frac{1}{\Delta}\int_{t-\Delta}^{t} q_2(s)\mathrm{d}s\right)} \\ &\quad - \mu q_2(t)\end{aligned} \tag{3.62}$$

where we assume that q_1 and q_2 start at different initial functions $\varphi_1(t)$ and $\varphi_2(t)$ on the interval $[-\Delta, 0]$.

On the onset, this problem is seemingly more difficult than the constant delay setting since the ratio now depends on a moving average of the queue length during a delay period Δ. To simplify the notation, we find it useful to define the moving average of the i^{th} queue over the time interval $[t-\Delta, t]$ as

$$m_i(t, \Delta) = \frac{1}{\Delta}\int_{t-\Delta}^{t} q_i(s)\mathrm{d}s. \tag{3.63}$$

This representation of the moving average leads to a key observation where we discover that the moving average itself solves a linear delay differential equation. In fact, by differentiating Eq. (3.63) with respect to time, it can be shown that the moving average of the i^{th} queue is the solution to the following delay differential equation

$$\dot{m}_i(t, \Delta) = \frac{1}{\Delta} \cdot (q_i(t) - q_i(t-\Delta)), \quad i \in \{1, 2\}. \tag{3.64}$$

Leveraging the above delay equation for the moving average, we can describe our moving average fluid model with the following four dimensional system of delay differential equations

$$\begin{aligned}\dot{q}_1 &= (\lambda + \lambda \cdot \alpha \cdot \epsilon \cdot \sin(\gamma t)) \\ &\quad \cdot \frac{\exp(-m_1(t))}{\exp(-m_1(t)) + \exp(-m_2(t))} - \mu q_1(t)\end{aligned} \tag{3.65}$$

$$\begin{aligned}\dot{q}_2 &= (\lambda + \lambda \cdot \alpha \cdot \epsilon \cdot \sin(\gamma t)) \\ &\quad \cdot \frac{\exp(-m_2(t))}{\exp(-m_1(t)) + \exp(-m_2(t))} - \mu q_2(t)\end{aligned} \tag{3.66}$$

$$\dot{m}_1 = \frac{1}{\Delta} \cdot (q_1(t) - q_1(t-\Delta)) \tag{3.67}$$

$$\dot{m}_2 = \frac{1}{\Delta} \cdot (q_2(t) - q_2(t-\Delta)). \tag{3.68}$$

In the moving average model, it is also critical to understand the case when the arrival rate is constant and does not depend on time. In Pender et al. [31], the authors show that the critical delay for the moving average model can be determined from the model parameters and the following theorem is from Pender et al. [31].

Theorem 3.1 *For the moving average fluid model given by Eqs. 3.65–3.68, the critical delay parameter is the solution to the following transcendental equation*

$$\sin\left(\Delta \cdot \sqrt{\frac{\lambda}{\Delta} - \mu^2}\right) + \frac{2\mu\Delta}{\lambda} \cdot \sqrt{\frac{\lambda}{\Delta} - \mu^2} = 0. \tag{3.69}$$

Proof See Pender et al. [31]. □

In order to begin our analysis of the delay differential equations with a time-varying rate, we need to first understand the case where $\epsilon = 0$ and the arrival rate is constant. Also like in the constant delay setting, this analysis has been carried out in Pender et al. [30,31] and we give a brief outline of the analysis for the reader's convenience.

The first step to understanding the case when $\epsilon = 0$ to compute the equilibrium in this case. The first part of the proof is to compute an equilibrium for the solution

to the delay differential equations. In our case, the delay differential equations given in Eqs. (3.65)–(3.68) are symmetric. Moreover, in the case where there is no delay, the two equations converge to the same point since in equilibrium each queue will receive exactly one half of the arrivals and the two service rates are identical. This is also true in the case where the arrival process contains delays in the queue length since in equilibrium, the delayed queue length is equal to the non-delayed queue length. It can be shown that there is only one equilibrium where all of the states are equal to each other. One can prove this by substituting $q_2 = \lambda/\mu - q_1$ in the steady-state version of Eq. (3.65) and solving for q_1. One eventually sees that $q_1 = q_2$ is the only solution since any other solution does not obey Eq. (3.65). Thus, we have in equilibrium that

$$q_1(t) = q_2(t) = \frac{q_\infty(t)}{2} \quad \text{as } t \to \infty \tag{3.70}$$

and

$$m_1(t) = m_2(t) = \frac{1}{\Delta}\int_{t-\Delta}^{t} \frac{q_\infty(s)}{2}\mathrm{d}s \quad \text{as } t \to \infty. \tag{3.71}$$

Now that we know the equilibrium for Eqs. (3.65)–(3.68), we need to understand the stability of the delay differential equations around the equilibrium. The first step in doing this is to set each of the queue lengths to the equilibrium values plus a perturbation. Thus, we set each of the queue lengths to

$$q_1(t) = \frac{q_\infty(t)}{2} + u(t) \tag{3.72}$$

$$q_2(t) = \frac{q_\infty(t)}{2} - u(t) \tag{3.73}$$

$$m_1(t) = \frac{1}{\Delta}\int_{t-\Delta}^{t} \frac{q_\infty(s)}{2}\mathrm{d}s + w(t) \tag{3.74}$$

$$m_2(t) = \frac{1}{\Delta}\int_{t-\Delta}^{t} \frac{q_\infty(s)}{2}\mathrm{d}s - w(t) \tag{3.75}$$

Substitute Eqs. (3.72)–(3.75) into (3.65)–(3.68) and solve for $\dot{q}_\infty$, $\dot{u}$ and $\dot{w}$.

$$\dot{q}_\infty = \lambda + \lambda\alpha\epsilon\sin(\gamma t) - \mu q_\infty(t) \tag{3.76}$$

$$\dot{u} = -\tfrac{\lambda}{2}\left(1 + \alpha\epsilon\sin(\gamma t)\right)\tanh(w(t)) - \mu u(t) \tag{3.77}$$

$$\dot{w} = \tfrac{1}{\Delta}\left(u(t) - u(t-\Delta)\right) \tag{3.78}$$

Equation 3.76 can be solved explicitly, to give the steady-state solution

$$q_\infty(t) = ce^{-\mu t} + \frac{\lambda}{2} \times \left(\frac{1}{\mu} + \frac{\alpha\epsilon(\mu\sin(\gamma t) - \gamma\cos(\gamma t))}{\gamma^2 + \mu^2}\right) \tag{3.79}$$

where

$$c = q_\infty(0) - \frac{\lambda}{2}\left(\frac{1}{\mu} - \frac{\alpha\gamma\epsilon}{\gamma^2 + \mu^2}\right) \tag{3.80}$$

To determine the stability of the system, we linearize about the point $u(t) = w(t) = 0$, giving

$$\dot{u} = -\tfrac{\lambda}{2}\left(1 + \alpha\epsilon\sin(\gamma t)\right)w(t) - \mu u(t) \tag{3.81}$$

$$\dot{w} = \tfrac{1}{\Delta}\left(u(t) - u(t-\Delta)\right) \tag{3.82}$$

First consider the unperturbed case ($\epsilon = 0$):

$$\dot{u} = -\tfrac{\lambda}{2}w(t) - \mu u(t) \tag{3.83}$$

$$\dot{w} = \tfrac{1}{\Delta}\left(u(t) - u(t-\Delta)\right) \tag{3.84}$$

To study Eqs. (3.83) and (3.84), we set

$$u = A\exp(rt) \tag{3.85}$$

$$w = B\exp(rt). \tag{3.86}$$

These solutions imply the following relationships between the constants A, B, and r.

$$Ar = -\frac{\lambda}{2}B - \mu A \tag{3.87}$$

$$Br = \frac{1}{\Delta}(A - A\exp(-r\Delta)) \tag{3.88}$$

solving for A yields

$$A = -\frac{\lambda}{2(\mu + r)}B \tag{3.89}$$

and rearranging yields the following equation for r

$$r = \frac{\lambda}{2\Delta\cdot r}(\exp(-r\Delta) - 1) - \mu. \tag{3.90}$$

Now it remains for us to understand the transition between stable and unstable solutions once again.

To find the transition between stable and unstable solutions, set $r = i\omega$, giving us the following equation

$$i\omega = \frac{\lambda}{2\Delta i\omega}(\exp(-i\omega\Delta) - 1) - \mu. \tag{3.91}$$

Multiplying both sides by $i\omega$ and using Euler's identity, we have that

$$\frac{\lambda}{2\Delta}(\cos(\omega\Delta) - i\sin(\omega\Delta) - 1) - \mu i\omega + \omega^2 = 0. \tag{3.92}$$

Writing the real and imaginary parts of Eq. (3.92), we get:

$$\cos(\omega\Delta) = 1 - \frac{2\Delta\omega^2}{\lambda} \tag{3.93}$$

for the real part and

$$\sin(\omega\Delta) = -\frac{2\Delta\mu\omega}{\lambda} \tag{3.94}$$

Once again by squaring and adding $\sin\omega\Delta$ and $\cos\omega\Delta$ together, we get:

$$\omega = \sqrt{\frac{\lambda}{\Delta} - \mu^2} \tag{3.95}$$

Finally by substituting the expression for ω into Eqs. 3.94 and 3.93 we obtain the final expression for the critical delay Δ_{cr}. The expression for the critical delay Δ_{cr} is also the simultaneous solution to the following transcendental equations:

$$\sin\left(\Delta_{\text{cr}}\sqrt{\frac{\lambda}{\Delta_{\text{cr}}} - \mu^2}\right) + \frac{2\mu\Delta_{\text{cr}}}{\lambda}\sqrt{\frac{\lambda}{\Delta_{\text{cr}}} - \mu^2} = 0 \tag{3.96}$$

$$\cos\left(\Delta_{\text{cr}}\sqrt{\frac{\lambda}{\Delta_{\text{cr}}} - \mu^2}\right) + 1 - \frac{2\mu^2\Delta_{\text{cr}}}{\lambda} = 0 \tag{3.97}$$

Squaring Eqs. (3.96) and (3.97) and adding them, we see that they are satisfied simultaneously when

$$2 + \left(2 - \frac{4\Delta_{\text{cr}}\mu^2}{\lambda}\right)\cos\left(\Delta_{\text{cr}}\sqrt{\frac{\lambda}{\Delta_{\text{cr}}} - \mu^2}\right) + \frac{4\Delta_{\text{cr}}\mu}{\lambda}\sqrt{\frac{\lambda}{\Delta_{\text{cr}}} - \mu^2}\sin\left(\Delta_{\text{cr}}\sqrt{\frac{\lambda}{\Delta_{\text{cr}}} - \mu^2}\right) = 0 \tag{3.98}$$

3.1 Asymptotic expansions for moving average model

Now that we understand the case where $\epsilon = 0$, it remains for us to understand the general case. Recall that we are analyzing the stability of the linearized system

$$\dot{u} = -\tfrac{\lambda}{2}\left(1 + \alpha\epsilon\sin(\gamma t)\right)w(t) - \mu u(t) \tag{3.81}$$

$$\dot{w} = \tfrac{1}{\Delta}\left(u(t) - u(t-\Delta)\right) \tag{3.82}$$

It is useful to convert the system of two first-order equations to a single second-order equation, by differentiating Eq. (3.81) and substituting in expressions for $w(t)$ and $\dot{w}(t)$ from Eqs. (3.81) and (3.82). We obtain

$$\begin{aligned}\ddot{u} = &\left(\frac{\alpha\gamma\epsilon\cos(\gamma t)}{\alpha\epsilon\sin(\gamma t) + 1} - \mu\right)\dot{u} \\ &+ \left(\frac{\alpha\gamma\mu\epsilon\cos(\gamma t)}{\alpha\epsilon\sin(\gamma t) + 1} - \frac{\lambda + \alpha\lambda\epsilon\sin(\gamma t)}{2\Delta}\right)u \\ &+ \left(\frac{\lambda + \alpha\lambda\epsilon\sin(\gamma t)}{2\Delta}\right)u(t-\Delta)\end{aligned} \tag{3.99}$$

However, since the arrival rate is not constant this time, we do not have a simple way to find the stability of the equation. Therefore, we will exploit the fact that the time-varying arrival rate is near the constant arrival rate and use the two-variable expansion method.

Theorem 3.2 *The only resonant frequency γ of the time-varying arrival rate function for the first-order two-variable expansion is $\gamma = 2\omega_{\text{cr}}$. For this value of γ, the change in stability occurs at $\Delta)_{\text{mod}}$ where*

$$\Delta_{\text{mod}} = \Delta_{\text{cr}} \pm \epsilon\sqrt{\frac{\alpha^2\Delta_{\text{cr}}^2}{\Delta_{\text{cr}}\lambda + 4\Delta_{\text{cr}}\mu + 4}} \tag{3.100}$$

where the sign of the ϵ term is positive if $\Delta_{\text{cr}} > \frac{\lambda - 2\mu}{2\mu^2}$ and negative if $\Delta_{\text{cr}} < \frac{\lambda - 2\mu}{2\mu^2}$.

Proof We expand time into two variables ξ and η that represent regular and slow time, respectively, i.e.,

$$\xi = t \quad \text{(regular time)} \quad \text{and} \quad \eta = \epsilon t \quad \text{(slow time)}. \tag{3.101}$$

Therefore, $u(t)$ now becomes $u(\xi, \eta)$. Moreover,

$$\begin{aligned}\dot{u} &= \frac{\mathrm{d}u}{\mathrm{d}t} = \frac{\partial u}{\partial\xi}\frac{\mathrm{d}\xi}{\mathrm{d}t} + \frac{\partial u}{\partial\eta}\frac{\mathrm{d}\eta}{\mathrm{d}t} \\ &= \frac{\partial u}{\partial\xi} + \epsilon\frac{\partial u}{\partial\eta}\end{aligned} \tag{3.102}$$

$$\begin{aligned}\ddot{u} &= \frac{\mathrm{d}^2u}{\mathrm{d}t^2} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial u}{\partial\xi} + \epsilon\frac{\partial u}{\partial\eta}\right) \\ &= \frac{\mathrm{d}\xi}{\mathrm{d}t}\frac{\partial}{\partial\xi}\left(\frac{\partial u}{\partial\xi} + \epsilon\frac{\partial u}{\partial\eta}\right) + \frac{\mathrm{d}\eta}{\mathrm{d}t}\frac{\partial}{\partial\eta}\left(\frac{\partial u}{\partial\xi} + \epsilon\frac{\partial u}{\partial\eta}\right)\end{aligned}$$

$$= \frac{\partial^2 u}{\partial \xi^2} + 2\epsilon \frac{\partial^2 u}{\partial \xi \partial \eta} + \epsilon^2 \frac{\partial^2 u}{\partial \eta^2} \tag{3.103}$$

Additionally, we have that

$$u(t-\Delta) = u(\xi - \Delta, \eta - \epsilon\Delta) \tag{3.104}$$

In discussing the dynamics of 2.25, we will detune the delay Δ off of its critical value:

$$\Delta = \Delta_{\text{cr}} + \epsilon\Delta_1 + O(\epsilon^2) \tag{3.105}$$

Substituting Eqs. (3.105) into (3.104) and expanding as a series in ϵ, we get

$$u(t-\Delta) = \bar{u} - \epsilon\Delta_1 \frac{\partial \bar{u}}{\partial \xi} - \epsilon\Delta_{\text{cr}} \frac{\partial \bar{u}}{\partial \eta} + O(\epsilon^2) \tag{3.106}$$

where

$$\bar{u} \equiv u(\xi - \Delta_{\text{cr}}, \eta).$$

Now, we expand u in a power series in terms ϵ:

$$u = u_0 + \epsilon u_1 + O(\epsilon^2) \tag{3.107}$$

Substituting Eqs. (3.102, 3.103, 3.106) and (3.107) into Eq. (3.99), expanding as a series in ϵ, collecting like terms, and equating like powers of ϵ, we get

$$\frac{\partial^2 u_0}{\partial \xi^2} + \frac{\partial u_0}{\partial \xi} + \frac{\lambda}{2\Delta_{\text{cr}}}(u_0 - \bar{u}_0) = 0 \tag{3.108}$$

$$\begin{aligned}
&\frac{\partial^2 u_1}{\partial \xi^2} + \frac{\partial u_1}{\partial \xi} + \frac{\lambda}{2\Delta_{\text{cr}}}(u_1 - \bar{u}_1) \\
&= \left(\alpha\gamma\mu\cos(\gamma\xi) + \frac{\lambda\left(\Delta_1 - \alpha\Delta_{\text{cr}}\sin(\gamma\xi)\right)}{2\Delta_{\text{cr}}^2}\right) u_0 \\
&\quad + \frac{\lambda\left(\alpha\Delta_{\text{cr}}\sin(\gamma\xi) - \Delta_1\right)}{2\Delta_{\text{cr}}^2}\bar{u}_0 - \mu\frac{\partial u_0}{\partial \eta} - \frac{\lambda}{2}\frac{\partial \bar{u}_0}{\partial \eta} \\
&\quad + \alpha\gamma\cos(\gamma\xi)\frac{\partial u_0}{\partial \xi} - \frac{\lambda\Delta_1}{2\Delta_{\text{cr}}}\frac{\partial \bar{u}_0}{\partial \xi} - 2\frac{\partial^2 u_0}{\partial \xi \partial \eta}
\end{aligned} \tag{3.109}$$

Equation (3.108) is linear, constant-coefficient, homogeneous, and does not involve any derivatives with respect to η. In fact, it is the equation that results from converting the $\epsilon = 0$ system (Eqs. 3.83–3.84) to a single second-order equation. So, we write down the solution:

$$u_0 = A(\eta)\cos(\omega_{\text{cr}}\xi) + B(\eta)\sin(\omega_{\text{cr}}\xi) \tag{3.110}$$

The functions $A(\eta)$ and $B(\eta)$ give the slow flow of the system. We find differential equations on $A(\eta)$ and $B(\eta)$ by substituting Eqs. (3.110) into (3.109) and eliminating the resonant terms. We compute $\bar{u}_0$ and its partial derivatives using expressions for $\cos(\Delta_{\text{cr}}\omega_{\text{cr}})$ and $\sin(\Delta_{\text{cr}}\omega_{\text{cr}})$ given by Eqs. (3.96) and 3.97. For example:

$$\begin{aligned}
\bar{u}_0 &= A(\eta)\cos(\omega_{\text{cr}}(\xi - \Delta_{\text{cr}})) + B(\eta)\sin(\omega_{\text{cr}}(\xi - \Delta_{\text{cr}})) \\
&= \left(A(\eta)\cos(\Delta_{\text{cr}}\omega_{\text{cr}}) - B(\eta)\sin(\Delta_{\text{cr}}\omega_{\text{cr}})\right)\cos(\omega_{\text{cr}}\xi) \\
&\quad + \left(A(\eta)\sin(\Delta_{\text{cr}}\omega_{\text{cr}}) + B(\eta)\cos(\Delta_{\text{cr}}\omega_{\text{cr}})\right)\sin(\omega_{\text{cr}}\xi) \\
&= \left(B(\eta)\frac{2\Delta_{\text{cr}}\mu}{\lambda}\sqrt{\frac{\lambda - \Delta_{\text{cr}}\mu^2}{\Delta_{\text{cr}}}} - A(\eta)\frac{\left(\lambda - 2\Delta_{\text{cr}}\mu^2\right)}{\lambda}\right) \\
&\quad \times \cos(\omega_{\text{cr}}\xi) \\
&\quad + \left(-A(\eta)\frac{2\Delta_{\text{cr}}\mu}{\lambda}\sqrt{\frac{\lambda - \Delta_{\text{cr}}\mu^2}{\Delta_{\text{cr}}}} - B(\eta)\frac{\left(\lambda - 2\Delta_{\text{cr}}\mu^2\right)}{\lambda}\right) \\
&\quad \times \sin(\omega_{\text{cr}}\xi)
\end{aligned} \tag{3.111}$$

After substituting these expressions into Eq. (3.109) and using angle-sum identities, the remaining trigonometric terms are of the forms

$$\begin{aligned}
&\cos(\omega_{\text{cr}}\xi), \quad \sin(\omega_{\text{cr}}\xi), \quad \cos((\omega_{\text{cr}} \pm \gamma)\xi), \\
&\sin((\omega_{\text{cr}} \pm \gamma)\xi)
\end{aligned} \tag{3.112}$$

Notice that $\gamma = 2\omega_{\text{cr}}$ is the only resonant frequency for the arrival function. For any other value of γ, the terms involving γ at $O(\epsilon)$ are non-resonant, and the first-order two-variable expansion method does not capture any effect from the time-varying arrival function. Therefore, we set $\gamma = 2\omega_{\text{cr}}$, and Eq. (3.109) becomes

$$\begin{aligned}
&\frac{\partial^2 u_1}{\partial \xi^2} + \frac{\partial u_1}{\partial \xi} + \frac{\lambda}{2\Delta_{\text{cr}}}(u_1 - \bar{u}_1) \\
&= \left[c_1 A'(\eta) + c_2 B'(\eta) + c_3 A(\eta) + c_4 B(\eta)\right]\cos(\omega_{\text{cr}}\xi) \\
&\quad + \left[c_5 A'(\eta) + c_6 B'(\eta) + c_7 A(\eta) + c_8 B(\eta)\right]\sin(\omega_{\text{cr}}\xi) \\
&\quad + \text{non-resonant terms}
\end{aligned} \tag{3.113}$$

where the coefficients c_i depend on $\lambda, \mu, \alpha, \Delta_{\text{cr}}$ and Δ_1.

Elimination of secular terms in Eq. (3.113) gives the slow flow equations on $A(\eta)$ and $B(\eta)$:

$$\frac{\mathrm{d}A}{\mathrm{d}\eta} = K_1 A(\eta) + K_2 B(\eta) \tag{3.114}$$

$$\frac{\mathrm{d}B}{\mathrm{d}\eta} = K_3 A(\eta) + K_4 B(\eta) \tag{3.115}$$

where

$$K_1 = \frac{\alpha\Delta_{\rm cr}\sqrt{\frac{\lambda}{\Delta_{\rm cr}}-\mu^2}\left(\mu\Delta_{\rm cr}\left(-4\mu^2\Delta_{\rm cr}+3\lambda-6\mu\right)+4\lambda\right)-2\Delta_1\left(\lambda-\mu^2\Delta_{\rm cr}\right)\left(\lambda-2\mu\left(\mu\Delta_{\rm cr}+1\right)\right)}{\Delta_{\rm cr}\left(\Delta_{\rm cr}\left(8\mu^3\Delta_{\rm cr}-\lambda^2-12\lambda\mu+12\mu^2\right)-16\lambda\right)} \tag{3.116}$$

$$K_2 = \frac{\alpha\Delta_{\rm cr}\left(\lambda-\mu^2\Delta_{\rm cr}\right)\left(-4\mu^2\Delta_{\rm cr}+\lambda-6\mu\right)+\Delta_1\sqrt{\frac{\lambda}{\Delta_{\rm cr}}-\mu^2}\left(\Delta_{\rm cr}\left(-4\mu^3\Delta_{\rm cr}+\lambda^2+8\lambda\mu-4\mu^2\right)+8\lambda\right)}{\Delta_{\rm cr}\left(\Delta_{\rm cr}\left(8\mu^3\Delta_{\rm cr}-\lambda^2-12\lambda\mu+12\mu^2\right)-16\lambda\right)} \tag{3.117}$$

$$K_3 = \frac{\alpha\Delta_{\rm cr}\left(\lambda-\mu^2\Delta_{\rm cr}\right)\left(-4\mu^2\Delta_{\rm cr}+\lambda-6\mu\right)+\Delta_1\sqrt{\frac{\lambda}{\Delta_{\rm cr}}-\mu^2}\left(\Delta_{\rm cr}\left(4\mu^3\Delta_{\rm cr}-\lambda^2-8\lambda\mu+4\mu^2\right)-8\lambda\right)}{\Delta_{\rm cr}\left(\Delta_{\rm cr}\left(8\mu^3\Delta_{\rm cr}-\lambda^2-12\lambda\mu+12\mu^2\right)-16\lambda\right)} \tag{3.118}$$

$$K_4 = \frac{\alpha\Delta_{\rm cr}\sqrt{\frac{\lambda}{\Delta_{\rm cr}}-\mu^2}\left(\mu\Delta_{\rm cr}\left(4\mu^2\Delta_{\rm cr}-3\lambda+6\mu\right)-4\lambda\right)-2\Delta_1\left(-2\mu^2\Delta_{\rm cr}+\lambda-2\mu\right)\left(\lambda-\mu^2\Delta_{\rm cr}\right)}{\Delta_{\rm cr}\left(\Delta_{\rm cr}\left(8\mu^3\Delta_{\rm cr}-\lambda^2-12\lambda\mu+12\mu^2\right)-16\lambda\right)} \tag{3.119}$$

The equilibrium point $A(\eta) = B(\eta) = 0$ of the slow flow corresponds to a periodic solution for u_0, and the stability of the equilibrium corresponds to the stability of that periodic solution. The stability is determined by the eigenvalues of the matrix

$$K = \begin{bmatrix} K_1 & K_2 \\ K_3 & K_4 \end{bmatrix} \tag{3.120}$$

If both eigenvalues have negative real part, the equilibrium is stable. Since the eigenvalues are cumbersome to work with directly, we use the Routh–Hurwitz stability criterion:

Denote the characteristic polynomial of K by

$$\det(K - rI) = a_0 + a_1 r + a_2 r^2 = 0 \tag{3.121}$$

Then both eigenvalues have negative real part if and only if all the coefficients satisfy $a_i > 0$. From Eqs. (3.116)–(3.119), we have

$$a_0 = -\frac{\lambda\left(\lambda-\Delta_{\rm cr}\mu^2\right)\left(\Delta_1^2\left(\Delta_{\rm cr}(\lambda+4\mu)+4\right)-\alpha^2\Delta_{\rm cr}^2\right)}{\Delta_{\rm cr}^3\left(-\Delta_{\rm cr}\lambda^2-4\lambda(3\Delta_{\rm cr}\mu+4)+4\Delta_{\rm cr}\mu^2(2\Delta_{\rm cr}\mu+3)\right)} \tag{3.122}$$

$$a_1 = \frac{4\Delta_1\left(\lambda-\Delta_{\rm cr}\mu^2\right)\left(\lambda-2\mu(\Delta_{\rm cr}\mu+1)\right)}{\Delta_{\rm cr}\left(-\Delta_{\rm cr}\lambda^2-4\lambda(3\Delta_{\rm cr}\mu+4)+4\Delta_{\rm cr}\mu^2(2\Delta_{\rm cr}\mu+3)\right)} \tag{3.123}$$

$$a_2 = 1 \tag{3.124}$$

Recall from Eq. (3.95) that ω is only positive when $0 < \Delta_{\rm cr} < \lambda/\mu^2$. Using this restriction, we find that the coefficients are all positive when

$$\begin{aligned} &0 < \lambda \le 2\mu \\ &\Delta_1 > \sqrt{\frac{\alpha^2\Delta_{\rm cr}^2}{\Delta_{\rm cr}\lambda+4\Delta_{\rm cr}\mu+4}} \end{aligned} \tag{3.125}$$

or alternatively when

$$\begin{aligned} &\lambda > 2\mu \\ &0 < \Delta_{\rm cr} < \frac{\lambda-2\mu}{2\mu^2} \\ &\Delta_1 < -\sqrt{\frac{\alpha^2\Delta_{\rm cr}^2}{\Delta_{\rm cr}\lambda+4\Delta_{\rm cr}\mu+4}} \end{aligned} \tag{3.126}$$

or when

$$\begin{aligned} &\lambda > 2\mu \\ &\frac{\lambda-2\mu}{2\mu^2} < \Delta_{\rm cr} < \frac{\lambda}{\mu^2} \\ &\Delta_1 > \sqrt{\frac{\alpha^2\Delta_{\rm cr}^2}{\Delta_{\rm cr}\lambda+4\Delta_{\rm cr}\mu+4}} \end{aligned} \tag{3.127}$$

Thus, the change of stability occurs at

$$\Delta = \Delta_{\rm cr} \pm \epsilon\sqrt{\frac{\alpha^2\Delta_{\rm cr}^2}{\Delta_{\rm cr}\lambda+4\Delta_{\rm cr}\mu+4}} \tag{3.128}$$

where the sign of the ϵ term depends on $\Delta_{\rm cr}$, λ and μ as in Eqs. (3.125)–(3.127). □

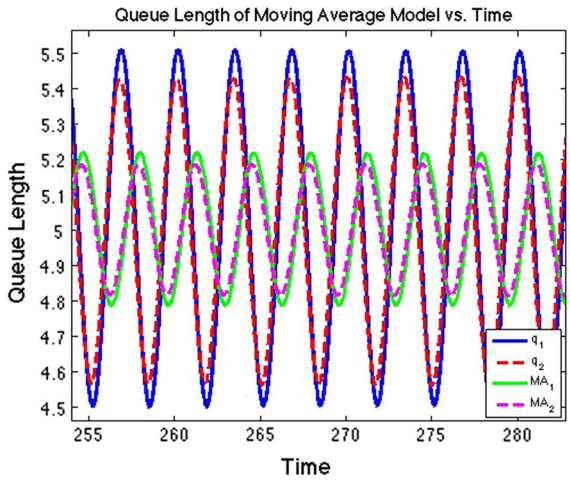

Fig. 6 $\Delta_{cr} = 2.1448$, $\Delta_{mod} = 2.2183$. $\lambda = 10$, $\mu = 1$, $\alpha = 1$, $\epsilon = .2$, $\gamma = \sqrt{10/\Delta_{cr} - 1}$, $\Delta = 2.18$, $\varphi_1([-\Delta, 0]) = 3$,$\varphi_2([-\Delta, 0]) = 4$

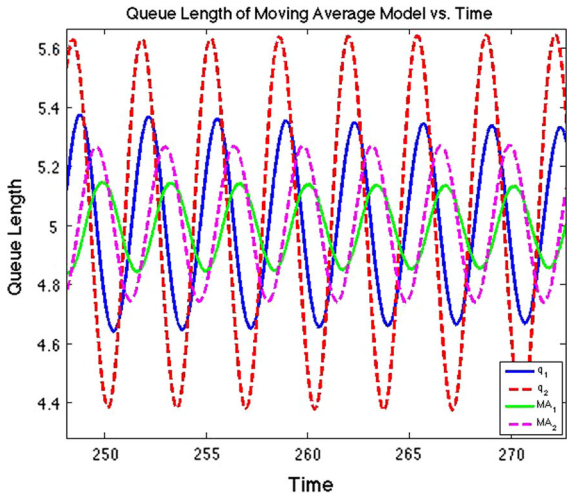

Fig. 7 $\Delta_{cr} = 2.1448$, $\Delta_{mod} = 2.2183$. $\lambda = 10$, $\mu = 1$, $\alpha = 1$, $\epsilon = .2$, $\gamma = \sqrt{10/\Delta_{cr} - 1}$, $\Delta = 2.25$, $\varphi_1([-\Delta, 0]) = 3.9$, $\varphi_2([-\Delta, 0]) = 4$

3.2 Numerics for moving average queueing model

In this section, we numerically integrate the delay two examples of delay differential equations with moving averages and compare the asymptotic results for determining the Hopf bifurcation that occurs. On the left of Fig. 6, we numerically integrate the two queues and plot the queue length as a function of time. In this example, our lag in information is given by $\Delta = 2.18$. We see that the two equations eventually converge to the same limit as time is increased toward infinity. This implies that the system is stable and no oscillations or asynchronous dynamics will occur due to instability in this case. On the right of Fig. 6 is a zoomed-in version of the figure on the left. It is clear that the two delay equations are converging toward one another and this system is stable. However, in Fig. 7 we use the same parameters, but we make the lag in information $\Delta = 2.25$. This is below the critical delay in the constant case and above the modified critical delay when the time-varying arrival rate is taken into account. On the right of Fig. 7, we display a zoomed-in version of the figure on the left. We see that in this case the two queues oscillate and asynchronous behavior is observed. Thus, the asymptotic analysis performed works well at predicting the change in stability.

4 Conclusion and future research

In this paper, we analyze two new two-dimensional fluid models that incorporate customer choice, delayed queue length information, and time-varying arrival rates. The first model considers the customer choice as a multinomial logit model where the queue length information given to the customer is delayed by a constant Δ. In the second model, we consider customer choice as a multinomial logit model where the queue length information given to the customer is a moving average over an interval of Δ. In the constant arrival case for both models, it is possible to derive an explicit threshold for the critical delay where below the threshold the two queues are balanced and converge to the equilibrium. However, when the arrival rate is time varying, this problem of finding the threshold is more difficult. When the time variation is small, we show using asymptotic techniques that the new threshold changes when the arrival rate frequency is twice that of the critical delay frequency. It is important for operators of queues to determine and know these thresholds since using delayed information can have such a large impact on the dynamics of the business.

Since our analysis is the first of its kind in the queueing literature, there are many extensions that are worthy of future study. One extension that we would like to explore is the use of different customer choice functions and incorporating customer preferences in the model. With regard to customer preferences, this is non-trivial problem because the equilibrium solution is no longer a simple expression, but the solution to a transcenden-

tal equation. This presents new challenges for deriving analytical formulas that determine synchronous or asynchronous dynamics. A detailed analysis of these extensions will provide a better understanding of what types of information and how the information that operations managers provide to their customers will affect the dynamics of the system. However, we might be able to use asymptotic techniques for these extensions if we expand around a solution that we know well. Finally, we would like to generalize the arrival and service distribution to follow general distributions. An extension to general distributions would aid in showing how the non-exponential distributions affect the dynamics of the empirical process. Recent work by Ko and Pender [24,25], Pender and Ko [29] provides a Poisson process representation of phase-type distributions and Markovian arrival processes. This work might be useful in deriving delay differential equation systems for the queueing process with non-renewal arrival and service processes. We plan to explore these extensions in subsequent work.

References

1. Allon, G., Bassamboo, A.: The impact of delaying the delay announcements. Oper. Res. **59**(5), 1198–1210 (2011)
2. Allon, G., Bassamboo, A., Gurvich, I.: "We will be right with you": managing customer expectations with vague promises and cheap talk. Oper. Res. **59**(6), 1382–1394 (2011)
3. Armony, M., Maglaras, C.: On customer contact centers with a call-back option: customer decisions, routing rules, and system design. Oper. Res. **52**(2), 271–292 (2004)
4. Armony, M., Shimkin, N., Whitt, W.: The impact of delay announcements in many-server queues with abandonment. Oper. Res. **57**(1), 66–81 (2009)
5. Armony, M., Israelit, S., Mandelbaum, A., Marmor, Y.N., Tseytlin, Y., Yom-Tov, G.B., et al.: On patient flow in hospitals: a data-based queueing-science perspective. Stoch. Syst. **5**(1), 146–194 (2015)
6. Ben-Akiva, M., Bierlaire, M.: Discrete choice methods and their applications to short term travel decisions. Handb. Transp. Sci. **23**, 5–33 (1999)
7. Dong, J., Yom-Tov, E., Yom-Tov, G.B.: The impact of delay announcements on hospital network coordination and waiting times. Technical report, Working Paper (2015)
8. Eick, S.G., Massey, W.A., Whitt, W.: Mt/G/∞ queues with sinusoidal arrival rates. Manag. Sci. **39**(2), 241–252 (1993a)
9. Eick, S.G., Massey, W.A., Whitt, W.: The physics of the Mt/G/∞ queue. Oper. Res. **41**(4), 731–742 (1993b)
10. Guo, P., Zipkin, P.: Analysis and comparison of queues with different levels of delay information. Manag. Sci. **53**(6), 962–970 (2007)
11. Guo, P., Zipkin, P.: The impacts of customers'delay-risk sensitivities on a queue with balking. Probab. Eng. Inf. Sci. **23**(03), 409–432 (2009)
12. Hassin, R.: Information and uncertainty in a queuing system. Probab. Eng. Inf. Sci. **21**(03), 361–380 (2007)
13. Hui, M.K., Tse, D.K.: What to tell consumers in waits of different lengths: an integrative model of service evaluation. J. Mark. **60**, 81–90 (1996)
14. Hul, M.K., Dube, L., Chebat, J.C.: The impact of music on consumers' reactions to waiting for services. J. Retail. **73**(1), 87–104 (1997)
15. Ibrahim, R., Whitt, W.: Real-time delay estimation in call centers. In Proceedings of the 40th Conference on Winter Simulation. Winter Simulation Conference, pp. 2876–2883 (2008)
16. Ibrahim, R., Whitt, W.: Real-time delay estimation in overloaded multiserver queues with abandonments. Manag. Sci. **55**(10), 1729–1742 (2009)
17. Ibrahim, R., Whitt, W.: Real-time delay estimation based on delay history in many-server service systems with time-varying arrivals. Prod. Oper. Manag. **20**(5), 654–667 (2011a)
18. Ibrahim, R., Whitt, W.: Wait-time predictors for customer service systems with time-varying demand and capacity. Oper. Res. **59**(5), 1106–1118 (2011b)
19. Ibrahim, R., Armony, M., Bassamboo, A.: Does the past predict the future? the case of delay announcements in service systems (2015)
20. Jennings, O.B., Pender, J.: Comparisons of ticket and standard queues. Queueing Syst. **84**, 145–202 (2016)
21. Jouini, O., Dallery, Y., Akşin, Z.: Queueing models for full-flexible multi-class call centers with real-time anticipated delays. Int. J. Prod. Econ. **120**(2), 389–399 (2009)
22. Jouini, O., Aksin, Z., Dallery, Y.: Call centers with delay information: models and insights. Manuf. Serv. Oper. Manag. **13**(4), 534–548 (2011)
23. Kevorkian, J., Cole, J.D.: Perturbation Methods in Applied Mathematics, vol. 34. Springer, Berlin (2013)
24. Ko, Y.M., Pender, J.: Strong Approximations for Time-Varying Infinite-Server Queues with Non-Renewal Arrival and Service Processes. Cornell University, Ithaca (2016)
25. Ko, Y.M., Pender, J.: Diffusion limits for the (MAPt/Pht/∞) N queueing network. Oper. Res. Lett. **45**(3), 248–253 (2017)
26. Munichor, N., Rafaeli, A.: Numbers or apologies? Customer reactions to telephone waiting time fillers. J. Appl. Psychol. **92**(2), 511 (2007)
27. Ng, L., Rand, R.: Nonlinear effects on coexistence phenomenon in parametric excitation. Nonlinear Dyn. **31**(1), 73–89 (2003)
28. Pender, J.: A poisson-charlier approximation for nonstationary queues. Oper. Res. Lett. **42**(4), 293–298 (2014)
29. Pender, J., Ko, Y.M.: Approximations for the queue length distributions of time-varying many-server queues. INFORMS J. Comput **29**(4), 688–704 (2017)
30. Pender, J., Rand, R.H., Wesson, E.: Managing information in queues: the impact of giving delayed information to customers. arXiv preprint arXiv:1610.01972 (2016)
31. Pender, J., Rand, R.H., Wesson, E.: Queues with choice via delay differential equations. Int. J. Bifurc. Chaos **27**(04), 1730016 (2017)

32. Pruyn, A., Smidts, A.: Effects of waiting on the satisfaction with the service: beyond objective time measures. Int. J. Res. Mark. **15**(4), 321–334 (1998)
33. Ruelas, R.E., Rand, D.G., Rand, R.H.: Nonlinear parametric excitation of an evolutionary dynamical system. Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci. **226**(8), 1912–1920 (2012)
34. Sarel, D., Marmorstein, H.: Managing the delayed service encounter: the role of employee action and customer prior experience. J. Serv. Mark. **12**(3), 195–208 (1998)
35. Taylor, S.: Waiting for service: the relationship between delays and evaluations of service. J. Mark. pp 56–69 (1994)
36. Train, K.E.: Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge (2009)
37. Whitt, W.: Improving service by informing customers about anticipated delays. Manag. Sci. **45**(2), 192–207 (1999)