**2569**

■

# Peakedness-based staffing for call center outsourcing

Jean-Christophe Van den Schrieck, Zeynep Akşin
and Philippe Chevalier

**CORE**
Voie du Roman Pays 34, L1.03.01
B-1348 Louvain-la-Neuve, Belgium.
Tel (32 10) 47 43 04
Fax (32 10) 47 43 01
E-mail: corestat-library@uclouvain.be
http://www.uclouvain.be/en-44509.html

POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT POMS

Vol. 23, No. 3, March 2014, pp. 504–524
ISSN 1059-1478 | EISSN 1937-5956 | 14 | 2303 | 0504
DOI 10.1111/poms.12049
© 2013 Production and Operations Management Society

# Peakedness-Based Staffing for Call Center Outsourcing

Jean-Christophe Van den Schrieck
n-Side s.a., Rue Granbonpré, 11, B-1348, Louvain-la-Neuve, Belgium, jvs@n-side.com

Zeynep Akşin
College of Administrative Sciences and Economics, Koç University, Istanbul, Turkey, zaksin@ku.edu.tr

Philippe Chevalier
Center for Operations Research and Econometrics, Louvain School of Management, Université catholique de Louvain, B-1348, Louvain-la-Neuve, Belgium, philippe.chevalier@uclouvain.be

This study considers the staffing problem of a vendor call center in a co-sourcing setting. The aim is to take short-term variability and correlations in time for call arrivals at such a vendor call center into account. To do so, peakedness is proposed as a useful measure of the burstiness in the arrival stream. The study empirically demonstrates the presence of bursty arrivals at a call center and proposes an approach to the measurement of the peakedness of the arrival stream making use of standard call center data. The problematic nature of bursty arrivals in the context of call center co-sourcing is demonstrated along with an asymptotic result establishing that the problem persists in large call centers. The study then analyzes two peakedness-based staffing methods: one which is a well known extension of the square root staffing rule and another which makes use of the Hayward approximation principles. Both approaches are simple and enable the vendor to improve its staffing procedure with good accuracy.

*Key words:* bursty arrivals; peakedness measurement; call center; outsourcing
*History*: Received: December 2011; Accepted: December 2012 by Michael Pinedo, after 3 revisions.

## 1. Introduction

Understanding service demand is essential for subsequent staffing decisions in a call center. Empirical evidence on call center arrivals from the literature, Chen and Henderson (2001), Avramidis et al. (2004), Brown et al. (2005), and Steckley et al. (2005), clearly shows that (i) call arrivals are typically overdispersed relative to a Poisson distribution as shown by a larger variance than the mean in any given time interval; (ii) arrival rates are time dependent, showing similar patterns on similar days; and (iii) call arrival counts between days are correlated, having nonzero correlation both between different time periods in a day and between days.

The first and third observations motivate our analysis. A focus on variability and correlations in time leads us to an analysis of burstiness in call arrivals. Bursty arrivals imply interarrival time patterns where a run of short interarrival times can be followed by a long interarrival time (Jagerman 1997). Making use of call arrival data from a retail bank, we demonstrate that call center arrivals can exhibit such burstiness, even after controlling for time-dependent arrivals mentioned under the second observation above. Thus, our analysis provides evidence for the presence of short-term or unpredictable variability in addition to the well-known predictable variability, as well as correlations in time not just between days or time intervals but between calls. To our knowledge, this is the first documentation of burstiness of call arrivals in a call center.

While our analysis, making use of bank call arrival data, illustrates the presence of burstiness in the natural variability of call arrivals faced by a call center, certain contexts further exacerbate the burstiness of call arrivals. The call variability faced by vendors in call center outsourcing is one instance of such additional burstiness. This setting frequently implies overflowing calls from the client call center to a vendor's call center, leading to overflow traffic that is known to be bursty in nature. We focus on the staffing problem of a vendor in call center outsourcing facing such a demand stream to highlight the importance of measuring burstiness in call center arrivals and taking it into account in subsequent staffing decisions.

In call center outsourcing settings, one of the biggest challenges that the vendor faces is the unpredictable variability of arrivals to be handled. Even when the arrival rate is contractually determined, many operational decisions from the outsourcing company

can have a significant impact on the short-term variability, rendering these unpredictable. These decisions can range from call routing or call sharing mechanisms to the management of promotional activities. The effect of short-term variability can be quite dramatic on the performance of the vendor. In a very simple example we show that, depending on the variability, the performance levels change by an order of magnitude at least. In this study we propose a way to measure this short-term variability from arrivals data and develop staffing methods that take this variability into account.

We focus on outsourcing settings where calls are shared between a client firm and a vendor firm. Many firms prefer to share their calls with vendors, a practice known as co-sharing (or co-sourcing). According to a survey conducted by ICMI (Dawson 2007), call centers that use outsourcing predominantly rely on such partial sharing of their calls: out of the outsourcing contact centers surveyed, 42.9% outsource only 1–20% of customer contacts the company receives, 12.7% outsource 21–40% of the workload, 19% outsource 41–60% of total volume, 17.5% outsource 61–80% of contacts, and 7.9% report using a vendor to handle most or all (81–100%) of the customer contacts received by the company. This highlights the prevalence of co-sourcing agreements and the associated need to understand call center operations under such agreements.

In many co-sharing settings (see, e.g., Akşin et al. 2008, Gans and Zhou 2007, Keblis and Chen 2006), the client firm will serve incoming calls, frequently modeled as Poisson arrivals, in-house up to a given threshold determined by its own internal capacity and will route any calls in excess of this threshold to the service provider. This type of overflow handling by outsourcer firms is quite common. According to Dawson (2007) overflow contacts are among the most common outsourced customer contacts (25.4%). Among the key reasons for outsourcing, handling the overflow is cited by 41.3%. In such settings the vendor call center will experience an overflow type arrival process from each client or a combination of overflow arrival streams from multiple clients in the presence of pooled capacity. Burstiness of this type of call stream will be high. Capacity optimization at the vendor should take this variability into account.

It is natural to expect large call centers with uncertain demand to outsource overflow contacts, possibly to multiple smaller vendors. While partnerships between client and vendor firms may exist, a client firm operating in a high uncertainty environment dealing with several vendors having a cost minimization objective in mind will typically treat the vendors as standard subcontractors. In the latter situation, information sharing between the two parties is minimal: the client does not share details of its own operations with the vendor, such as its own demand forecast or its own service capacity. Instead the client announces an average call volume that the vendor can expect to receive, coupled with required service levels and pricing information that will be stipulated in a contract. Some co-sourcing contracts may also impose minimum call volumes that have to be outsourced to a vendor. In such a setting with minimal information sharing, the ability to measure burstiness from actual arrival data is essential for subsequent capacity optimization.

Even with a client that shares more information, vendors will find it hard to deduce the variability of the arrival stream they will receive based on this information. Possible reasons include the fact that the variability of the arrivals to the client may often already be hard to evaluate or the fact that calls may be dynamically shared with multiple vendors. In environments where promotional activities at the client firm can affect the burstiness of call volumes, a similar problem exists for the vendor, who will typically not be informed of these activities or, even when informed, will find it hard to quantify the effects of such promotions on the burstiness of call arrivals.

To alleviate this problem we propose a technique based on the peakedness to measure burstiness that is well adapted to the outsourcing setting. We then show how, based on this measure, the staffing policy can be optimized. This measure makes it also possible to improve contracting, as it is a good indicator of how "difficult" the arrival stream is to handle. The contracting will not be pursued further in this article, as it is beyond our scope.

Subsequent to a review of the literature, in section 3 we present a motivating numerical example highlighting the problematic nature of the vendor's capacity optimization. After demonstrating that arrival burstiness constitutes an important part of the problem, different metrics for measuring this burstiness are reviewed and the peakedness metric is adopted. Section 4 presents a method to measure the peakedness from real life data and explores its applicability in practice. Section 5 demonstrates that burstiness, as measured by the peakedness, persists in large-sized systems. This result is important in demonstrating that the problem is not an artifact of the particular system size analyzed in the motivating example. Section 6 proposes ways to incorporate a measure of burstiness in staffing. The proposed staffing is validated in section 7, along with an empirical documentation of burstiness in real call arrival data of a bank. The validation section also presents a comparison between peakedness based staffing and coefficient of variation based staffing, thus showing that using the coefficient of variation as a measure of variability has

shortcomings when arrivals are bursty in nature. Implication of these results for managers are discussed in section 8. We end the paper with concluding remarks in section 9.

## 2. Literature Review

The literature on call center outsourcing is relatively young. Studies that focus on contracting problems constitute an important part of this literature (Akan et al. 2011, Akşin et al. 2008, Hasija et al. 2008, Ren and Zhou 2008, Ren and Zhang 2009). The contracts studied in Akşin et al. (2008) are the only ones that allow for co-sharing of calls. Such contracts are also mentioned by Keblis and Chen (2006). The focus in Hasija et al. (2008), Ren and Zhang (2009), and Akan et al. (2011) is on contracts for complete outsourcing (i.e., all calls are outsourced to a vendor) under various forms of information asymmetry between outsourcers and service providers.

In a related stream of literature, operational problems like staffing and call routing that arise from various outsourcing settings and contracts are analyzed (Baron and Milner 2009, Gans and Zhou 2007, Keblis and Chen 2006, Milner and Olsen 2008). A vendor firm's staffing problem under different service-level agreements is analyzed in Baron and Milner (2009). The capacity optimization problem of a client company, Amazon.com, is considered by Keblis and Chen (2006), where some calls are shared with vendors in different co-sourcing agreements. The co-sourcers are used to handle overflows, and the overall optimization imposes constraints on the fraction of calls that can be allocated to them. It is the staffing problem faced by such co-sourcing vendor call centers that is studied herein.

Both Gans and Zhou (2007) and Milner and Olsen (2008) consider settings with two types of customers, a low and a high type or a contract versus non-contract type. It is shown in Milner and Olsen (2008) how certain types of service-level agreements might lead to inferior delay performance for contract type customers. This aspect of the problem is not considered in the current study, where we focus on the simplest case with a single client. In Gans and Zhou (2007), high type calls are taken in-house, while the low type customers can be outsourced to a vendor. Minimization of the variable costs for outsourcing implies that the client firm will want to take low type calls in-house whenever its capacity allows it to do so. Different call sharing and routing schemes of low type calls to the vendor are analyzed.

Gans and Zhou (2007) is the only other study in the literature that considers a co-sourcing setting as well as the operational problem of staffing at the vendor that may receive overflow type calls. Unlike the setting we consider, all information is symmetric. A queueing model with an interrupted Poisson process as arrivals is used to analyze the staffing problem. Instead, we consider staffing using approximations that rely on a measure of burstiness known as the peakedness. We discuss different approaches to measuring burstiness in arrivals in section 4.

More generally, the issue of analyzing queues in the presence of overflows or their superpositions has been addressed in a stream of literature focusing on renewal approximations (Araghi and Balcioglu 2008, Jagerman et al. 2004, Whitt 1982). While these approximations are applicable in the setting being analyzed herein, they are more complex and require knowledge about the interarrival time distribution which may not be easily available in a call center setting where call data is aggregated by 15–30-minute time intervals (see section 4).

## 3. Motivation

In this section, we start by presenting a very simple example to show how even in a very simple setting the client–vendor organization and in particular the variability of the arrival flow to the vendor can adversely affect the staffing at the vendor. We then review some approaches to measure and model this variability.

### 3.1. A Motivating Example

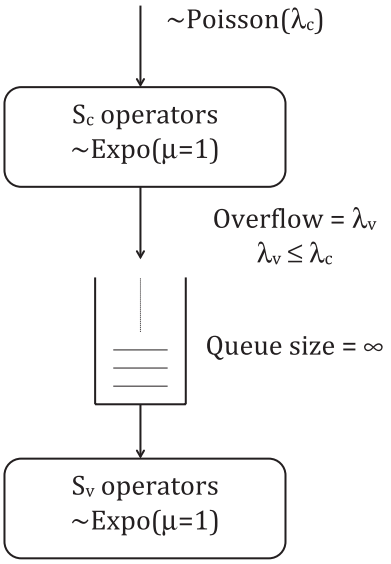
There are many ways for clients to share the load with vendors. In some circumstances the sharing mechanism might be very complex, such as described in Keblis and Chen (2006). In order to develop some understanding about the implications this might have for the vendor, we will analyze, in this section, a simple example where the sharing mechanism is based on overflow when all internal operators are busy.

Figure 1 illustrates a schematic representation of the setting being considered in this section, with a single client and a vendor with a single pool of servers. We assume that the client firm receives calls arriving according to a Poisson process with rate $\lambda_c$. Calls are served in-house as long as one of the $S_c$ servers is available and overflown to the vendor otherwise. Assuming exponential service times, the client level of this system operates as an Erlang-B system.

The vendor needs to determine a staffing level $S_v$ in response to the announced call volume $\lambda_v$ in order to fulfill the agreed service level.

In our illustrative example, we set the overflow volume to be $\lambda_v = 20$. We consider different cases where the client call center may experience different call volumes $\lambda_c = \{20, 60, 100, 140\}$. These correspond to cases where a decreasing percentage of calls are outsourced by the client, ranging from 100% to 14% as

**Figure 1 The System with One Client and One Vendor**




we go from $\lambda_c = 20$ to $\lambda_c = 140$. The varying $\lambda_c$ values can be interpreted as variations in call arrivals experienced by a single client at different points in time. Interpreted this way, we would have call volumes reaching 140 at the peak. Another way of interpreting the $\lambda_c$ values would be as different clients of the same vendor, each having a different call volume, but all promising the same call volume of 20 to the vendor. In such a multi-client setting, despite differences between the clients, the vendor could perceive them to be identical clients due to the common overflow volume $\lambda_v = 20$.

In our example, it is assumed that the client finds the corresponding staffing levels for each $\lambda_c$ value, to ensure an overflow with an average rate of 20. Service times are assumed to be exponential at both the client and vendor. To simplify the exposition, and without loss of generality, we set the service rates equal to 1. Without going into the details of the staffing procedure at the vendor, let us exogenously set $S_v = 30$ in the example. Note that given $\lambda_v = 20$, this capacity choice would ensure high service levels if arrivals did not have an overflow characteristic. Table 1 tabulates the waiting probability (WP) the expected waiting time (E[WT]) and the service level (SL) at level 0.05 experienced at the vendor as a result of this staffing, obtained via simulation. A service level at level 0.05 is defined as the proportion of calls that have to wait less than 0.05 time units before being served. We observe that, despite the fact that the arrival rate is constant at the vendor, the different service level measures show a large differential as we go from $\lambda_c = 20$ to $\lambda_c = 140$, with more than an order of magnitude of difference between the extreme cases for the waiting probability, expected waiting time or probability to wait more than 0.05 time units.

We next consider the same example, but rather than assuming a staff level for the vendor, we determine via simulation (analytically for the case of $\lambda_c = 20$) the minimum number of servers that ensure a particular service level. The staffing levels are determined respectively to achieve a constant waiting probability (WP) of 0.05, constant mean waiting time (E[WT]) of 0.05, or constant service level (SL) of 0.90. This setting can be seen as a theoretical benchmark or as representing the case where the vendor has perfect information about the client operations. The results are tabulated in Table 2. Irrespective of which service level measure is chosen as the objective, we see a big difference in the number of servers required at the vendor. This difference can be as high as 40% of the number of servers in the Poisson arrivals case.

The examples in this section illustrate the impact short-term variability can have on the operations of the service provider that has to comply with a service-level agreement. In Figure 2 we analyze the origin of this phenomenon. There are in fact two reasons for the staffing difference observed in the different cases. One effect is that the staffing at the vendor has to cope with the tail of a distribution with increasing variability. The other effect is linked to the fact that splitting the service in two pools brings some inefficiency. Indeed, once a call is overflown to the vendor, it will not be handled by the main pool even though an operator might become available before one is available at the vendor.

To estimate the relative importance of both effects, we used the Erlang C formula to compute the exact number of operators needed to obtain the target performance in the case of single pool of operators. From this number we subtract the number of operators at the client $S_c$ and compare this result to the number of

**Table 1 Constant Staffing Level for Different Overflow Proportions**

| $\lambda_c$ | 20 | 60 | 99 | 140 |
|---|---|---|---|---|
| $S_c$ | 0 | 42 | 83 | 125 |
| $\lambda_v$ | 20 | 20 | 20 | 20 |
| $S_v$ | 30 | 30 | 30 | 30 |
| WP | 0.025 | 0.157 | 0.256 | 0.327 |
| E[WT] | 0.002 | 0.034 | 0.082 | 0.137 |
| SL at 0.05 | 0.985 | 0.875 | 0.781 | 0.71 |

**Table 2 Staffing Adapted for Overflow Proportions**

| $\lambda_c$ | 20 | 60 | 99 | 140 |
|---|---|---|---|---|
| $S_c$ | 0 | 42 | 83 | 125 |
| $\lambda_v$ | 20 | 20 | 20 | 20 |
| $S_v \mid \text{WP} = 0.05$ | 29 (0.04) | 35 (0.043) | 39 (0.044) | 42 (0.046) |
| $S_v \mid \text{E[WT]} = 0.05$ | 25 (0.042) | 29 (0.048) | 32 (0.046) | 35 (0.039) |
| $S_v \mid \text{SL at } 0.05 = 0.9$ | 27 (0.932) | 31 (0.905) | 34 (0.901) | 37 (0.91) |

operators needed to handle the same load in isolation. This gives us the difference between handling a marginal load of $\lambda$ compared to an isolated load of $\lambda$. This is in effect the additional difficulty of handling the tail of an increasingly large load. The other part of the staffing difference in Table 2 can be attributed to the efficiency loss created by the overflow structure. It is striking to observe very different patterns depending on the performance measure we base the staffing on. If the target is to obtain a fixed waiting probability, most of the difference can be linked to the tail of the distribution effect. The converse is observed if the staffing is based on reaching a target expected waiting time. In this case the tail effect is even slightly negative, implying that the economies of scale dominate the variability of the tail effect. Finally, if the objective is to reach a given service level, we observe an intermediary situation where the effect of the overflow structure dominates, but still we have a significant tail effect too.

EXAMPLE. In order to explain Figure 2, we provide some details on the computations. To reach a waiting probability of 5% with a Poisson process of rate 20, we need 28.51 servers (we consider the continuous version of the Erlang-C formula). For a Poisson process with a rate of 140 the number of operators needed is 161.4. Subtracting the 125 operators of the first line pool $S_c$ (that will handle a load of 120), we find 161.4−125 = 36.4 operators are needed in order to handle the remaining load of 20 with a waiting probability of 5%. Compared with the 28.51 operators needed for a Poisson stream of rate 20 this represents an increase of 7.89. This is the value of the blue area in the left graph. We attribute this increase to the fact that a Poisson process with rate 140 presents more variability than one with rate 20.

From our simulation we know that we need 42 operators to handle the overflow load. We compare this value to the 28.7 operators needed for a Poisson process to reach the same waiting probability (given the integrality constraint the waiting probability is 4.6%). This represents now an increase of 13.3 operators; the green area is 13.3–7.89 and is attributed to the inefficiency of the two-level structure.

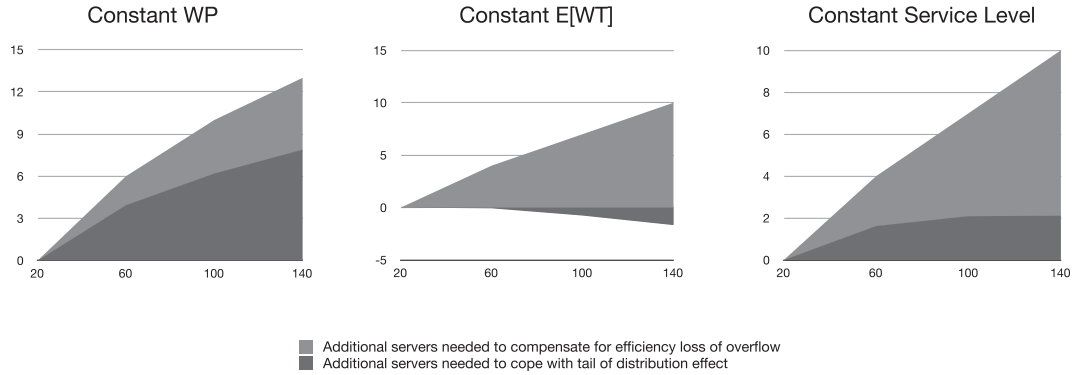
In a simple situation such as the example with overflow based call sharing presented here, one possible method to perform the staffing at the vendor is to use information from the client regarding the overall arrival rate and the local staffing level. The vendor could then either use an exact formula like the Erlang C formula or an approximate formula such as the square root staffing rule introduced by Halfin and Whitt (1981). Such an approach could account for the increasing tail effect, however, inferring the inefficiency introduced by the two-level structure would be difficult. If the setting at the client is more complex than a single first-level pool, the staffing for the vendor becomes even more difficult, particularly if the variability of the flow arriving to the vendor is not measured directly.

As mentioned earlier, it is quite frequent that a call center faces high variability in arrivals (more variable than a Poisson process; Avramidis et al. 2004, Chen and Henderson 2001, Steckley et al. 2005). The variability of arrivals has often been attributed to nonconstant arrival rates. In our example we saw that even with a constant arrival rate, the variability of the arrival process can have a strong influence on service quality as experienced by the customer.

### 3.2. Modeling Variability

**3.2.1. Squared Coefficient of Variability.** The squared coefficient of variability (SCV) is widely used for stochastic models in manufacturing environments. From a theoretical point of view this measure assumes that interarrival times are IID, which is not the case for an overflow process. While easy to

**Figure 2 Additional Staffing Needed in an Overflow System to Keep the Waiting Probability, the Expected Waiting Time, or the Service Level Constant Compared to Poisson Arrivals for Different Arrival Rates to the Client if the Overflow Rate Is Kept Constant at $\lambda = 20$**



compute, models that use the SCV in order to compute the performance of a queueing system are based on the IID assumption and cannot exploit the burstiness of the arrivals. Since this is the standard approach to measuring variability in many settings and is easily understood by practitioners, we use it as a benchmark for our staffing results in section 7.

**3.2.2. Peakedness.** The peakedness is a measure of variability that does not make any assumption on the interarrival times. This measure was first derived in the telecommunications literature to study overflows in telecommunication networks. The peakedness is in fact an indirect measure of variability. Variability is measured based on the number of servers that would be busy if the arrival stream was fed to an infinite server pool. The peakedness is defined as the ratio of the variance divided by the mean of the number of busy servers from this infinite pool. See, for example, Wolff (1989) for an introduction to peakedness.

Models for call centers that use the peakedness to measure variability are presented in Chevalier and Tabordon (2003), Chevalier and Van den Schrieck (2008), and Pot et al. (2008). Although it might seem that the peakedness would be hard to measure, in practice it is not much harder than computing the SCV. We elaborate on the measurement of peakedness in section 4.1.

**3.2.3. Interrupted Poisson Process.** Another approach used to model arrival processes that are not IID is to use an interrupted Poisson process (IPP). Kuczura (1973) presents a method to fit the parameters of an IPP based on the first three moments of the interarrival times. The computations for this method are much harder than for the peakedness, and the numerical tests of Tabordon (2002) show that the accuracy of this method for call center models is in general not as good as that of models using the peakedness.

**3.2.4. Doubly Stochastic Poisson Process.** One way of modeling a process that is more variable than a Poisson process is to imagine that the arrival process is actually a Poisson process with an arrival rate that varies according to some random process. This type of process can model a large variety of situations. A typical interpretation is to view this randomness as a consequence of forecasting difficulties (Steckley et al. 2005). Another interpretation for such processes is to model short-term variability, which is the focus of this study. Even though doubly stochastic Poisson processes allow for sophisticated models, in practice one would need to make many assumptions on the random process that generates the arrival rates. Furthermore, the statistical estimation of the parameters of this process will be difficult.

**Table 3 A Comparison of Different Approaches to Model Variability for Arrival Flows**

| Approach | Advantages | Disadvantages |
|---|---|---|
| SCV | Easy to measure | Ignores time correlation |
| Peakedness | Model used successfully for call centers | Less intuitive measure than SCV |
| IPP | Burstiness explicitly modeled | Complex computations |
| Doubly stochastic | Most general model | Parameter estimation very hard in practice |

*Note:* SCV, squared coefficient of Variability; IPP interrupted Poisson process

Table 3 presents a short summary of the advantages and disadvantages of the different approaches. The peakedness is the measure of burstiness that we use in the coming sections. Although less intuitive than the square coefficient of variation, it is a richer measure and has interesting features that make it much easier to use in practice than other approaches.

## 4. Measuring the Burstiness

In this section, we focus on peakedness as a measure of variability and present an approach to measuring the peakedness of an arrival process directly from call arrival data.

### 4.1. A Method to Measure the Peakedness from Data

There exist analytical formulae to determine the peakedness of an overflow when the characteristics of the first level arrivals and server pool are known (see, e.g., Wolff 1989). Our objective is to be able to estimate this peakedness from call data without having access to operational details of the first level. Consequently, we present a way to measure the peakedness directly from the arrival stream data.

Recall that the peakedness $z$ is the ratio of the variance to the mean of the number of busy servers if the arrivals were handled by infinitely many servers. One immediate way of measuring the peakedness is then to simulate an infinite pool of servers, using the historical arrival times and simulated random service times. Following Fredericks (1980), the infinite pool of servers is assumed to have a service time that is exponentially distributed with the same service rate as that of the vendor call center. Thus, throughout this section $\mu$ corresponds to the service rate at the vendor. The corresponding peakedness can be denoted by $z(\mu)$.

EXAMPLE. We illustrate this approach with an example. Computational details of this example can be found in Appendix B. Arrivals are assumed to be

Poisson with rate 10/hour. Service times are exponentially distributed with parameter 2/hour (Tables A2 and A3 in the Appendix show the arrival and service times used). Setting the initial number of busy servers as $S(0) = \lambda/\mu = 5$, $S(t)$, the number of busy servers at time $t$ for the first 15 time units of the time horizon is shown in Figure 3. We find that $E[S] = 4.81$ and $\text{Var}[S] = 5.04$. $z$ is thus equal to 1.05, not far from the theoretical value of $z = 1$ for a Poisson process.

The major drawback of the simulation method is that the randomness of the service times will add some noise to the estimator of the peakedness. In order to circumvent this difficulty, we will use the approach developed by Macq (2005) based on a fluid version of the peakedness. In this approach, an arrival is virtually divided into infinitesimal requests with processing times that follow the service time distribution. This means that if $G(x)$ is the cumulative service time distribution, an arrival at time $\tau$ will "use" $1 - G(t - \tau)$ servers at time $t$ (for $t \geq \tau$). Thus the number of busy servers does not include any variability that is due to the service process. The fluid peakedness $\bar{z}$ is defined as the ratio of the variance to the mean of the number of busy servers when service is "fluid." There is a direct relationship between the fluid peakedness $\bar{z}$ and the peakedness $z$. This relationship is given by

$$\bar{z} = z - 1 + \mu\rho_{G^c}(0),$$

where $\rho_{G^c}(t) = \int_{\max(0,-t)}^{+\infty} G^c(\tau)G^c(t+\tau)d\tau$ is the autocorrelation function of the complementary service time distribution $G^c(t) = 1 - G(t)$. When the service time distribution is exponential this relationship simplifies to

$$\bar{z} = z - \frac{1}{2}. \tag{1}$$

EXAMPLE CONTINUED. We continue our illustrative example and estimate the fluid peakedness for the sample of arrivals in Table A2. $\bar{S}(t)$, the fluid number of busy servers at time $t$, suffers an exponential decay with parameter $\mu = 2$. Figure 4 shows the evolution of $\bar{S}(t)$ over the first 15 time units of the time

**Figure 3** ***S*(*t*) as Observed with the 400 Arrivals and Service Times of Tables A2 and A3 over the First 15 Time Units of the Time Horizon**

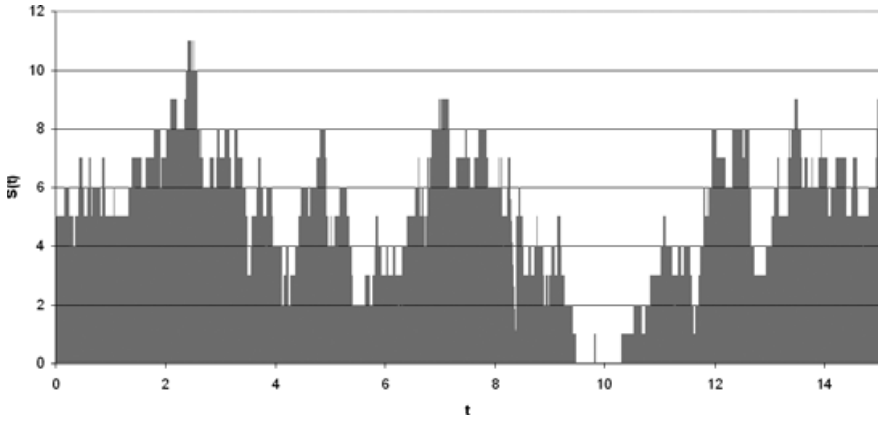
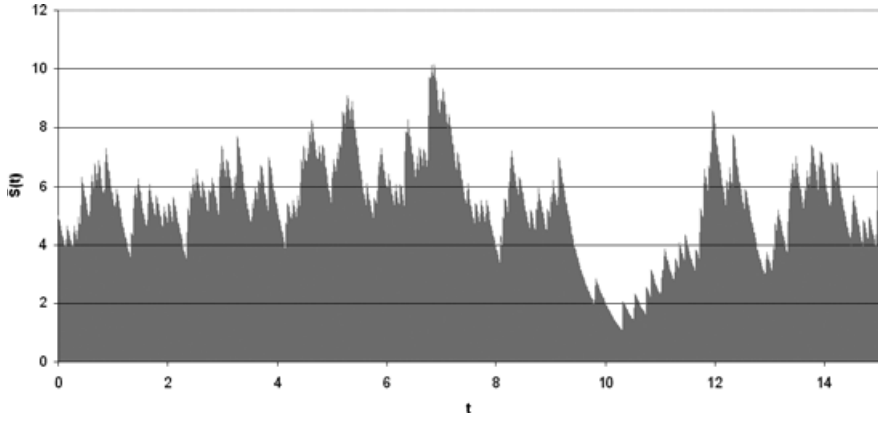



**Figure 4** $\bar{S}(t)$ **as Observed with the 400 Arrivals of Table A2 for the First 15 Hours of the Considered Time Horizon**



horizon. We randomly choose 50 time points on the time horizon to compute $\bar{S}(t)$. To limit the bias at the beginning of the time horizon we set $\bar{S}(0) = \lambda/\mu = 5$.

In practice call centers keep reports of past arrivals as aggregated demand over time intervals. For example, reporting software may collect the number of calls over 15-minute time intervals. Although some information is lost in the process, it is possible to compute an accurate estimate of the peakedness based on such data. Such a peakedness approximation is presented by Chevalier et al. (2010) in the context of batch arrivals in a supply chain. We adapt their approach to call center data herein.

With the aggregated data, we observe arrivals at specific epochs, every 15 or 30 minutes, depending on the time interval of the aggregation. In the arrival process with aggregated data, $B_n$ is the set of arrivals in the interval $[(n-1)T,nT)$, and $|B_n|$ is the number of arrivals in this interval. $\bar{S}_n = \bar{S}_{n-1}e^{-\mu T} + |B_n|$ denotes the number of busy fluid servers at time $nT$. Proposition 1 shows that the fluid peakedness $\bar{z}$ of the full arrival process can be estimated from the aggregated data.

PROPOSITION 1. *Under some mild assumptions (stated in the Appendix), the fluid peakedness of the unaggregated arrival process can be estimated from the aggregated data using the following relationship:*

$$\bar{z} \simeq \frac{1}{2} - \frac{1}{\mu T} \cdot \frac{1-e^{-\mu T}}{1+e^{-\mu T}} + \frac{\mathrm{Var}[\mathrm{S_n}]}{E[\bar{S}_n]} \left( \frac{1-e^{-\mu T}}{\mu T} \right). \quad (2)$$

The details of the derivation can be found in the Appendix. In deriving this result we make the assumption that the arrival process is in essence independent from the aggregation periods.

EXAMPLE CONTINUED. Going back to the illustrative example, we aggregate the Poisson arrival data in time intervals of $T = 0.25$, corresponding to 15-minute time intervals (displayed in Table A5). For each interval $n$ we compute $\bar{S}_n$. To decrease the bias brought by warm-up effects, we assume $\bar{S}_0 = \frac{\lambda T/\mu}{1-e^{-\mu T}} = 6.35$. Figure 5 depicts the fluid number of busy servers with batch arrivals.

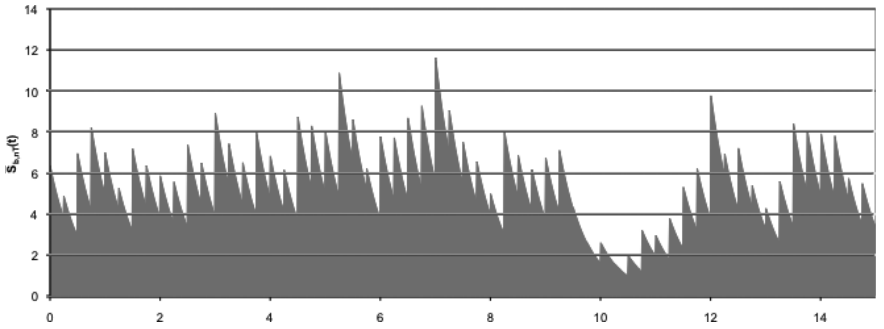
Using Equation (1) with the computed values of $E[\bar{S}_n] = 6.13$ and $\mathrm{Var}[\bar{S}_n] = 3.74$, we find $\bar{z} = 0.49$ and $z = 0.99$.

### 4.2. Applicability of the Proposed Method

The example of this section assumes Poisson arrivals with a steady rate. Our measurement approach assumes wide-sense stationarity of call arrivals. Thus, the result of Proposition 1 would not be applied to a time interval where call arrivals are non-stationary. In practice, however, call arrival rates are time dependent. This implies that we envision the use of the measurement approach to distinct time intervals where stationarity is assumed. This raises the question of the number of intervals needed to have a reliable estimate. The derivation of statistical confidence intervals is beyond the scope of this article, but experience suggests that 100–200 intervals give reliable estimates. In the example used in the previous subsection, the estimate was based on approximately 160 intervals. If we imagine in practice to have arrivals aggregated on 15-minute intervals, this implies having 40 hours of historical data. Note that this must be data over a period where the arrivals can be considered stationary. If, for example, we observe that during the afternoon from 2 to 6 pm no systematic pattern exists in the arrival process and consequently the arrival process can be considered stationary in a wide sense, then 10 days of observations will be needed to gather 160 samples. Obviously, if the arrival process can only be considered stationary over a 2-hour period, then 20 days of observations will be needed to gather 160 samples of 15 minutes aggregated arrivals. This seems quite compatible with a practical application.

Another assumption made in the above example is that service times are exponentially distributed. Empirical evidence exists showing that call center

**Figure 5 The Fluid Number of Busy Servers as Observed with the 400 Arrivals of Table A2 for the First 15 Time Units of the Considered Time Horizon**



service times may have a lognormal distribution (Brown et al. 2005). Simulations of the peakedness when service times are assumed to be lognormal instead of exponential, presented in Van den Schrieck (2010), show that peakedness values are not sensitive to the service time distribution; however, the value of the coefficient of variation of service times ($CV_s$) does matter. In particular, when $CV_s$ is close to 1, the difference in the peakedness induced by assuming exponential service times versus lognormal service times is minimal. In one set of examples in Van den Schrieck (2010) we see that systems where the infinite server pool service times are assumed to be lognormal with $CV_s = 1$, the relative error vis-a-vis an exponential service time benchmark case is around 4%. The empirical evidence presented in Brown et al. (2005) is encouraging in this respect, as it shows that service times in practice can have $CV_s$ values that are close to 1. Using the information presented in table 1 of that article, we find the coefficient of variation values of service times for different types of calls as follows: regular service calls $CV_s = 1.05$, potential customer calls $CV_s = 1.27$, Internet calls $CV_s = 1.18$, and stock trading calls $CV_s = 1.12$. Sensitivity to the assumed service rate $\mu$ is explored in a set of simulations not reported herein. We see that the peakedness value is not too sensitive to the service rate value, showing an error of around 1% when the service rate value is off by around 10%.

The following section proposes two staffing methods that take the measured peakedness into account. A natural question that may come to mind at this point is why one does not determine the appropriate staffing directly by using simulation, making use of the arrivals data that is being used in the measurement of peakedness. In a simulation, the effect of variability introduced by arrivals would be hard to disentangle from other types of variability in the system. The proposed fluid peakedness measurement approach allows eliminating the service time variability, making arrival variability harder to dispute. This in turn would be useful in a business contracting setting. The fluid approach, which leads to a formula for the peakedness (Proposition 1), will make conveying information on the peakedness much easier, and its effect on staffing is made more transparent by a staffing rule that depends on the value of this measured peakedness. Compared to a staffing result that is generated by a simulation, this added transparency has value in practice. An additional challenge of the simulation approach comes from the fact that arrival data may not be readily available at the individual call level, as call center software typically collects these by aggregating data over 15- or 30-minute intervals. Proposition 1 helps in dealing with such a setting where data are aggregated.

# 5. Peakedness in Large-Size Call Centers

One relevant question for practitioners is whether the effect of peakedness illustrated by our example persists in large systems. In particular, it is well known in call center practice that the effects of supply side variability diminishes with the size of a call center. In this section, we illustrate that peakedness is different and will persist in large systems.

As before, we choose to analyze the effect of size in a simple setting in order to gain intuition about the effect of peakedness for large size call centers. Kosten (1973) shows that the peakedness of the overflow process of a loss system with Poisson arrivals with rate $\lambda$ that are serviced by a pool with $s$ operators with exponential service times is given by

$$z = 1 - \rho B_E(s,\rho) + \frac{\rho}{1 + \lambda B_E(s,\rho) - \rho + s}. \qquad (3)$$

Consider a call center where the ratio of the incoming load over the number of operators is equal to $\alpha$, i.e., $\rho = \alpha s$, the overflow of this call center is sent to the external service provider. We suppose that $\alpha > 1$ as the overflow rate would quickly be negligible for large size systems if $\alpha < 1$.

The following lemma provides an asymptotic expansion for $B_E(s,\alpha s)^{-1}$ when $s \to \infty$.

LEMMA 1.

$$B_E(s,\alpha s)^{-1} \sim \sum_{i=0}^{\infty} \frac{\alpha}{\alpha - 1} s^{-i} \frac{\delta^i}{\delta \alpha^i}\left(\frac{\alpha}{\alpha - 1}\right), \qquad (4)$$

*$s \to \infty$, $s > 0$, $\alpha > 1$, and $\frac{\delta^i}{\delta \alpha^i} f(\alpha)$ is the $i^{th}$ derivative of $f(\alpha)$.*

PROOF. See Jagerman (1974), Theorem 13. □

We can exploit this result to find $z$ when $s$ tends to $\infty$.

PROPOSITION 2. *Suppose $\alpha > 1$. Then,*

$$\begin{aligned} \lim_{s\to\infty} z(s,\alpha s) &= \lim_{s\to\infty} \left(1 - \alpha s B_E(s,\alpha s) \right. \\ &\qquad \left. + \frac{\alpha s}{1 + s - \alpha s + \alpha s B_E(s,\alpha s)}\right) \\ &= \frac{\alpha}{\alpha - 1}. \end{aligned} \qquad (5)$$

PROOF. Using the result of Lemma 1 in Equation (3) yields

$$\begin{aligned}\lim_{s\to\infty} z(s,\alpha s) &= 1+\lim_{s\to\infty}\frac{-\alpha s}{\frac{\alpha}{\alpha-1}\left[1-\frac{s^{-1}}{(\alpha-1)^2}+\frac{(2\alpha+1)s^{-1}}{(\alpha-1)^4}+o(s^{-3})\right]}+\lim_{s\to\infty}\frac{\frac{\alpha^2 s}{\alpha-1}\left[1-\frac{s^{-1}}{(\alpha-1)^2}+\frac{(2\alpha+1)s^{-1}}{(\alpha-1)^4}+o(s^{-3})\right]}{\frac{\alpha}{\alpha-1}\left[1-\frac{s^{-1}}{(\alpha-1)^2}+\frac{(2\alpha+1)s^{-1}}{(\alpha-1)^4}+o(s^{-3})\right](1+s-\alpha s)+\alpha s}\\ &= 1+\lim_{s\to\infty}\frac{-s(\alpha-1)}{1-\frac{s^{-1}}{(\alpha-1)^2}+o(s^{-2})}+\lim_{s\to\infty}\frac{\alpha s-\frac{\alpha}{(\alpha-1)^2}+\frac{2\alpha^2+\alpha}{(\alpha-1)^4}+o(s^{-2})}{\frac{\alpha}{\alpha-1}+\frac{3\alpha s^{-1}}{(\alpha-1)^3}+o(s^{-2})}\\ &= 1+\lim_{s\to\infty}\frac{-\alpha s+\frac{3\alpha}{(\alpha-1)^2}+\alpha s-\frac{\alpha}{(\alpha-1)^2}-\frac{\alpha}{(\alpha-1)^2}+o(s^{-1})}{\frac{\alpha}{\alpha-1}+o(s^{-1})}\\ &= 1+\frac{1}{\alpha-1}=\frac{\alpha}{\alpha-1}.\end{aligned}$$

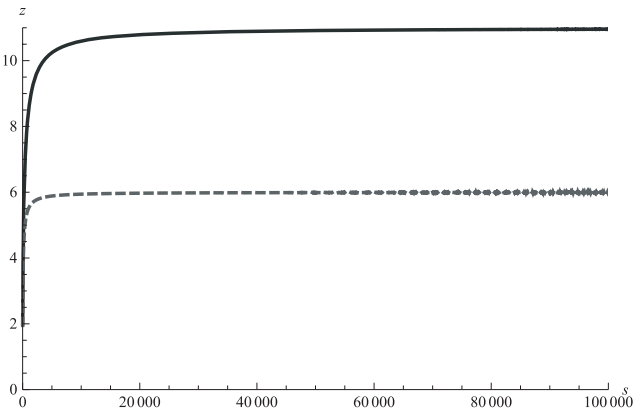
□

Figure 6 shows the evolution of $z$ when $s$ increases for two values of $\alpha$. It is noticeable that the peakedness of the overflow will always stay higher than 1; this means that the overflow will always exhibit more variability than a Poisson process. Equation (5) also confirms the observation of our initial example that the smaller the fraction of calls that is overflown the larger its peakedness will tend to be. Observe also that the asymptotic peakedness increases when $\alpha$ tends to 1. This means that the impact of the peakedness will be maximal in systems that operate close to a full capacity regime.

# 6. Staffing by Measuring the Peakedness

We now present two staffing methods that incorporate the peakedness. The objective is to derive staffing rules that can be based on the variability of the arrival flow of calls as measured with the peakedness. We start with an extension of the square-root staffing formula adapted to take the peakedness of the arrival stream into account. We next present the Hayward Approximation (see Wolff 1989, pp. 354–355). This approximation widely used in the telecommunication industry has been extensively studied for loss models (i.e., zero queue length servers); we show here that this approximation can be readily extended for queueing models with quite satisfactory results.

**Figure 6 The Evolution of the Peakedness $z$ when $s$ Grows; the Blue Continuous Line Corresponds to $\alpha = 1.1$, the Red Dotted Line Corresponds to $\alpha = 1.2$**




## 6.1. A Peakedness Adjusted Square-Root Staffing Formula

A well-known rule of thumb for staffing a call center with a single homogeneous pool of agents is known as the square root safety staffing rule. This rule simply states that the buffer capacity in an Erlang-C system should be proportional to the square root of the offered load. The proportionality factor depends on the desired service level. First evoked by Erlang (1948) and Newell (1973), the rule was explored in the study by Whitt (1992) for an Erlang-C system and shown to perform well for an Erlang-A system in Garnett et al. (2002). We refer to Borst et al. (2004) for a detailed history of the formula. Other related references can be found in the survey articles Gans et al. (2003) and Akşin et al. (2007).

Given an offered load $\lambda/\mu$ and a desired service level $\beta$, the square root staffing rule states that the number of servers $S$ should be set as

$$S=\frac{\lambda}{\mu}+\beta\sqrt{\frac{\lambda}{\mu}}.$$

This rule is based on exponential inter-arrival times and service times. Whitt (1992) draws upon the relationship with an infinite server model to derive a square root staffing rule in the case of general inter-arrival times. This leads to the following adjustment of the original square root staffing rule for a setting with arrivals having peakedness $z$:

$$S=\frac{\lambda}{\mu}+\beta\sqrt{z}\sqrt{\frac{\lambda}{\mu}}. \tag{6}$$

It is interesting to note that this expression corresponds to dividing the arrival rate and the number of

servers by $z$, in a sense *de-scaling* the system by the peakedness. This is the same idea as Hayward's approximation presented in the next section.

There is no direct way to determine the value of $\beta$. We propose here to fix it by starting from a system where we keep the original arrival rate but assume exponential inter-arrival times (corresponding to a peakedness of 1). We then use the M/M/s formulae to determine the exact number of operators needed $S_{MMs}$. From this number we can determine $\beta$ through

$$\beta = \left(S_{MMs} - \frac{\lambda}{\mu}\right) / \sqrt{\frac{\lambda}{\mu.}} \tag{7}$$

EXAMPLE. Suppose we want to determine the value of $\beta$ to obtain an average waiting time of 0.05. From the M/M/s model we establish that 24.69 operators are needed for a load of 20. From Equation (7) we determine a value of $\beta = 1.048$.

### 6.2. The Hayward Approximation

The Hayward approximation (see Fredericks 1980, Wilkinson 1956) permits us to approximate the performance of a loss system when the arrival rate is not Poisson. The idea is that if the incoming flow of rate $\lambda$ is not a Poisson process but a general point process with a peakedness $z$, it will behave in a pool of $s$ operators similarly to a Poisson arrival flow of rate $\lambda/z$ in a pool with $s/z$ operators. The blocking (or loss) probability can be approximated by

$$B(\rho, s, z) \approx B_E(\rho/z, s/z), \tag{8}$$

where $B(\rho,s,z)$ denotes the probability of blocking and $B_E(.,.)$ is the Erlang-B formula,

$$B_E(\rho, s) = \frac{\frac{\rho^s}{s!}}{\sum_{i=0}^{s} \frac{\rho^i}{i!}}, \tag{9}$$

and $\rho = \frac{\lambda}{\mu}$ is the offered load of the system.

Although the idea of Hayward's *de-scaling* with the peakedness can be extended in a straight forward way to the Erlang-C formula for a queueing system, the quality of the resulting approximation has, to our knowledge, never been investigated in the scientific literature. For this reason we start by presenting our extension of Hayward's formulae and perform some simulation experiments to validate it.

We propose thus to use the following approximation for the Waiting Probability (*WP*), which is the same as that proposed in Whitt (1992):

$$WP \approx C_E(\rho/z, s/z), \tag{10}$$

where $C_E(\rho, s)$ is the standard Erlang C formula. In order to compute the Erlang C for noninteger values of the number of servers, it is convenient to use the algebraic relationship between the Erlang B and Erlang C formulae:

$$C_E(\rho, s) = \frac{sB_E(\rho, s)}{s - \rho(1 - B_E(\rho, s))}. \tag{11}$$

To estimate the waiting time distribution we use the fact that for a queueing system with a general arrival process (including overflow arrival processes) and an exponential service time distribution, the waiting time distribution of a call, provided that the call has to wait, is exponential (Kleinrock 1975). The expectation of the exponential waiting time distribution depends on the distribution of the interarrival times. No closed form expression exists for overflow type processes. We propose to once again continue the generalization of the Hayward approximation, dividing the load and staff levels by the peakedness of the flow. This yields the following approximations for the Waiting Time (*WT*):

$$P(WT \le t | WT > 0) \approx 1 - e^{-\frac{s\mu - \lambda}{z}t}. \tag{12}$$

$$P(WT \le t) \approx 1 - C_E(\rho/z, s/z)e^{-\frac{s\mu - \lambda}{z}t}. \tag{13}$$

$$E[WT] \approx \frac{zC_E(\rho/z, s/z)}{s\mu - \lambda}. \tag{14}$$

In order to validate formulae (11)–(14) we carry out a set of numerical examples. We also compare, where appropriate, the corresponding approximations in Whitt (1992) as a further benchmark. The set consists of 144 configurations where we vary the number of operators and the characteristics of the arrival flow in terms of rate and variability (see Table A4 in the Appendix). The arrival processes are generated as overflows of a loss system. All service times are exponentially distributed with a mean value of 1. The range of arrival rates goes from 3.2 to 75. The range of peakedness values for the arrival flow varies from 1.13 (close to a Poisson process having peakedness 1) to 5.10.

In Table 4 we present the mean absolute error, the mean error, and the minimum and maximum errors observed when comparing the simulated results with the results obtained with formula 11 on the 144

**Table 4 Percent Errors Observed with Formula (11) in Comparison to Simulation Results for the Data Set Presented in the Appendix**

| | |
|---|---|
| Mean absolute error | 3.46 |
| Mean error | 3.46 |
| Minimum error | 0.01 |
| Maximum error | 12.24 |

**Table 5 Percent Errors Observed with Formula (12) (in Parentheses the Errors Observed with Formulas (19) and (20) in Whitt 1992) in Comparison to Simulation Results for the Data Set Presented in the Appendix**

| $SL\|WT>0$ | 0.05 | 0.1 | 0.25 | 0.5 | 1 | $E[WT\|WT>0]$ |
|---|---|---|---|---|---|---|
| Mean abs error | 0.37 (2.36) | 0.59 (4.04) | 0.88 (7.06) | 0.97 (9.24) | 0.90 (10.04) | 5.07 (58.40)) |
| Mean error | 0.31 (2.36) | 0.46 (4.03) | 0.56 (7.05) | 0.41 (9.22) | −0.02 (10.01) | 0.83 (58.40) |
| Min error | −0.17 (0.28) | −0.33 (0.54) | −0.72 (1.22) | −1.17 (2.12) | −1.71 (3.04) | −5.32 (9.39) |
| Max error | 2.81 (10.09) | 3.35 (15.00) | 4.89 (24.88) | 3.94 (31.16) | 2.85 (36.70) | 62.86 (186.73) |

**Table 6 Percent Errors Observed with Formula (13) in Comparison to Simulations Results for the Data Set Presented in the Appendix (Corrsponding Results for Whitt 1992 through Unconditioning)**

| SL | 0.05 | 0.1 | 0.25 | 0.5 | 1 | E[WT] |
|---|---|---|---|---|---|---|
| Mean abs error | 3.22 (2.72) | 3.03 (2.30) | 2.60 (1.92) | 2.12 (1.94) | 1.61 (3.20) | 9.55 (43.45) |
| Mean error | −3.22 (−2.56) | −3.01 (−1.81) | −2.56 (−0.13) | −2.07 (1.58) | −1.57 (3.14) | 7.66 (−43.34) |
| Min error | −11.6 (−10.45) | −11.03 (−8,81) | −10.20 (−5.28) | −9.10 (−2.83) | −7.95 (0.00) | −61.88 (−150.34) |
| Max error | 0.15 (0.97) | 0.25 (2.24) | 0.32 (5.76) | 0.40 (10.04) | 0.64 (13.94) | 37.69 (8.18) |

configurations. The mean absolute error is the difference between the simulation and approximation results in absolute value. The mean error is computed by subtracting the estimated result from the simulated one. In the table, we report percent errors.

We observe that the mean absolute error lies between 3% and 4%. Also, we see that the mean error is identical to the mean absolute error. This indicates that the approximation underestimates the actual waiting probability. A similar behavior was observed in loss models by Chevalier and Tabordon (2003).

We now use the same data set to verify the accuracy of formula (12). This is testing whether the parameter $(s\mu-\lambda)/z$ is a good proxy for the real parameter of the exponential distribution. We compute the proportion of the waiting calls that wait less than 5%, 10%, 25%, 50%, and 100% of the service time and compare the results to what is observed in simulation. In addition to the results on the service level, we also compute the average waiting time of the waiting calls, $E[WT|WT>0]$. According to our method, it can be estimated by $E[WT|WT>0] \approx \frac{z}{s\mu-\lambda}$. As this performance measure is not a probability, we present the relative error, that is, the difference in absolute value between the simulation result and the approximation result divided by the simulation result. We also compare formulae (19) and (20) in Whitt (1992) for the conditional waiting time distribution and conditional mean waiting time, respectively, with the simulation results. Both sets of comparisons are presented in Table 5, where the numbers in parentheses indicate the errors under the approximations in Whitt (1992).

We obtain excellent results: the mean absolute error is < 1% for the five service levels considered. As the mean error shows, there is no significant bias. The maximum error is quite low as well. The conditional waiting time is accurately estimated, with on average a 5% error and no significant bias in the estimation. The maximum error observed is close to 63%. This is large but corresponds to an extreme case where the load at the vendor is about 99.9%. There are in fact three cases with extremely high loads for the vendor where absolute errors of about 30% are observed. Aside from those cases where the system virtually reaches saturation, the approximation performs well. The comparison with the Whitt (1992) benchmark is also favorable throughout.

In Table 6 we present the results for the validation of formula (13). The average absolute error is at 2% or 3%. Here again we observe a tendency to overestimate the service level. The estimation for the average waiting time is not as accurate as the other estimations but remains of good quality, with an average relative error of < 10%, with a tendency to underestimate the actual waiting time. The numbers in parentheses are the corresponding values from Whitt (1992) where we have used Equation (11) to uncondition the results obtained with formulae (19) and (20) of Whitt (1992). The results are comparable, though the Whitt (1992) results are slightly better for lower values of SL while Formula 13 leads to a better estimate for higher values of SL and for $E[WT]$. As a conclusion we believe that formulae (11)–(14) yield an excellent trade-off between simplicity of utilization and accuracy.

## 7. Experiments and Validation

In this section, we compare both methods introduced in the previous sections and show that both enable us to improve the staffing decisions in the presence of peakedness. We also wish to address the question of whether just a measurement of the SCV of arrivals and staffing based on the SCV might not be sufficient. To this end, we propose a SCV-based staffing

approach making use of earlier results by Whitt (1993). We first analyze the different staffing methods for the examples presented in section 1. This will show us that in a perfectly controlled environment the proposed methods work as expected. In a second subsection we use real-life data from a large bank call center to show the usefulness of our proposed approach in a real environment.

### 7.1. Controlled Experiment

For queues with general arrival streams, it is difficult to do an exact staffing analysis with performance targets like the SL measure we consider herein without an explicit understanding of the distribution of inter-arrival times. In order to perform staffing that only makes use of the SCV of arrivals as measured from inter-arrival time data, we resort to approximations developed in Whitt (1993). In particular, Equation (2.24) in Whitt (1993) is used to approximate $E[WT]$, Equation (3.4) in Whitt (1993) is used to approximate the $WP$, and Equation (4.6) in Whitt (1993) is used to approximate the $SL$.

We consider staffing for the examples presented in section 1. For each scenario we determine the staffing obtained using the square root rule with peakedness (z-Square Root), the Hayward approximation applied to queueing systems method (Ha-Queueing), as well as the approximations of Whitt (1993) making use of the SCV instead of the peakedness (SCV-based). Because the number of operators must be integer, the simulation results are not exactly equal to the desired performance level: the staffing has been chosen in order to ensure the minimum number of operators that outperform the desired performance level. Therefore, we do not set the desired performance level, but the effective performance reached in the simulations as the objective.

Table 7 summarizes the 12 scenarios. It is divided into three parts, one for each of the performance measures. In each of the three parts, the first line repeats the simulation results presented in Table 2 and gives the observed performance that serves as a target for the two methods. The second line presents the staffing obtained with the z-Square Root method, the third line gives the staffing from the Ha-Queueing method, and the last line gives the staffing from the SCV-based staffing. We use here the formulae presented in the previous section where $\lambda$ is replaced by $\lambda_v$. We observe that the introduction of the peakedness improves the estimation of the staffing. The SCV-based approach that measures variability through the SCV instead of the peakedness clearly demonstrates the need to take into account the dependence between arrivals present in an overflow type arrival stream when staffing, especially when having a $SL$ target in mind.

**Table 7 Comparison of Peakedness Methods with Full Information**

| | | | | |
|---|---|---|---|---|
| $\lambda_c$ | 20 | 60.2 | 99.6 | 140.07 |
| $S_c$ | 0 | 42 | 83 | 125 |
| $\lambda_v$ | 20 | 20 | 20 | 20 |
| $z_v$ | 1 | 2.49 | 3.63 | 4.62 |
| Simulated, $S_v \mid WP = 0.05$ | 29 (0.04) | 35 (0.043) | 39 (0.044) | 42 (0.046) |
| z-Square Root, $S_v$ | 29 | 33.54 | 36.27 | 38.05 |
| Ha-Queueing, $S_v$ | 29 | 34.58 | 37.94 | 40.35 |
| SCV-based, $S_v$ | 29 | 35 | 37 | 37 |
| Simulated $S_v \mid E[WT] = 0.05$ | 25 (0.042) | 29 (0.049) | 32 (0.046) | 35 (0.039) |
| z-Square Root, $S_v$ | 25 | 27.44 | 29.16 | 31.01 |
| Ha-Queueing, $S_v$ | 25 | 28.96 | 31.88 | 34.8 |
| SCV-based, $S_v$ | 29 | 27 | 28 | 28 |
| Simulated $S_v \mid SLat0.05 = 0.9$ | 27 (0.93) | 31 (0.91) | 34 (0.9) | 37 (0.91) |
| z-Square Root, $S_v$ | 27 | 29.86 | 31.74 | 33.69 |
| Ha-Queueing, $S_v$ | 27 | 30.67 | 33.18 | 35.7 |
| SCV-based, $S_v$ | 29 | 28 | 28 | 29 |

SCV, Squared coefficient of variability; WP, Waiting probability; WT, Waiting time.

### 7.2. Real-Life Experiment

This experiment is based on real-life data collected from the call center of a major bank. The data are aggregated data over 15-minute periods. The call center operates 24/7 with huge variations in demand over the week or year. We focus our analysis on 1 week and more particularly on the period of 1–4 pm where the traffic, although variable, shows no particular pattern and would warrant a constant staffing level (see Figure 7). This bank offers several different services through its call center. Our analysis is performed for a case where each service has a dedicated pool of operators.

We test our method for three of the services offered by the bank. For each service we compute the mean arrival rate, the estimated peakedness of the arrival rate, and the mean service time (based on the aggregate records by 15-minute interval). We also calculate the SCV of inter-arrival times. As is typical for call centers, the data do not contain inter-arrival times. We approximate the SCV of inter-arrival times using the 15-minute arrival count data. In particular, with $\sigma_N$ denoting the standard deviation of the call count data and $E[N]$ its mean, we can estimate SCV as $\sigma_N^2/E[N]$. We first present, as a benchmark, the naive approach that corresponds to assuming that the arrivals are Poisson distributed and using the Erlang-C formula to determine the appropriate staffing. We then compute the staffing level using Ha-Queueing, the z-Square root, and SCV-based Whitt (1993) methods in order to reach a 90% probability of being immediately answered. Finally we simulated the system assuming that the arrivals were distributed according to a Poisson process with a rate varying

**Figure 7 Profile of Daily Demand for 5 Days of a Week and for Three Services**

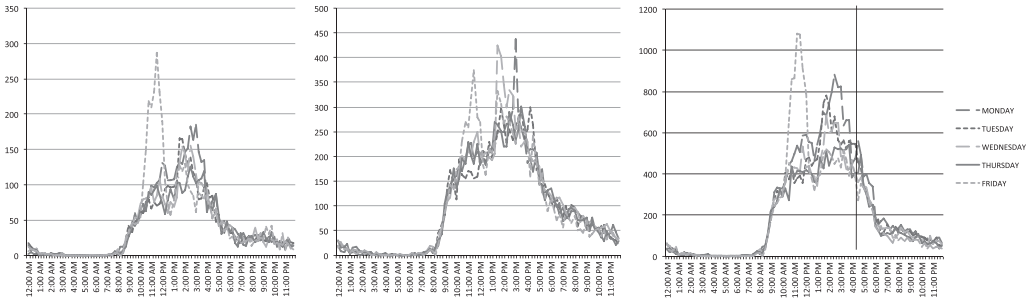



**Table 8 Comparison of Peakedness Methods on Real-Life Data**

| | Service 1 | Service 2 | Service 3 |
|---|---|---|---|
| Arrival rate (per minute) | 7.77 | 17.72 | 37.3 |
| Service time (in minutes) | 3.72 | 2.24 | 4.03 |
| Peakedness estimate | 2.43 | 2.2 | 8.01 |
| Approximated SCV | 6.91 | 9.18 | 27.77 |
| **Staffing for 10% waiting probability** | | | |
| Erlang-C staffing | 37 | 49 | 169 |
| Ha-Queueing | 42 | 54 | 204 |
| z-Square root | 42 | 53 | 209 |
| SCV-based | 46 | 61 | 203 |
| Simulation | 44 | 58 | 218 |
| **Staffing for 90% service level (30s)** | | | |
| Erlang-C staffing | 35 | 45 | 160 |
| Ha-Queueing | 39 | 50 | 191 |
| z-Square root | 38 | 47 | 178 |
| SCV-based | 38 | 49 | 174 |
| Simulation | 42 | 56 | 212 |

every 15 minutes in order to correspond with the observed arrivals of each interval.

From the results of Table 8, we observe that arrivals are indeed much more variable than a Poisson process. We also see that all approaches improve the staffing estimate compared to the classic Erlang-C formula. The results confirm what is observed in our controlled experiment. The method based on the peakedness and Hayward's approximation is more robust and nicely complements the square root staffing or SCV-based staffing methods when the arrivals get very variable (peakedness $\geq 2.5$).

## 8. Managerial Implications

Our results are applicable in co-sourcing relationships where calls are shared with a vendor. In addition to providing simple quantifications of the effect of arrivals with short-term variability on the vendor's operations, we provide a data-based approach to its measurement. Our analysis highlights the variability of arrivals as an important issue to be managed for successful vendor operations.

In the analysis we considered the example of overflow type arrivals at the vendor as a prototypical case of arrivals with high short-term variability, characterized by an arrival stream with peakedness $> 1$. Overflows are not the only form of call sharing encountered in such settings. Our approach enables the measurement of the peakedness of the outsourced call stream. If peakedness is found to be different from 1 (corresponding to non-Poisson arrivals) the proposed staffing approximations will apply. Depending on how call sharing is actually implemented in real time, the vendor may observe arrival streams with different peakedness for the same base volume objective (or other objective) of the client. Thus, the possibility of measurement is a critical success factor for the vendor. We next discuss how this issue will affect outsourcing contracts between the client and vendor.

In a setting with co-sourcing, our analysis has shown that the vendor's staffing costs depend on the mean demand as well as the peakedness of the arrival flow. Recalling the peakedness-adjusted square root staffing rule, we observe that in addition to the economies of scale in arrival rate, the staffing at the vendor exhibits dis-economies of scale in the peakedness $z$, albeit as a second-order effect. In other words, the vendor firm will prefer a client that shares a higher demand volume with lower peakedness. Depending on the relationship between the two parties, the client may promise an average call volume over an agreed upon time interval or may provide minimum and maximum levels within which call volumes can fluctuate to aid in the vendor's staffing decision. However, in a typical setting where operational details are not shared, the vendor will not know the peakedness of the underlying demand, and the client will have an incentive to misrepresent it.

Taking a contracting approach to the problem, the vendor would want to screen clients for their call volumes as well as the peakedness of their call streams. In a setting with information asymmetry regarding call volumes, Akan et al. (2011) show that a two-part tariff can screen different types of clients. This result relies on the underlying economies of scale. In a setting with information asymmetry and call sharing as considered in this study, this screening would have to be done for both the arrival rate and its peakedness.

Screening will be difficult as the peakedness will subdue some of the economies of scale. There will be two dimensions of information asymmetry and the two dimensions are not perfectly correlated. Our measurement based approach provides a practical alternative to this difficult screening situation. Instead of trying to screen the clients up front, vendors may include measurement based peakedness penalties to compensate for the dis-economies of scale these cause in their operations. Contracts that disregard the peakedness issue can be extended to include conditions regarding measured peakedness.

## 9. Concluding Remarks

We analyze the capacity optimization problem faced by a call center service provider (vendor) in service supply chains. The main challenge for the vendor is that it receives a call stream that has more variability than a Poisson arrival stream and cannot forecast or estimate this variability up front. If the vendor staffs according to average demand, as a result of the lack of information regarding variability, this ignored variability will lead to undesirable service levels. Without measuring the variability of the arrival rate the vendor is left to guess how to staff. To overcome the problem, we propose to measure the variability of the overflow arrivals, making use of the peakedness, and use this to staff the second tier call center. We propose two approaches to staffing, one which is an adjustment of the square root staffing rule and the other that makes use of an extension of Hayward's approximation for the queueing system of the vendor. While both methods give very satisfactory results in terms of estimating the required workforce to cope with arrivals that exhibit a moderate peakedness, for higher values ($\geq 2.5$), the second method appears to be superior.

In modeling the vendor's operations, we have considered an infinite queue with patient customers. In practice, call center customers are known to abandon, so that the vendor's system will also have abandonments. As shown in earlier literature (see, e.g., Garnett et al. 2002) abandonments have a smoothing effect on service performance. Abandoning customers free system resources up for non-abandoning ones, thus improving service levels experienced by those who stay. Qualitatively, this implies that with abandonment, the differences in service levels or in staff levels that we have for the different cases in our examples will be smaller.

Even though we focused on the simplest case with a vendor having a single client, the qualitative nature of our results will continue to hold when there are several clients. If it is possible for the vendor to pool calls from several clients in the multi-client case, it will face an arrival stream that is the sum of several overflow streams. While the vendor will benefit from pooling on the supply side, this will not necessarily reduce the peakedness. Thus the effects we have shown in the examples throughout the paper will still be present. Moreover, since our methods are based on the measured peakedness, the vendor can directly measure the peakedness after calls have been pooled and implicitly take into account correlations that might exist between different arrival streams. If the vendor cannot pool calls from different clients, this implies that it will be managing several small pools of servers, one for each client. As often with stochastic systems, smaller pools will exacerbate the effect of the variability on performance.

## Aknowledgments

The authors gratefully acknowledge the support of the Belgian French-speaking Community (ARC project 08-013).

## Appendix A

### Proof of Proposition 1

PROPOSITION 1. *If the arrival process is non-arithmetic (there does not exist a real $d$ such that arrivals can only occur at times that are multiples of $d$) and the interarrival process is independent of the data collection intervals, then*

$$\bar{z} \simeq \frac{1}{2} - \frac{1}{\mu T} \cdot \frac{1 - e^{-\mu T}}{1 + e^{-\mu T}} + \frac{\mathrm{Var}[\bar{S}_n]}{E[\bar{S}_n]} \left( \frac{1 - e^{-\mu T}}{\mu T} \right). \quad \text{(A1)}$$

PROOF. The conditions on the arrival process ensure that we can estimate $\bar{z}$ from the number of busy virtual operators in the virtual fluid pool when sampled at times $0,T,2T,\ldots$; in other words

$$\bar{z} \simeq \frac{\mathrm{Var}[\bar{S}(nT)]}{E[\bar{S}(nT)]}.$$

We can compute $\bar{S}(nT)$ using the following equation:

$$\bar{S}(nT) = \bar{S}((n-1)T)e^{-\mu T} + \sum_{i \in B_n} e^{-\mu(nT - t_i)}, \quad \text{(A2)}$$

where $B_n$ is the set of arrivals in the $n$th batch (i.e., arrivals in the interval $((n-1)T,nT]$), and $t_i$ is the time of arrival $i$. On the other hand, the number of busy servers when arrivals are aggregated is given by

$$\bar{S}_n = \bar{S}_{n-1}e^{-\mu T} + |B_n|. \quad \text{(A3)}$$

For each arrival the difference between the two expressions is the factor $e^{-\mu(nT-t_i)}$ that is replaced by 1 in the second case. Under the conditions of the proposition, the distribution of the arrival time $t_i$ of any arrival with respect to the batch interval is uniform. Put another way, the difference $\tau_i = nT - t_i$ is a random variable with the same distribution as $\tau$, which is uniformly distributed over the interval $[0,T]$.

From Equation (A3) we find that

$$E[\bar{S}_n] = \frac{E[|B_n|]}{1-e^{-\mu T}}.$$

And, for the variance we get

$$\begin{aligned}\mathrm{Var}[\bar{S}_n] &= \mathrm{Var}[\bar{S}_n]e^{-2\mu T} + \mathrm{Var}[|B_n|] + 2\,\mathrm{Cov}[\bar{S}_n, |B_n|]e^{-\mu T} \\ &= \frac{\mathrm{Var}[|B_n|] + 2\,\mathrm{Cov}[\bar{S}_n, |B_n|]e^{-\mu T}}{1-e^{-2\mu T}}.\end{aligned}$$

$$\mathrm{Var}[\bar{S}(nT)] \simeq \frac{\mathrm{Var}[e^{-\mu\tau}]E[|B_n|] + E^2[e^{-\mu\tau}]\mathrm{Var}[|B_n|] + 2\,\mathrm{Cov}[\bar{S}_n, |B_n|]E^2[e^{-\mu\tau}]e^{-\mu T}}{1-e^{-2\mu T}}.$$

Dividing $\mathrm{Var}[\bar{S}_n]$ by $E[\bar{S}_n]$ yields

$$\frac{\mathrm{Var}[\bar{S}_n]}{E[\bar{S}_n]} = \frac{\mathrm{Var}[|B_n|] + 2\,\mathrm{Cov}[\bar{S}_n, |B_n|]e^{-\mu T}}{E[|B_n|](1+e^{-\mu T})}.$$

We now focus on $\bar{S}(nT)$. From Equation (A2), we immediately find

$$\begin{aligned}E[\bar{S}(nT)] &= \frac{E[\sum_{i\in B_n} e^{-\mu\tau_i}]}{1-e^{-\mu T}} \\ &= \frac{E[|B_n|]E[e^{-\mu\tau}]}{1-e^{-\mu T}}.\end{aligned}$$

And, for the variance,

$$\begin{aligned}&\mathrm{Var}[\bar{S}(nT)] = \mathrm{Var}[\bar{S}(nT)]e^{-2\mu T} + \mathrm{Var}\left[\sum_{i\in B_n} e^{-\mu\tau_i}\right] + 2\,\mathrm{Cov}[\bar{S}(nT), \\ &\sum_{i\in B_n} e^{-\mu\tau_i}]e^{-\mu T} = \frac{\mathrm{Var}[\sum_{i\in B_n} e^{-\mu\tau_i}] + 2\,\mathrm{Cov}[\bar{S}(nT), \sum_{i\in B_n} e^{-\mu\tau_i}]e^{-\mu T}}{1-e^{-2\mu T}}.\end{aligned}$$

We use the law of total variance to solve $\mathrm{Var}[\sum_{i\in B_n} e^{-\mu\tau_i}]$. This gives

$$\begin{aligned}\mathrm{Var}[\sum_{i\in B_n} e^{-\mu\tau_i}] &= E\left[\mathrm{Var}\left[\sum_{i\in B_n} e^{-\mu\tau_i}|B_n\right]\right] \\ &\quad + \mathrm{Var}\left[E\left[\sum_{i\in B_n} e^{-\mu\tau_i}|B_n\right]\right].\end{aligned}$$

Assuming that all $\tau_i$'s are independent, this expression is equivalent to

$$\mathrm{Var}\left[\sum_{i\in B_n} e^{-\mu\tau_i}\right] = \mathrm{Var}[e^{-\mu\tau}]E[|B_n|] + E^2[e^{-\mu\tau}]\mathrm{Var}[|B_n|].$$

As the introduction of any arrival in $S(nT)$ is delayed by a time uniformly distributed over $[0,T]$, the amount of fluid operators answering any arrival in $\bar{S}(nT)$ is on average smaller by a factor $E[e^{-\mu\tau}]$ than the amount of operators in $S_n$. We therefore assume that $\mathrm{Cov}[\bar{S}(nT), \sum_{i\in B_n} e^{-\mu\tau_i}] = \mathrm{Cov}[\bar{S}_n, |B_n|]E^2[e^{-\mu\tau}]$. Therefore,

We obtain an approximation for $\bar{z}$ by dividing this last expression by $E[\bar{S}(nT)]$:

$$\begin{aligned}\bar{z} &\simeq \frac{\mathrm{Var}[e^{-\mu\tau}]}{E[e^{-\mu\tau}](1+e^{-\mu T})} \\ &\quad + E[e^{-\mu\tau}]\left(\frac{\mathrm{Var}[|B_n|] + 2\,\mathrm{Cov}[\bar{S}_n, |B_n|]e^{-\mu T}}{E[|B_n|](1+e^{-\mu T})}\right) \\ &\simeq \frac{\mathrm{Var}[e^{-\mu\tau}]}{E[e^{-\mu\tau}](1+e^{-\mu T})} + E[e^{-\mu\tau}]\frac{\mathrm{Var}[\bar{S}_n]}{E[\bar{S}_n]}.\end{aligned}$$

As,

$$\begin{aligned}E[e^{-\mu\tau}] &= \frac{1}{T}\int_0^T e^{-\mu t}dt \\ &= \frac{1-e^{-\mu T}}{\mu T} \\ \mathrm{Var}[e^{-\mu\tau}] &= \frac{1}{T}\int_0^T e^{-2\mu t}dt - \left(\frac{1-e^{-\mu T}}{\mu T}\right)^2 \\ &= \frac{1-e^{-2\mu T}}{2\mu T} - \left(\frac{1-e^{-\mu T}}{\mu T}\right)^2,\end{aligned}$$

we can write

$$\bar{z} \simeq \frac{1}{2} - \frac{1}{\mu T}\cdot\frac{1-e^{-\mu T}}{1+e^{-\mu T}} + \frac{\mathrm{Var}[\bar{S}_n]}{E[\bar{S}_n]}\left(\frac{1-e^{-\mu T}}{\mu T}\right).$$

**Table A1 Proportion of Time Each State Is Reached over the Time Horizon of the 400 Arrivals of Table A2**

| $p_0$ | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ | $p_7$ | $p_8$ | $p_9$ | $p_{10}$ | $p_{11}$ | $p_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0416 | 0.0288 | 0.0717 | 0.1296 | 0.169 | 0.1945 | 0.1501 | 0.1054 | 0.0616 | 0.0212 | 0.0189 | 0.0048 | 0.0027 |

## Appendix B. Measurement Example

### Computing the Peakedness

In the following example we illustrate the different ways of computing the peakedness. Table A2 gives the arrival times of 400 arrivals of a Poisson process with rate $\lambda = 10$/hour. In Table A3 we show the service times for these arrivals in the fictitious pool with an infinite number of operators. Service times are exponentially distributed with parameter $\mu = 2$/hour. Figure 3 shows $S(t)$, the number of busy servers at time $t$, for the first 15 time units of the time horizon. Note that the initial number of busy servers has been set to $S(0) = \lambda/\mu = 5$.

We compute $E[S] = \sum_{i=0}^{\infty} ip_i$ (where $p_i$ is the proportion of time when $S(t) = i$) and $\mathrm{Var}[S] = E[S^2] - E^2[S]$. In the example we observe that the maximum value $S(t)$ takes over the time horizon of 41.39 is equal to 12. Table A1 presents the values of $p_i$ for $i \in \{0,1,\ldots,12\}$. We find that $E[S] = 4.81$, $E[S^2] = 28.14$ and $\mathrm{Var}[S] = 5.04$. $z$ is thus equal to 1.05, which is not far from the theoretical value of $z = 1$ for a Poisson process.

### Computing the Fluid Peakedness

We continue our illustrative example and estimate the fluid peakedness for the sample of arrivals in Table A2. This time the services times of Table A3 are not required. $\bar{S}(t)$ suffers an exponential decay with parameter $\mu = 2$. Figure 4 shows the evolution of $\bar{S}(t)$ over the first 15 time units of the time horizon. We start out by computing the number of busy operators for each arrival time. Denote by $t_i$ the arrival time of the $i^{th}$ arrival; $\bar{S}(t_i)$ is computed using the property that $\bar{S}(t_i) = \bar{S}(t_{i-1})\, e^{-\mu(t_i - t_{i-1})} + 1$. To limit the bias at the beginning of the time horizon we set $\bar{S}(0) = \lambda/\mu = 5$. We randomly choose 50 time points on the time horizon and for each of these 50 epochs $t^*$ we compute the value of $\bar{S}(t^*)$ using the fact that the process suffers an exponential decay between arrivals. We thus have 50 measurements of $\bar{S}(t)$ at random times, and we compute the mean and variance of this sample. We obtain $E[\bar{S}] = 4.98$ and $\mathrm{Var}[\bar{S}] = 2.52$, resulting in $\bar{z} = 0.51$. Adding 1/2, according to Equation (1), yields a value of 1.01.

### Computing the Peakedness from Aggregate Data

We next aggregate the Poisson arrival data in time intervals of $T = 0.25$, corresponding to 15-minute time intervals. The number of arrivals in each interval, $X_n$, is displayed in Table A5. Note that the last interval considered in this example goes from 41 to 41.25. The two arrivals after 41.25 are neglected. For each interval $n$ we compute $\bar{S}_n$. To decrease the bias brought by warm-up effects we assume $\bar{S}_0 = \frac{\lambda T}{\mu}\frac{e^{-\mu T}}{1-e^{-\mu T}} = 3.85$. Figure 5 depicts the fluid number of busy servers with batch arrivals.

From the computed values of $\bar{S}_n$ we find $E[\bar{S}_n] = 6.09$ and $\mathrm{Var}[\bar{S}_n] = 3.79$. Using (1) with $\mu = 2$ and $T = 0.25$, we find $\bar{z} = 0.5$.

**Table A2 List of 400 Arrivals of a Poisson Process of Rate 10**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0.1574 | 4.5578 | 8.6309 | 14.4950 | 20.3141 | 25.2501 | 29.2269 | 35.0697 |
| 0.2750 | 4.5888 | 8.7309 | 14.5185 | 20.4512 | 25.4198 | 29.4422 | 35.1320 |
| 0.3587 | 4.6274 | 8.7637 | 14.7140 | 20.5882 | 25.5880 | 29.4901 | 35.1807 |
| 0.4207 | 4.7549 | 8.9381 | 14.8188 | 20.7536 | 25.5906 | 29.5842 | 35.2351 |
| 0.4285 | 4.8250 | 8.9908 | 14.9402 | 20.9248 | 25.8435 | 29.9743 | 35.3311 |
| 0.5885 | 4.9993 | 9.0382 | 14.9647 | 21.0401 | 25.8508 | 30.0158 | 35.3379 |
| 0.6000 | 5.0261 | 9.1333 | 14.9733 | 21.1407 | 25.8788 | 30.1745 | 35.4270 |
| 0.6599 | 5.0858 | 9.1444 | 14.9906 | 21.2343 | 25.9190 | 30.3488 | 35.4722 |
| 0.7246 | 5.1388 | 9.8160 | 15.2009 | 21.3500 | 26.0994 | 30.3863 | 35.6719 |
| 0.8576 | 5.1764 | 10.3124 | 15.3000 | 21.3642 | 26.3528 | 30.4315 | 35.7917 |
| 0.8582 | 5.1979 | 10.5366 | 15.3644 | 21.3868 | 26.3674 | 30.5611 | 36.1994 |
| 1.0507 | 5.2409 | 10.7432 | 15.6688 | 21.4341 | 26.3938 | 30.7024 | 36.4228 |
| 1.3369 | 5.2801 | 10.8241 | 15.8770 | 21.4641 | 26.4009 | 30.8238 | 36.5656 |
| 1.3769 | 5.3515 | 11.0285 | 15.9266 | 21.7823 | 26.6006 | 31.0316 | 36.6438 |
| 1.3873 | 5.6367 | 11.0657 | 16.0155 | 21.8438 | 26.8660 | 31.2814 | 36.6921 |
| 1.4422 | 5.7749 | 11.2731 | 16.1458 | 22.0639 | 27.0049 | 31.4123 | 36.8318 |
| 1.6333 | 5.8224 | 11.3418 | 16.2899 | 22.0970 | 27.0668 | 31.5042 | 36.8499 |
| 1.6501 | 5.8467 | 11.4427 | 16.6229 | 22.1069 | 27.0958 | 31.5895 | 36.9219 |
| 1.7894 | 5.8881 | 11.6511 | 16.6886 | 22.1150 | 27.1705 | 31.9654 | 37.0212 |
| 1.9289 | 6.0341 | 11.7110 | 16.7953 | 22.3769 | 27.2530 | 32.0442 | 37.1467 |

(continued)

**Table A2 Continued**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2.0178 | 6.1512 | 11.7359 | 16.8104 | 22.4889 | 27.2791 | 32.1361 | 37.4496 |
| 2.0988 | 6.2462 | 11.7928 | 17.0903 | 22.5244 | 27.2801 | 32.3390 | 37.5531 |
| 2.3432 | 6.3227 | 11.8096 | 17.1096 | 22.5593 | 27.3962 | 32.3680 | 37.5637 |
| 2.3651 | 6.3243 | 11.8770 | 17.1318 | 22.5613 | 27.4849 | 32.5971 | 37.8305 |
| 2.4027 | 6.3490 | 11.9110 | 17.2223 | 22.5675 | 27.6439 | 32.6354 | 37.9119 |
| 2.4535 | 6.3892 | 11.9362 | 17.3454 | 22.7333 | 27.7660 | 32.6931 | 38.0334 |
| 2.5150 | 6.5491 | 11.9517 | 17.3586 | 22.8352 | 27.8511 | 32.7406 | 38.0634 |
| 2.6320 | 6.5985 | 12.2008 | 17.4712 | 23.0093 | 27.9772 | 32.8626 | 38.1468 |
| 2.7538 | 6.6754 | 12.2563 | 17.5475 | 23.1630 | 28.1486 | 32.9572 | 38.3943 |
| 2.8011 | 6.7485 | 12.3221 | 17.6608 | 23.3303 | 28.1680 | 32.9915 | 38.4677 |
| 2.9373 | 6.7669 | 12.3265 | 17.6938 | 23.3452 | 28.2053 | 33.0808 | 38.5895 |
| 2.9473 | 6.7735 | 12.5571 | 17.8699 | 23.6793 | 28.2438 | 33.0851 | 38.6182 |
| 2.9724 | 6.7862 | 12.9381 | 18.3092 | 23.8721 | 28.3308 | 33.1017 | 38.7564 |
| 3.0811 | 6.8197 | 13.0536 | 18.3112 | 23.9624 | 28.3389 | 33.2992 | 39.5334 |
| 3.2195 | 6.8653 | 13.0882 | 18.3779 | 24.0301 | 28.3705 | 33.3270 | 39.5514 |
| 3.2527 | 6.9886 | 13.1469 | 18.4379 | 24.0370 | 28.4407 | 33.3957 | 39.6261 |
| 3.2677 | 7.0241 | 13.3357 | 18.6989 | 24.1307 | 28.4973 | 33.5079 | 40.1965 |
| 3.5476 | 7.1311 | 13.3475 | 18.7792 | 24.1822 | 28.5184 | 33.5955 | 40.3122 |
| 3.5828 | 7.2917 | 13.3636 | 18.8187 | 24.2895 | 28.5984 | 33.6374 | 40.3713 |
| 3.6549 | 7.4623 | 13.4003 | 18.8297 | 24.3434 | 28.6371 | 33.6486 | 40.4198 |
| 3.6963 | 7.6231 | 13.4642 | 19.0971 | 24.3907 | 28.6511 | 33.9078 | 40.5301 |
| 3.8420 | 7.7129 | 13.6398 | 19.1702 | 24.4091 | 28.7556 | 34.1390 | 40.5689 |
| 3.8474 | 7.8130 | 13.6794 | 19.3322 | 24.5045 | 28.7646 | 34.1785 | 40.6883 |
| 4.1552 | 8.0790 | 13.7326 | 19.4460 | 24.7358 | 28.9345 | 34.2260 | 40.9937 |
| 4.1995 | 8.1167 | 13.7683 | 19.4626 | 24.9090 | 29.0361 | 34.5206 | 41.0377 |
| 4.2835 | 8.1534 | 13.8978 | 19.4748 | 24.9398 | 29.0498 | 34.5840 | 41.1254 |
| 4.3678 | 8.2170 | 13.9320 | 19.8083 | 24.9638 | 29.0854 | 34.7196 | 41.2111 |
| 4.4180 | 8.2482 | 14.1278 | 19.8644 | 24.9723 | 29.1229 | 34.7774 | 41.2167 |
| 4.4338 | 8.2749 | 14.1339 | 19.8659 | 25.1153 | 29.1282 | 34.8338 | 41.3906 |
| 4.4746 | 8.4260 | 14.2194 | 20.2900 | 25.1768 | 29.1917 | 35.0178 | 41.3970 |

**Table A3 List of 405 Services Times from an Exponential Distribution of Parameter $\mu = 2$; The Five First Arrivals in the First Column Are the Service Times of the Five Calls in Service at $t = 0$**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0.2750 | 0.0735 | 0.0061 | 0.0040 | 0.0836 | 0.0558 | 0.1402 | 0.5596 | 0.9375 |
| 0.4737 | 0.0307 | 0.0044 | 0.3427 | 0.1803 | 0.4076 | 1.0783 | 0.2197 | 1.0077 |
| 0.6147 | 2.2521 | 0.4911 | 0.1319 | 0.3842 | 0.2466 | 2.1402 | 0.0921 | 1.6581 |
| 0.4962 | 2.0606 | 0.1623 | 0.0113 | 0.3454 | 0.2452 | 0.0672 | 1.2138 | 0.0275 |
| 0.8685 | 0.2895 | 0.1129 | 0.4183 | 0.3101 | 0.0909 | 0.4156 | 1.0533 | 0.4945 |
| | 0.3023 | 0.0238 | 0.1676 | 0.1130 | 0.0090 | 0.2292 | 1.4557 | 0.2742 |
| | 0.0465 | 0.3188 | 0.3472 | 0.3225 | 0.1808 | 0.1441 | 0.1792 | 0.1920 |
| | 1.2372 | 0.1531 | 0.0336 | 0.0637 | 0.0454 | 0.0470 | 0.0502 | 0.5944 |
| | 0.0893 | 0.2012 | 0.0088 | 0.9115 | 2.2115 | 0.1515 | 0.1181 | 0.0342 |
| | 2.0162 | 0.0993 | 0.3396 | 0.5938 | 0.0451 | 0.0105 | 0.5408 | 0.0950 |
| | 0.1972 | 0.5239 | 0.6823 | 0.6935 | 1.2905 | 0.7136 | 1.0686 | 0.0847 |
| | 0.3130 | 0.0610 | 0.3566 | 0.0284 | 0.1025 | 0.2026 | 0.0783 | 0.4413 |
| | 0.0678 | 0.6430 | 0.4436 | 0.3258 | 1.6942 | 1.4437 | 0.7915 | 0.7891 |
| | 0.2474 | 0.0500 | 0.3506 | 1.5866 | 0.0222 | 1.0083 | 0.0217 | 1.4288 |
| | 0.1786 | 0.6817 | 0.4817 | 1.4415 | 1.7809 | 0.3627 | 0.1291 | 0.2175 |
| | 1.3016 | 0.2725 | 0.3345 | 0.3399 | 0.1619 | 0.2750 | 0.3112 | 0.9304 |
| | 0.8128 | 0.0532 | 0.2286 | 0.2771 | 0.4258 | 0.0745 | 0.3338 | 0.4303 |
| | 0.9273 | 0.0122 | 0.3702 | 0.0021 | 0.0379 | 1.0523 | 0.1295 | 0.0821 |
| | 0.0950 | 0.3411 | 0.2428 | 0.9715 | 0.3329 | 0.4779 | 0.1823 | 1.3085 |
| | 1.2456 | 0.1383 | 0.9211 | 0.2181 | 0.0596 | 1.3308 | 0.2762 | 0.3466 |
| | 0.1763 | 0.4684 | 0.9756 | 0.7181 | 0.0578 | 0.7779 | 0.1426 | 1.1536 |
| | 0.5798 | 0.4657 | 0.3996 | 0.7366 | 1.4370 | 0.2914 | 0.6224 | 0.8100 |
| | 0.3017 | 0.3701 | 0.4446 | 0.3745 | 0.1039 | 1.1990 | 0.8208 | 0.3611 |
| | 0.1547 | 0.0129 | 0.1474 | 0.2925 | 0.4383 | 0.3144 | 0.0516 | 0.0217 |
| | 0.8071 | 1.9291 | 1.2664 | 1.0741 | 0.1358 | 0.1159 | 0.3260 | 0.3971 |
| | 1.2750 | 0.7836 | 0.5698 | 0.0126 | 0.0199 | 0.2836 | 0.2181 | 1.0057 |

(continued)

**Table A3 Continued**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0.0690 | 0.0586 | 0.2510 | 0.4023 | 0.3487 | 0.0085 | 1.0167 | 0.0428 |
| 0.8603 | 0.9357 | 0.4419 | 0.6275 | 2.2196 | 0.8666 | 0.3616 | 0.1778 |
| 0.7404 | 0.4365 | 2.3478 | 1.1876 | 0.4428 | 2.5290 | 0.0409 | 0.7374 |
| 0.3533 | 1.3877 | 0.3187 | 0.1038 | 1.0719 | 0.8183 | 0.1410 | 0.3353 |
| 0.0351 | 0.0157 | 0.3170 | 0.3935 | 0.9509 | 0.1833 | 0.0249 | 0.5775 |
| 0.0303 | 0.0233 | 1.0117 | 0.4092 | 2.5741 | 0.4433 | 0.7271 | 0.0596 |
| 0.4103 | 0.6884 | 0.2258 | 0.2551 | 1.1799 | 0.3048 | 0.5559 | 0.0362 |
| 0.2304 | 1.0605 | 1.0798 | 0.4760 | 0.0125 | 0.7506 | 1.0976 | 0.1252 |
| 0.6000 | 0.3088 | 1.3182 | 0.1221 | 0.1209 | 0.2965 | 0.0079 | 0.2963 |
| 0.6669 | 0.0134 | 0.3535 | 0.2929 | 0.1964 | 1.7118 | 0.4400 | 0.2476 |
| 0.1817 | 0.7643 | 0.3245 | 0.3598 | 0.2346 | 0.9730 | 0.2745 | 0.0501 |
| 0.2082 | 0.0066 | 0.0172 | 0.3008 | 0.2852 | 0.3318 | 0.0506 | 0.7760 |
| 0.5373 | 0.5339 | 0.3257 | 0.8716 | 0.0103 | 0.2780 | 0.4739 | 0.0210 |
| 0.7653 | 1.4050 | 0.1934 | 0.3912 | 0.1262 | 0.8895 | 0.7682 | 0.6255 |
| 0.2560 | 0.5099 | 0.6032 | 0.0826 | 0.0727 | 0.1567 | 0.2752 | 0.5248 |
| 0.2721 | 0.5560 | 0.1442 | 0.3186 | 0.0025 | 0.2641 | 0.6862 | 0.9174 |
| 0.3949 | 0.2735 | 0.0314 | 0.0639 | 0.0392 | 0.0317 | 0.2004 | 0.3181 |
| 0.0754 | 0.6851 | 0.2168 | 1.1455 | 0.1814 | 0.2851 | 0.2331 | 1.0527 |
| 1.1831 | 0.0063 | 0.2689 | 0.7292 | 0.0429 | 0.0121 | 1.2645 | 0.3760 |
| 0.2505 | 0.2390 | 0.5061 | 0.1476 | 0.0396 | 0.7254 | 0.9941 | 0.2920 |
| 0.8267 | 1.0515 | 1.1483 | 0.0364 | 2.1235 | 0.7277 | 0.2074 | 0.0288 |
| 0.4831 | 0.2662 | 1.0418 | 0.6949 | 0.0887 | 0.2884 | 0.5565 | 1.0341 |
| 0.4850 | 0.2314 | 1.4747 | 1.8666 | 0.4720 | 0.2767 | 0.2175 | 0.2246 |
| 0.5365 | 0.0315 | 0.9623 | 1.2813 | 1.1480 | 1.1593 | 0.8941 | 0.1132 |

**Table A4 Simulation DataSet**

| Example | $\lambda$ | $z$ | $S_v$ | Example | $\lambda_v$ | $z$ | $S_v$ | Example | $\lambda_v$ | $z$ | $S_v$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3.20 | 1.13 | 4 | 49 | 12.28 | 1.24 | 13 | 97 | 45.32 | 1.30 | 46 |
| 2 | 3.20 | 1.13 | 5 | 50 | 12.28 | 1.24 | 14 | 98 | 45.32 | 1.30 | 50 |
| 3 | 3.20 | 1.13 | 6 | 51 | 12.28 | 1.24 | 15 | 99 | 45.32 | 1.30 | 54 |
| 4 | 3.20 | 1.13 | 7 | 52 | 12.28 | 1.24 | 16 | 100 | 45.32 | 1.30 | 58 |
| 5 | 2.46 | 1.28 | 3 | 53 | 8.72 | 1.56 | 9 | 101 | 30.89 | 1.81 | 31 |
| 6 | 2.46 | 1.28 | 4 | 54 | 8.72 | 1.56 | 10 | 102 | 30.89 | 1.81 | 35 |
| 7 | 2.46 | 1.28 | 5 | 55 | 8.72 | 1.56 | 11 | 103 | 30.89 | 1.81 | 39 |
| 8 | 2.46 | 1.28 | 6 | 56 | 8.72 | 1.56 | 12 | 104 | 30.89 | 1.81 | 43 |
| 9 | 1.80 | 1.42 | 2 | 57 | 5.48 | 1.98 | 6 | 105 | 17.17 | 2.73 | 18 |
| 10 | 1.80 | 1.42 | 3 | 58 | 5.48 | 1.98 | 7 | 106 | 17.17 | 2.73 | 20 |
| 11 | 1.80 | 1.42 | 4 | 59 | 5.48 | 1.98 | 8 | 107 | 17.17 | 2.73 | 22 |
| 12 | 1.80 | 1.42 | 5 | 60 | 5.48 | 1.98 | 9 | 108 | 17.17 | 2.73 | 24 |
| 13 | 1.24 | 1.54 | 2 | 61 | 2.80 | 2.40 | 3 | 109 | 5.78 | 4.08 | 6 |
| 14 | 1.24 | 1.54 | 3 | 62 | 2.80 | 2.40 | 4 | 110 | 5.78 | 4.08 | 8 |
| 15 | 1.24 | 1.54 | 4 | 63 | 2.80 | 2.40 | 5 | 111 | 5.78 | 4.08 | 10 |
| 16 | 1.24 | 1.54 | 5 | 64 | 2.80 | 2.40 | 6 | 112 | 5.78 | 4.08 | 12 |
| 17 | 6.24 | 1.19 | 7 | 65 | 15.29 | 1.25 | 16 | 113 | 60.32 | 1.31 | 62 |
| 18 | 6.24 | 1.19 | 8 | 66 | 15.29 | 1.25 | 17 | 114 | 60.32 | 1.31 | 66 |
| 19 | 6.24 | 1.19 | 9 | 67 | 15.29 | 1.25 | 18 | 115 | 60.32 | 1.31 | 70 |
| 20 | 6.24 | 1.19 | 10 | 68 | 15.29 | 1.25 | 19 | 116 | 60.32 | 1.31 | 74 |
| 21 | 4.60 | 1.41 | 5 | 69 | 10.76 | 1.61 | 11 | 117 | 40.92 | 1.85 | 41 |
| 22 | 4.60 | 1.41 | 6 | 70 | 10.76 | 1.61 | 12 | 118 | 40.92 | 1.85 | 45 |
| 23 | 4.60 | 1.41 | 7 | 71 | 10.76 | 1.61 | 13 | 119 | 40.92 | 1.85 | 49 |
| 24 | 4.60 | 1.41 | 8 | 72 | 10.76 | 1.61 | 14 | 120 | 40.92 | 1.85 | 53 |
| 25 | 3.12 | 1.66 | 4 | 73 | 6.60 | 2.09 | 7 | 121 | 22.31 | 2.89 | 23 |
| 26 | 3.12 | 1.66 | 5 | 74 | 6.60 | 2.09 | 8 | 122 | 22.31 | 2.89 | 26 |
| 27 | 3.12 | 1.66 | 6 | 75 | 6.60 | 2.09 | 9 | 123 | 22.31 | 2.89 | 29 |
| 28 | 3.12 | 1.66 | 7 | 76 | 6.60 | 2.09 | 10 | 124 | 22.31 | 2.89 | 32 |
| 29 | 1.88 | 1.89 | 2 | 77 | 3.18 | 2.61 | 4 | 125 | 6.73 | 4.62 | 7 |
| 30 | 1.88 | 1.89 | 3 | 78 | 3.18 | 2.61 | 5 | 126 | 6.73 | 4.62 | 8 |
| 31 | 1.88 | 1.89 | 4 | 79 | 3.18 | 2.61 | 6 | 127 | 6.73 | 4.62 | 9 |
| 32 | 1.88 | 1.89 | 5 | 80 | 3.18 | 2.61 | 7 | 128 | 6.73 | 4.62 | 10 |
| 33 | 9.27 | 1.22 | 10 | 81 | 30.31 | 1.29 | 31 | 129 | 75.32 | 1.31 | 85 |

(continued)

**Table A4 Continued**

| Example | $\lambda$ | $z$ | $S_v$ | Example | $\lambda_v$ | $z$ | $S_v$ | Example | $\lambda_v$ | $z$ | $S_v$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 34 | 9.27 | 1.22 | 11 | 82 | 30.31 | 1.29 | 34 | 130 | 75.32 | 1.31 | 90 |
| 35 | 9.27 | 1.22 | 12 | 83 | 30.31 | 1.29 | 37 | 131 | 75.32 | 1.31 | 95 |
| 36 | 9.27 | 1.22 | 13 | 84 | 30.31 | 1.29 | 40 | 132 | 75.32 | 1.31 | 100 |
| 37 | 6.67 | 1.50 | 7 | 85 | 20.85 | 1.74 | 21 | 133 | 50.93 | 1.87 | 51 |
| 38 | 6.67 | 1.50 | 8 | 86 | 20.85 | 1.74 | 24 | 134 | 50.93 | 1.87 | 55 |
| 39 | 6.67 | 1.50 | 9 | 87 | 20.85 | 1.74 | 27 | 135 | 50.93 | 1.87 | 60 |
| 40 | 6.67 | 1.50 | 10 | 88 | 20.85 | 1.74 | 30 | 136 | 50.93 | 1.87 | 65 |
| 41 | 4.33 | 1.84 | 5 | 89 | 11.97 | 2.49 | 12 | 137 | 27.40 | 3.02 | 28 |
| 42 | 4.33 | 1.84 | 6 | 90 | 11.97 | 2.49 | 14 | 138 | 27.40 | 3.02 | 30 |
| 43 | 4.33 | 1.84 | 7 | 91 | 11.97 | 2.49 | 16 | 139 | 27.40 | 3.02 | 32 |
| 44 | 4.33 | 1.84 | 8 | 92 | 11.97 | 2.49 | 18 | 140 | 27.40 | 3.02 | 34 |
| 45 | 2.38 | 2.16 | 3 | 93 | 4.65 | 3.44 | 5 | 141 | 7.57 | 5.10 | 8 |
| 46 | 2.38 | 2.16 | 4 | 94 | 4.65 | 3.44 | 6 | 142 | 7.57 | 5.10 | 9 |
| 47 | 2.38 | 2.16 | 5 | 95 | 4.65 | 3.44 | 7 | 143 | 7.57 | 5.10 | 10 |
| 48 | 2.38 | 2.16 | 6 | 96 | 4.65 | 3.44 | 8 | 144 | 7.57 | 5.10 | 11 |

**Table A5 Number of Arrivals in Each 15-Minute Interval**

| Interval | No. arrivals | Interval | No. arrivals | Interval | No. rarrivals | Interval | No. arrivals |
|---|---|---|---|---|---|---|---|
| 0 – 0.25 | 1 | 10.5 – 10.75 | 2 | 20.75 – 21 | 2 | 31 – 31.25 | 1 |
| 0.25 – 0.5 | 4 | 10.75 – 11 | 1 | 21 – 21.25 | 3 | 31.25 – 31.5 | 2 |
| 0.5 – 0.75 | 4 | 11 – 11.25 | 2 | 21.25 – 21.5 | 5 | 31.5 – 31.75 | 2 |
| 0.75 – 1 | 2 | 11.25 – 11.5 | 3 | 21.5 – 21.75 | 0 | 31.75 – 32 | 1 |
| 1 – 1.25 | 1 | 11.5 – 11.75 | 3 | 21.75 – 22 | 2 | 32 – 32.25 | 2 |
| 1.25 – 1.5 | 4 | 11.75 – 12 | 6 | 22 – 22.25 | 4 | 32.25 – 32.5 | 2 |
| 1.5 – 1.75 | 2 | 12 – 12.25 | 1 | 22.25 – 22.5 | 2 | 32.5 – 32.75 | 4 |
| 1.75 – 2 | 2 | 12.25 – 12.5 | 3 | 22.5 – 22.75 | 5 | 32.75 – 33 | 3 |
| 2 – 2.25 | 2 | 12.5 – 12.75 | 1 | 22.75 – 23 | 1 | 33 – 33.25 | 3 |
| 2.25 – 2.5 | 4 | 12.75 – 13 | 1 | 23 – 23.25 | 2 | 33.25 – 33.5 | 3 |
| 2.5 – 2.75 | 2 | 13 – 13.25 | 3 | 23.25 – 23.5 | 2 | 33.5 – 33.75 | 4 |
| 2.75 – 3 | 5 | 13.25 – 13.5 | 5 | 23.5 – 23.75 | 1 | 33.75 – 34 | 1 |
| 3 – 3.25 | 2 | 13.5 – 13.75 | 3 | 23.75 – 24 | 2 | 34 – 34.25 | 3 |
| 3.25 – 3.5 | 2 | 13.75 – 14 | 3 | 24 – 24.25 | 4 | 34.25 – 34.5 | 0 |
| 3.5 – 3.75 | 4 | 14 – 14.25 | 3 | 24.25 – 24.5 | 4 | 34.5 – 34.75 | 3 |
| 3.75 – 4 | 2 | 14.25 – 14.5 | 1 | 24.5 – 24.75 | 2 | 34.75 – 35 | 2 |
| 4 – 4.25 | 2 | 14.5 – 14.75 | 2 | 24.75 – 25 | 4 | 35 – 35.25 | 5 |
| 4.25 – 4.5 | 5 | 14.75 – 15 | 5 | 25 – 25.25 | 2 | 35.25 – 35.5 | 4 |
| 4.5 – 4.75 | 3 | 15 – 15.25 | 1 | 25.25 – 25.5 | 2 | 35.5 – 35.75 | 1 |
| 4.75 – 5 | 3 | 15.25 – 15.5 | 2 | 25.5 – 25.75 | 2 | 35.75 – 36 | 1 |
| 5 – 5.25 | 6 | 15.5 – 15.75 | 1 | 25.75 – 26 | 4 | 36 – 36.25 | 1 |
| 5.25 – 5.5 | 2 | 15.75 – 16 | 2 | 26 – 26.25 | 1 | 36.25 – 36.5 | 1 |
| 5.5 – 5.75 | 1 | 16 – 16.25 | 2 | 26.25 – 26.5 | 4 | 36.5 – 36.75 | 3 |
| 5.75 – 6 | 4 | 16.25 – 16.5 | 1 | 26.5 – 26.75 | 1 | 36.75 – 37 | 3 |
| 6 – 6.25 | 3 | 16.5 – 16.75 | 2 | 26.75 – 27 | 1 | 37 – 37.25 | 2 |
| 6.25 – 6.5 | 4 | 16.75 – 17 | 2 | 27 – 27.25 | 4 | 37.25 – 37.5 | 1 |
| 6.5 – 6.75 | 4 | 17 – 17.25 | 4 | 27.25 – 27.5 | 5 | 37.5 – 37.75 | 2 |
| 6.75 – 7 | 6 | 17.25 – 17.5 | 3 | 27.5 – 27.75 | 1 | 37.75 – 38 | 2 |
| 7 – 7.25 | 2 | 17.5 – 17.75 | 3 | 27.75 – 28 | 3 | 38 – 38.25 | 3 |
| 7.25 – 7.5 | 2 | 17.75 – 18 | 1 | 28 – 28.25 | 4 | 38.25 – 38.5 | 2 |
| 7.5 – 7.75 | 2 | 18 – 18.25 | 0 | 28.25 – 28.5 | 5 | 38.5 – 38.75 | 2 |
| 7.75 – 8 | 1 | 18.25 – 18.5 | 4 | 28.5 – 28.75 | 4 | 38.75 – 39 | 1 |
| 8 – 8.25 | 5 | 18.5 – 18.75 | 1 | 28.75 – 29 | 3 | 39 – 39.25 | 0 |
| 8.25 – 8.5 | 2 | 18.75 – 19 | 3 | 29 – 29.25 | 7 | 39.25 – 39.5 | 0 |
| 8.5 – 8.75 | 2 | 19 – 19.25 | 2 | 29.25 – 29.5 | 2 | 39.5 – 39.75 | 3 |
| 8.75 – 9 | 3 | 19.25 – 19.5 | 4 | 29.5 – 29.75 | 1 | 39.75 – 40 | 0 |
| 9 – 9.25 | 3 | 19.5 – 19.75 | 0 | 29.75 – 30 | 1 | 40 – 40.25 | 1 |
| 9.25 – 9.5 | 0 | 19.75 – 20 | 3 | 30 – 30.25 | 2 | 40.25 – 40.5 | 3 |
| 9.5 – 9.75 | 0 | 20 – 20.25 | 0 | 30.25 – 30.5 | 3 | 40.5 – 40.75 | 3 |
| 9.75 – 10 | 1 | 20.25 – 20.5 | 3 | 30.5 – 30.75 | 2 | 40.75 – 41 | 1 |
| 10 – 10.25 | 0 | 20.5 – 20.75 | 1 | 30.75 – 31 | 1 | 41 – 41.25 | 4 |
| 10.25 – 10.5 | 1 | | | | | | |

## References

Akan, M., B. Ata, M. A. Lariviere. 2011. Asymmetric information and economies of scale in service contracting. *Manuf. Serv. Oper. Manag.* **13**(1): 58–72.

Akşin, Z., M. Armony, V. Mehrotra. 2007. The modern call center: A multi-disciplinary perspective on operations management research. *Prod. Oper. Manag.* **16**(6): 665–688.

Akşin, Z., F. de Véricourt, F. Karaesmen. 2008. Call center outsourcing contract analysis and choice. *Manage. Sci.*, **54**(2): 354–368.

Araghi, A., B. Balcioglu. 2008. A new renewal approximation for certain autocorrelated processes. *Oper. Res. Lett.* **36**: 114–161.

Avramidis, A. N., A. Deslauriers, P. Lecuyer. 2004. Modeling daily arrivals to a telephone call center. *Manage. Sci.* **50**(7): 896–908.

Baron, O., J. Milner. 2009. Staffing to maximize profit for call centers with alternate service level agreements. *Oper. Res.* **57**(3): 685–700.

Borst, S., A. Mandelbaum, M. Reiman. 2004. Dimensioning large call centers. *Oper. Res.* **52**(1): 17–34.

Brown, L., N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, L. Zhao. 2005. Statistical analysis of a telephone call center: A queueing science perspective. *J. Am. Statist. Assoc.* **100**(469): 36–50.

Chen, B. S., G. Henderson. 2001. Two issues in setting call center staffing levels. *Ann. Oper. Res.* **108**(1-4): 175–192.

Chevalier, P., N. Tabordon. 2003. Overflow analysis and cross-trained servers. *Int. J. Prod. Econ.* **85**(1): 47–60.

Chevalier, P., J.-C. Van den Schrieck. 2008. Optimizing the staffing and routing of small-size hierarchical call centers. *Prod. Oper. Manag.* **17**(3): 306–319.

Chevalier, P., J.-C. Van den Schrieck, Y. Wei. 2010. Measuring the variability in supply chains with the peakedness. Technical Report 67, CORE Discussion Paper.

Dawson, K. 2007. ICMI's contact center outsourcing report: Key findings. *Call Center Magazine*. Available at http://www.call-centermagazine.com/shared/article/showArticle jhtml?articleId=201805251.

Erlang, A. K. 1948. On the rational determination of the number of circuits. E. Brockmeyer, H. L. Halstrom, A. Jensen, eds. *The Life and Works of A. K. Erlang*. The Copenhagen Telephone Company.

Fredericks, A. A. 1980. Congestion in blocking systems: A simple approximation technique. *Bell Syst. Tech. J.*, **59**(6): 805–827.

Gans, N., G. Koole, A. Mandelbaum. 2003. Telephone Call centers: tutorial, review and research prospects. *Manuf. Serv. Oper. Manag.* **5**(2): 79–141.

Gans, N., Y.-P. Zhou. 2007. Call-routing schemes for call-center outsourcing. *Manuf. Serv. Oper. Manag.* **9**(1): 33–50.

Garnett, O., A. Mandelbaum, M. Reiman. 2002. Designing a call center with impatient customers. *Manuf. Serv. Oper. Manag.* **4**(3): 208–227.

Halfin, S., W. Whitt. 1981. Heavy-traffic limits for queues with many exponential servers. *Oper. Res.* **29**(3): 567–588.

Hasija, S., E. J. Pinker, R. A. Shumsky. 2008. Call center outsourcing contracts under information asymmetry. *Manage. Sci.* **54**(4): 793–807.

Jagerman, D., B. Balcioglu, T. Altiok, B. Melamed. 2004. Mean waiting time approximations in the G/G/1 Queue. *Queue. Syst.* **46**(3-4): 481–506.

Jagerman, D. L. 1974. Some properties of the Erlang loss function. *Bell Syst. Tech. J.*, **53**(3): 525–551.

Jagerman, D. L., B. Melamed, W. Willinger. 1997. Stochastic modeling of traffic processes. Technical Report 7. Rutgers Center for Operations Research (RUTCOR).

Keblis, M., M. Chen. 2006. Improving customer service operations at amazon.com. *Interfaces* **36**(5): 433–445.

Kleinrock, L. 1975. *Queueing Systems; Volume 1: Theory*. Wiley, New York.

Kosten, L. 1973. *Stochastic Theory of Service Systems*. Pergamon, Oxford.

Kuczura, A. 1973. The interrupted Poisson process as an overflow process. *Bell Syst. Tech. J.* **52**(3): 437–448.

Macq, J.-F. 2005. *Optimization of Multimedia Flows over Data Networks: The Core Location Problem and the Peakedness Characterization*. PhD thesis, Université catholique de Louvain - Faculté des Sciences Appliquées.

Milner, J. M., T. L. Olsen. 2008. Service-level agreements in call centers: Perils and prescriptions. *Manage. Sci.* **54**(2): 238–252.

Newell, G. F. 1973. *Approximate Stochastic Behavior of n-Server Service Systems with Large n*. Lecture Notes in Economics and Mathematical Systems, volume 87. Springer-Verlag, Berlin.

Pot, A., S. Bhulai, G. Koole. 2008. A simple staffing method for multi-skill call centers. *Manuf. Serv. Oper. Manag.* **10**(3): 421–428.

Ren, Z. J., F. Zhang. 2009. Service Outsourcing: Capacity, Quality and Correlated Costs. Technical report, Boston University.

Ren, Z. J., Y.-P. Zhou. 2008. Call center outsourcing: Coordinating staffing level and service quality. *Manage. Sci.* **54**(2): 369–383.

Steckley, S. G., S. G. Henderson, V. Mehrotra. 2005. Performance measures for service systems with a random arrival rate. M. E. Kuhl, N. M. Steiger, F. B. Armstrong, J. A. Joines, eds. *Proceedings of the 2005 Winter Simulation Conference*. IEEE, Piscataway, NJ.

Tabordon, N. 2002. *Modeling and Optimizing the Management of Operator Training in a Call Center*. PhD thesis, Université catholique de Louvain - Institut d'Administration et de Gestion.

Van den Schrieck, J.-C. 2010. *Multi-skill Queueing Models for Call Centers: Approximations and Performance Optimization*. PhD thesis, Louvain School of Management - Center for Operations Research and Econometrics.

Whitt, W. 1982. Approximating a point process by a renewal process: I. Two basic methods. *Oper. Res.* **30**(1): 125–147.

Whitt, W. 1992. Understanding the efficiency of multi-server service systems. *Manage. Sci.* **38**(5): 708–723.

Whitt, W. 1993. Approximations for the GI/G/m Queue. *Prod. Oper. Manag.* **2**(2): 114–161.

Wilkinson, R. 1956. Theories for toll traffic engineering in the U.S.A. *Bell Syst. Tech. J.* **35**(2): 421–514.

Wolff, R. W. 1989. *Stochastic Modeling and the Theory of Queues*. Prentice Hall, Englewood Cliffs, NJ.

## CORE Reprints -- Recent titles

2525. Dominik DORSCH, Hubertus Th. JONGEN and Vladimir SHIKHMAN. On structure and computation of generalized Nash equilibria. *SIAM Journal on Optimization,* 23(1), 452-474, 2013.
2526. Luc BAUWENS, Christian M. HAFNER and Diane PIERRET. Multivariate volatility modeling of electricity futures. *Journal of Applied Econometrics,* 28(5), 743-761, 2013.
2527. Jean-François CARPANTIER and Besik SAMKHARADZE. The asymmetric commodity inventory effect on the optimal hedge ratio. *The Journal of Futures Markets,* 33(9), 868-888, 2013.
2528. Elena DEL REY and Miguel-Angel LOPEZ-GARCIA. On welfare criteria and optimality in an endogenous growth model. *Journal of Public Economic Theory,* 14(6), 927-943, 2012.
2529. Adel HATAMI-MARBINI, Per J. AGRELL, Madjid TAVANA and Ali EMROUZNEJAD. A stepwise fuzzy linear programming model with possibility and necessity relations. *Journal of Intelligent & Fuzzy Systems,* 25(1), 81-93, 2013.
2530. Adel HATAMI-MARBINI, Madjid TAVANA, Saber SAATI and Fatemeh KANGI. A fuzzy group linear programming technique for multidimentional analysis of preference. *Journal of Intelligent & Fuzzy Systems,* 25(3), 723-735, 2013.
2531. Madjid TAVANA, Rashed KHANJANI SHIRAZ, Adel HATAMI-MARBINI, Per J. AGRELL and Khalil PARYAB. Chance-constrained DEA models with random fuzzy inputs and outputs. *Knowledge-based systems,* 52, 32-52, 2013.
2532. Adel HATAMI-MARBINI, Madjid TAVANA, Vahid HAJIPOUR, Fatemeh KANGI and Abolfazl KAZEMI. An extended compromise ratio method for fuzzy group multi-attribute decision making with SWOT analysis. *Applied Soft Computing,* 13(8), 3459-3472, 2013.
2533. Luc BAUWENS, Arnaud DUFAYS and Jeroen V.K. ROMBOUTS. Marginal likelihood for Markov-switching and change-point GARCH models. *Journal of Econometrics,* 178 (Part 3), 508-522, 2014.
2534. Manuel FÖRSTER, Michel GRABISCH and Agnieszka RUSINOWSKA. Anonymous social influence. *Games and Economic Behavior,* 82, 621-635, 2013.
2535. Grégory VANDENBULCKE, Isabelle THOMAS and Luc INT PANIS. Predicting cycling accident risk in Brussels: a spatial case-control approach. *Accident Analysis and Prevention,* 62, 341-357, 2014.
2536. Laurent VAN MALDEREN, Bart JOURQUIN, Claude PECHEUX, Isabelle THOMAS, Elien VAN DE VIJVER, Thomas VANOUTRIVE, Ann VERHETSEL and Frank WITLOX. Exploring the profession of mobility manager in Belgium and their impact on commuting. *Transportation Research A,* 55, 46-55, 2013.
2537. Ana MAULEON, Vincent VANNETELBOSCH and Cecilia VERGARI. Bargaining and delay in patent licensing. *International Journal of Economic Theory,* 9(4), 279-302, 2013.
2538. Mikael PETITJEAN. Bank failures and regulation: a critical review. *Journal of Financial Regulation and Compliance,* 21(1), 16-38, 2013.
2539. Jean-François CAULIER, Ana MAULEON, Jose J. SEMPERE-MONERRIS and Vincent VANNETELBSOCH. Stable and efficient coalitional networks. *Review of Economic Design,* 17(4), 249-271, 2013.
2540. Michel LE BRETON, Juan D. MORENO-TERNERO, Alexei SAVVATEEV and Shlomo WEBER. Stability and fairness in models with a multiple membership. *International Journal of Game Theory,* 42(3), 673-694, 2013.
2541. Jens Leth Hougaard, Juan D. MORENO-TERNERO and Lars Peter ØSTERDAL. Rationing with baselines the composition extension operator. *Annals of Operations Research,* 211(1), 179-191, 2013.
2542. Jens Leth Hougaard, Juan D. MORENO-TERNERO and Lars Peter ØSTERDAL. On the measurement of the (multidimensional) inequality of health distributions. In *Health and Inequality,* Research on Economic Inequality, Vol. 21, 111-129, 2013.
2543. Ali EMROUZNEJAD, Madjid TAVANA and Adel HATAMI-MARBINI. The state of the art in fuzzy data envelopment analysis. Chapter 1 in *Performance Measurement with Fuzzy Data Development Analysis,* Studies in Fuzziness and Soft Computing, 309, Springer-Verlag, 1-45, 2014.
2544. Marie-Louise LEROUX and Grégory PONTHIERE. Utilitarianism and unequal longevities: a remedy? *Economic Modelling,* 30, 888-899, 2013.
2545. Marie-Louise LEROUX and Pierre PESTIEAU. Social security and family support. *Canadian Journal of Economics/Revue canadienne d'Economique,* 47(1), 115-143, 2014.

2546. Paul BELLEFLAMME and Martin PEITZ. Asymmetric information and overinvestment in quality. *European Economic Review,* 66, 127-143, 2014.
2547. Alexander OSHARIN, Jacques-François THISSE, Philip USHCHEV and Valery VERBUS. Monopolistic competition and income dispersion. *Economics Letters,* 122(2), 348-352, 2014.
2548. Sudipto BHATTACHARYA, Claude d'ASPREMONT, Sergei GURIEV, Debapriya SEN and Yair TAUMANN. Cooperation in R&D: patenting, licensing, and contracting. Chapter 10 in K. Chatterjee and W. Samuelson (eds.), *Game Theory and Business Applications,* International Series in Operations Research & Management Science, 194, 265-286, 2014.
2549. Jean-François MERTENS and Anna RUBINCHIK. Essential properties of $L_{p,q}$ spaces (the amalgams) and the implicit function theorem for equilibrium analysis in continuous time. *Journal of Mathematical Economics,* 50, 187-196, 2014.
2550. Andreas EHRENMANN and Yves SMEERS. Risk adjusted discounted cash flows in capacity expansion models. *Mathematical Programming Series B,* 140(2), 267-293, 2013.
2551. Gauthier de MAERE d'AERTRYCKE and Yves SMEERS. Liquidity risks on power exchanges: a generalized Nash equilibrium model. *Mathematical Programming Series B,* 140(2), 381-414, 2013.
2552. Gauthier de MAERE d'AERTRYCKE, Alexander Shapiro and Yves SMEERS. Risk exposure and Lagrange multipliers of nonanticipativity constraints in multistage stochastic problems. *Mathematical Methods of Operations Research,* 77(3), 393-405, 2013.
2553. Giorgia OGGIONI and Yves SMEERS. Market failures of market coupling and counter-trading in Europe: an illustrative based discussion. *Energy Economics,* 35, 74-87, 2013.
2554. Giorgia OGGIONI, Frederic H. MURPHY and Yves SMEERS. Evaluating the impacts of priority dispatch in the European electricity market. *Energy Economics,* 42, 183-200, 2014.
2555. Per J. AGRELL and Adel HATAMI-MARBINI. Frontier-based performance analysis models for supply chain management: state of the art and research directions. *Computers & Industrial Engineering,* 66(3), 567-583, 2013.
2556. Amir H. SHOKOUHI, Hamid SHAHIRIARI, Per J. AGRELL and Adel HATAMI-MARBINI. Consistent and robust ranking in imprecise data envelopment analysis under perturbations of random subsets of data. *OR Spectrum,* 36(1), 133-160, 2014.
2557. Per J. AGRELL and Pooria NIKNAZAR. Structural and behavioral robustness in applied best practice regulation. *Socio-Economic Planning Sciences,* 48(1), 89-103, 2014.
2558. Claire DUJARDIN, Vincent LORANT and Isabelle THOMAS. Self-assessed health of elderly people in Brussels: Does the built environment matter? *Health & Place,* 27, 59-67, 2014.
2559. Pierre M. PICARD and Takatoshi TABUCHI. On microfoundations of the city. *Journal of Economic Theory,* 148, 2561-2582, 2013.
2560. Vladyslav NORA and Hiroshi UNO. Saddle functions and robust sets of equilibria. *Journal of Economic Theory,* 150, 866-877, 2014.
2561. Mattew O. JACKSON and Dunia LOPEZ-PINTADO. Diffusion and contagion in network heterogeneous agents and homophily. *Network Science,* 1(1), 49-67, 2013.
2562. Dunia LOPEZ-PINTADO. Public goods in directed networks. *Economics Letters,* 121(2), 160-162, 2013.
2563. Françoise FORGES. J.-F. Mertens (1946-2012): des jeux répétés à l'analyse "coût-bénéfice". *Revue Française d'Economie,* 3(XXVII), 19-34, 2013.
2564. Françoise FORGES. MD interview: interview with Jean-François Mertens (1946-2012). *Macroeconomic Dynamics,* 2013.
2565. Adel HATAMI-MARBINI, Ali EMROUZNEJAD and Per J. AGRELL. Interval data without sign restrictions in DEA. *Applied Mathematical Modelling* 38(7-8), 2028-2036, 2014.
2566. Tao WANG, François GLINEUR, Jérôme LOUVEAUX and Luc VANDENDORPE. Weighted sum rate maximization for downlink OFDMA with subcarrier-pair based opportunistic DF relaying. *IEEE Transactions on Signal Processing,* 61(10), 2512-2524, 2013.
2567. Nicolas GILLIS and François GLINEUR. A continuous characterization of the maximum-edge biclique problem. *Journal of Global Optimization,* 58(3), 439-464, 2014.
2568. Thibaut CALIMAN, Catherine D'HONDT and Mikael PETITJEAN. Determining an optimal multiplier in dynamic core-satellite strategies. *Journal of Asset Management,* 14(4), 210-227, 2013.
2569. Jean-Christophe VAN den SCHRIECK, Zeynep AKŞIN and Philippe CHEVALIER. Peakedness-based staffing for call center outsourcing. *Production and Operations Management,* 23(3), 504-524, 2014.