Economic Research-Ekonomska Istraživanja

ISSN: 1331-677X (Print) 1848-9664 (Online) Journal homepage: https://www.tandfonline.com/loi/rero20

Employee business at various levels of a hierarchy for organisations completing case work

Miao-Sheng Chen, Jing Chung & Po-Yu Chen

To cite this article: Miao-Sheng Chen, Jing Chung & Po-Yu Chen (2019) Employee business at various levels of a hierarchy for organisations completing case work, Economic Research-Ekonomska Istraživanja, 32:1, 2818-2828, DOI: 10.1080/1331677X.2019.1653210

To link to this article: https://doi.org/10.1080/1331677X.2019.1653210

Published online: 29 Aug 2019.

ECONOMIC RESEARCH-EKONOMSKA ISTRAŽIVANJA
2019, VOL. 32, NO. 1, 2818–2828
https://doi.org/10.1080/1331677X.2019.1653210

Employee business at various levels of a hierarchy for organisations completing case work

Miao-Sheng Chen[a], Jing Chung[b] and Po-Yu Chen[c]

[a]Department of Business Administration, Nanhua University, Dalin, Taiwan (R.O.C.); [b]Department of Mathematics and Information Education, National Taipei University of Education, Taipei, Taiwan (R.O.C.); [c]Department of Advertising and Strategic Marketing, Ming Chuan University, Taipei, Taiwan (R.O.C.)

ABSTRACT
This article describes a model for examining the contribution of supervisors to an organization by considering the case work they complete as a production system. The average delay in case work is referred to as the service level. At a given service level, the minimization of total wages within one hour can be studied as a cost function. With this cost function, wage spending on handled-case time and idle time can be formulated. The ratio between the handled-case time and idle time of all employees at the kth level within 1 hour is defined as the 'busy index' at the kth level. From the optimal hierarchical structure, we find the following two properties: (1) Given any two levels i and j, the ratio between the idle times of ith and jth levels is independent not only of the service level but also the rate of arriving cases; and (2) At each level, the busy indices are proportional to the square root of each level's wage rates. This implies that the busy indices decrease with the hierarchical level. Ultimately, when the wage rates at all levels are equal, the increment also becomes equal.

ARTICLE HISTORY
Received 19 May 2017
Accepted 21 January 2019

KEYWORDS
organisation design; cost function; idle time; hierarchy organisation

JEL CLASSIFICATION
L25

1. Introduction

For managing service, production and supply chain systems, understanding queueing systems is critical. Tasks or scheduling problems such as customer services involving tangible and intangible queues or crowds, loadings, case management time, procedural durations, and delivery plans are research topics related to queueing theory.

The basic concept of queuing theory originated in the queuing model proposed by A. K. Erlang in the early twentieth century for resolving traffic congestion problems in an automatic telephone system (Brockmeyer, Halstrom, & Jensen, 1948). In the early 1950s, D. G. Kendall developed a discrete-time Markov chain model and a queuing-pattern categorization method, thereby providing a theoretical foundation for queuing theory. Since the 1970s, investigating queuing networks and identifying

CONTACT Po-Yu Chen chenboy@mail.mcu.edu.tw

asymptotic solutions to complex queueing problems have become the new trend for studying modern queueing theory (Delasay, Ingolfsson, & Kolfal, 2016). Scudder and Hill (1998) and Gupta, Verma, and Victorino (2006) have indicated that the number of empirical studies on queueing systems increased considerably in the 1990s. According to previous studies, queueing theory has been studied from the following aspects: (1) solving problems related to statistical inference and establishing models based on data; (2) investigating the distribution of the probabilities of queueing-related quantitative indices; and (3) studying topics regarding the optimisation of queueing systems (i.e., how to correctly design and effectively implement various service systems and to optimise their effectiveness). Relevant research results can be widely applied to various random service systems, such as telecommunications, traffic engineering, production, transportation, and inventory systems (Dowdy, Almeida, & Menasce, 2004) as well as service designs for factories, stores, offices, and hospitals (Chen & Chung, 1990; Kesavan, Staats, & Gilland, 2014; Mayhew & Smith, 2008; Shah, Ball, & Netessine, 2017; Song, Tucker, & Murrell, 2015; Tan & Netessine, 2014; Tarny & Chen, 1988; Whitt, 2006).

Analysis of labor productivity or service time (service level) is a critical component of business operations. Previous studies have adopted mathematical models to discuss the relationship among variables related to organisational structure. They have observed the functions of organizations on the basis of the input-output concept and have used mathematical models to discuss more concretely some problems in an organization. For example, Beckmann (1977, 1982), Malone (1987), and Chen and Chung (1990), and Thompson (1992) have established production functions for organizations executing projects. Keren and Levhari (1979) and Tarny and Chen (1988) have studied the optimum span of control in organizational hierarchies. Whitt (2006) noted the essentialness of allocating appropriate levels of manpower to maximize work efficiency, hence the employment of optimization tools by organizations in workforce allocation. Personnel policies have also attracted widespread interest in related research, whereby optimization algorithms have been used to design the ideal personnel strategies for organizations (Tan & Netessine, 2014). Tiffin and Kunc (2011) suggested improving the microlevel problems of organizations to resolve the macrolevel problems of personnel allocation. Bendoly, Swink, and Simpson (2014) focused on the efficient organizational management of multiple projects.

In practice, employee service time is affected by their workload (Kc and Terwiesch, 2009). However, for businesses, employee service time, work quality, and service quality are equally important. Therefore, businesses face the trade-off between employee service time versus work and service quality (Tan & Netessine, 2014). Previous studies have used the variance of employees' average work time to represent their service standards (Berrio, Ospina, & Martnez, 2014; Gans, Koole, & Mandelbaum, 2003); thus, in the present study, employees' average amount of work per hour was adopted as an indicator of service standards. This study applied queueing theory in the context of operations research to specifically present the production and cost functions of an organisation implementing case work.

2. Model

The research method used in this study involved the modification and addition of assumption conditions in the queuing theory; specifically, the conditions were modified to imitate the concepts behind the 'production function' and 'cost function' in micro-economics. This method was used to establish the production and cost functions for the 'optimal hierarchical structure of the case work implementation organization'.

2.1. Symbols

Hierarchy parameters

H: organization level, with Hth representing the top level of the organization.
s: average case completion rate, defined as the service level of an organisation.
λ: average number of cases received in the organization per unit time, representing the average case arrival rate.
w_k: wage rate of an employee at the kth level (wage per unit of time).
μ_k: number of cases completed by an employee at the kth level, $k = 1, 2, \ldots, H$.
θ_k: average proportion of a kth case relative to all cases in the inventory, $k = 1, 2, \ldots, H$.

Decision variable

x_k: total work time of all employees at the kth level per unit time, $k = 1, 2, \ldots, H$.

2.2. Assumptions

This article considers the case work that supervisors complete within an organization as a production system. For the inputs, employees at different levels of the hierarchy are treated as different production factors. This means that an organization with H levels of employees will have H unique production factors. Concurrently, the term x_k is used to denote the total work time of all the kth level's employees within 1 hour, and this is considered the quantity of the kth production factor used. For the outputs, we assume that there are H types of work to be handled by the organization. Of these, the first type of work can be viewed as products of a single-level production process. This means that the work will be completed after being handled by a first-level employee. The second type of work can be viewed as products of a dual-level production process. This means that the work is first handled by a first-level employee and then passed to a second-level employee for completion. Normally, the kth type of work can be viewed as the product of an kth level production process, and it is completed once it is served by a kth level's employee. If θ_k, $\sum_{k=1}^{H} \theta_k = 1$ indicates the percentage of kth type of work and $T_k = T_k(x_1, x_2, \ldots, x_H)$ indicates the expected time that a kth type of work spends in the hierarchy, then $T = \sum_{k=1}^{H} \theta_k T_k$ is the expected time that case work will spend in the hierarchy. To compare the production efficiencies between different organizations and to have a common basis for comparison, we define the service level s of a hierarchy as:

$$s = \frac{T(\infty, \infty, \ldots, \infty)}{T(x_1, x_2, \ldots, x_H)},$$

and use it to indicate the production quantity of the organization.

The organization populates a given level of the hierarchy with employees with identical skill sets; and each employee has a well-defined area of competence, determined by technical criteria that are easy to verify. We consider the first level of the hierarchy as a queueing system (regarding case work and first-level employees as customers and servers, respectively) and assume that this queueing system has Poisson input process with a rate λ.

If all arriving case works from the kth type to the Hth type are assigned without delay to the kth level's employees, then the expected number of those case work received by a kth level's employee within one hour is $p_k \cdot \frac{\lambda}{x_k}$, where $p_k = \sum_{j=k}^{H} \theta_j$, and $1 = p_1 < p_2 < \cdots < p_H$. An elementary result of queueing theory (Gross, Shortle, Thompson, & Harris, 2008) yields that the expected time of a case work spent in kth level is:

$$\frac{1}{\mu_k - \frac{p_k \lambda}{x_k}}$$

if the completion time of a kth level's employee is exponentially distributed with a mean of $\frac{1}{\mu_k}$ and $\mu_k > \frac{p_k \lambda}{x_k}$. Therefore, T_k, the expected time that a kth type of work spends in the hierarchy, is given by:

$$T_k = \sum_{j=1}^{k} \frac{1}{\mu_j - \frac{p_j \lambda}{x_j}} \tag{1}$$

This implies that T, the expected time that case work spends in the hierarchy, is given by:

$$\begin{aligned} T &= \sum_{k=1}^{H} \theta_k T_k; \quad \text{by using (1)} \\ &= \sum_{k=1}^{H} \theta_k \left(\sum_{j=1}^{k} \frac{1}{\mu_j - \dfrac{p_j \lambda}{x_j}} \right) = \sum_{j=1}^{H} \left(\sum_{k=j}^{H} \theta_k \right) \left(\frac{1}{\mu_j - \dfrac{p_j \lambda}{x_j}} \right) \\ &= \sum_{j=1}^{H} \frac{1}{\dfrac{\mu_j}{p_j} - \dfrac{\lambda}{x_j}} \end{aligned} \tag{2}$$

if $\mu_j > \frac{p_j \lambda}{x_j}, \forall j = 1, 2, \ldots, H$

At a given service level s, the minimization of total wages within one hour can be studied as the cost function $C(s)$. With this cost function, the busyness of employees at various levels can be concretely discussed.

2.3. Mathematical model

According to (3), it is valid that the service level $s = s(x_1, x_2, \ldots, x_h)$ is given by:

$$s = \frac{T(\infty, \infty, \ldots, \infty)}{T(x_1, x_2, \ldots, x_h)}; \quad \text{by using (2)}$$
$$= \frac{\sum_{k=1}^{H} \frac{p_k}{\mu_k}}{\sum_{k=1}^{H} \frac{1}{\frac{\mu_k}{p_k} - \frac{\lambda}{x_k}}} \tag{3}$$

Let w_k be the wage of a kth level employee working 1 hour, then $\sum_{k=1}^{H} w_k x_k$ is the hourly wage of all employees in the entire hierarchy. Given a service level s, the minimization of total wages is referred as the cost function $C = C(s)$. This means that $C(s)$ is the minimized objective value of the following problem:

$$\begin{cases} \min \sum_{k=1}^{H} w_k x_k \\ \text{s.t.} \left[\sum_{k=1}^{H} \frac{p_k}{\mu_k}\right] \left[\sum_{k=1}^{H} \left(\frac{\mu_k}{p_k} - \frac{\lambda}{x_k}\right)^{-1}\right]^{-1} = s \end{cases} \tag{4}$$

3. Optimal solution

The Lagrangian of (4) with a multiplier v is:

$$L = \sum_{k=1}^{H} w_k x_k + v \left\{ s - \left[\sum_{k=1}^{H} \frac{p_k}{\mu_k}\right] \left[\sum_{k=1}^{H} \left(\frac{\mu_k}{p_k} - \frac{\lambda}{x_k}\right)^{-1}\right]^{-1} \right\}$$

Then, the necessary conditions for optimality are:

$$0 = \frac{\partial L}{\partial x_j} = w_j - v \left[\sum_{k=1}^{H} \frac{p_k}{\mu_k}\right] \left[\sum_{k=1}^{H} \left(\frac{\mu_k}{p_k} - \frac{\lambda}{x_k}\right)^{-1}\right]^{-2} \left[\frac{\mu_j}{p_j} - \frac{\lambda}{x_j}\right]^{-2} x_j^{-2} \lambda, \tag{5}$$
$$\forall j = 1, 2, \ldots, H$$

$$0 = \frac{\partial L}{\partial v} = s - \left[\sum_{k=1}^{H} \frac{p_k}{\mu_k}\right] \left[\sum_{k=1}^{H} \left(\frac{\mu_k}{p_k} - \frac{\lambda}{x_k}\right)^{-1}\right]^{-1} \tag{6}$$

A simple computation from (5) yields the following two properties:

$$\frac{\mu_j x_j}{p_j} - \lambda = w_j^{\frac{-1}{2}} \left[v \lambda \sum_{k=1}^{H} \frac{p_k}{\mu_k}\right]^{\frac{1}{2}} \left[\sum_{k=1}^{H} \left(\frac{\mu_k}{p_k} - \frac{\lambda}{x_k}\right)^{-1}\right]^{-1} \tag{7}$$

and:

$$w_j^{\frac{-1}{2}} x_j = \left[v\lambda \sum_{k=1}^{H} \frac{p_k}{\mu_k} \right]^{\frac{1}{2}} \left[\sum_{k=1}^{H} \left(\frac{\mu_k}{p_k} - \frac{\lambda}{x_k}\right)^{-1} \right]^{-1} \left[\frac{\mu_k}{p_k} - \frac{\lambda}{x_k} \right]^{-1} \tag{8}$$

Summing (8) from $j=1$ to $j=H$, we find that:

$$\sum_{j=1}^{H} w_j^{\frac{1}{2}} x_j = \left(v\lambda \sum_{k=1}^{H} \frac{p_k}{\mu_k} \right)^{\frac{1}{2}} \tag{9}$$

Combining this with (6) and (9) gives:

$$s^{-1}\left(\sum_{k=1}^{H} \frac{p_k}{\mu_k}\right) = \left[\sum_{k=1}^{H} \left(\frac{\mu_k}{p_k} - \frac{\lambda}{x_k}\right)^{-1} \right] \left[v\lambda \sum_{k=1}^{H} \frac{p_k}{\mu_k} \right]^{\frac{-1}{2}} \left[\sum_{j=1}^{H} w_j^{\frac{1}{2}} x_j \right]; \quad \text{by using (7)}$$

$$= \left[\sum_{k=1}^{H} \left(\frac{\mu_k}{p_k} - \frac{\lambda}{x_k}\right)^{-1} \right] \left[v\lambda \sum_{k=1}^{H} \frac{p_k}{\mu_k} \right]^{\frac{-1}{2}}$$

$$\times \left\{ \sum_{k=1}^{H} w_k^{\frac{1}{2}} \frac{p_k}{\mu_k} \left[w_k^{\frac{-1}{2}} \left(v\lambda \sum_{k=1}^{H} \frac{p_k}{\mu_k} \right)^{\frac{-1}{2}} \left(\sum_{k=1}^{H} \left(\frac{\mu_k}{p_k} - \frac{\lambda}{x_k} \right)^{-1} \right)^{-1} + \lambda \right] \right\}$$

$$= \sum_{k=1}^{H} \frac{p_k}{\mu_k} + \left[\sum_{k=1}^{H} \left(\frac{\mu_k}{p_k} - \frac{\lambda}{x_k}\right)^{-1} \right] \left[v\lambda \sum_{k=1}^{H} \frac{p_k}{\mu_k} \right]^{\frac{-1}{2}} \left[\sum_{k=1}^{H} w_k^{\frac{-1}{2}} \frac{p_k}{\mu_k} \right] \lambda \tag{10}$$

A computation from (10) yields:

$$\left[v\lambda \sum_{k=1}^{H} \frac{p_k}{\mu_k} \right]^{\frac{1}{2}} \left[\sum_{k=1}^{H} \left(\frac{\mu_k}{p_k} - \frac{\lambda}{x_k}\right)^{-1} \right]^{-1} = \frac{s}{1-s} \left[\sum_{k=1}^{H} \frac{p_k}{\mu_k} \right]^{-1} \left[\sum_{k=1}^{H} w_k^{\frac{1}{2}} \frac{p_k}{\mu_k} \right] \lambda \tag{11}$$

Here, the optimal solution x_j^* of (4) can be obtained by substituting (11) into (7):

$$x_j^* = \frac{p_j}{\mu_j} \lambda \left[1 + w_j^{\frac{-1}{2}} \cdot \frac{s}{1-s} \left(\sum_{k=1}^{H} \frac{p_k}{\mu_k} \right)^{-1} \left(\sum_{k=1}^{H} w_k^{\frac{1}{2}} \frac{p_k}{\mu_k} \right) \right], \quad \text{for } j = 1, 2, \ldots, H \tag{12}$$

It can be shown that the variable s, determined by (3), is jointly concave in variables $(x_1, x_2, \ldots, x_H)$; thus, the formulation (12) is indeed an optimal solution of the problem Equation (4). Substituting (12) into the cost function $C(s) = \sum_{k=1}^{H} w_k x_k^*$ gives

$$C(s) = \lambda \left[\sum_{k=1}^{H} w_k \frac{p_k}{\mu_k} + \frac{s}{1-s} \left(\sum_{k=1}^{H} \frac{p_k}{\mu_k} \right)^{-1} \left(\sum_{k=1}^{H} w_k^{\frac{1}{2}} \frac{p_k}{\mu_k} \right)^{2} \right] \tag{13}$$

4. Idel time and handle-case time

Because λ is the expected number of cases arriving in the hierarchy within one hour, $p_j = \sum_{k=j}^{H} \theta_k$ is the percentage of those cases passing through the jth level of the hierarchy, and μ_j^{-1} is the (expected) completion time of a case proceeded by a jth level employee; thus, t_j, the handled-case time of all employee at the jth level within 1 hour, is given by:

$$t_j = \lambda \ p_j \mu_j^{-1} \tag{14}$$

and hence I_j, the idle time of all employees at the jth level within 1 hour, is given by:

$$I_j = x_j^* - t_j \tag{15}$$

Combining this with (12) and (14) gives:

$$I_j = \left[\frac{p_j}{\mu_j} \ w_j^{\frac{-1}{2}}\right] \lambda \ \frac{s}{1-s} \left(\sum_{k=1}^{H} \frac{p_k}{\mu_k}\right)^{-1} \left(\sum_{k=1}^{H} w_k^{\frac{1}{2}} \ \frac{p_k}{\mu_k}\right) \tag{16}$$

From (14) and (16), we have:

$$\frac{I_i}{I_j} = \left(\frac{p_i}{\mu_i} \ w_i^{\frac{-1}{2}}\right) / \left(\frac{p_j}{\mu_j} \ w_j^{\frac{-1}{2}}\right) \tag{17}$$

$$= \frac{t_i}{t_j} \left(\frac{w_j}{w_i}\right)^{\frac{1}{2}} \tag{18}$$

and:

$$\frac{t_i}{t_j} \geq \frac{I_i}{I_j} \quad iff \quad w_i \geq w_j \tag{19}$$

If we use $b_k, b_k = t_k / I_k$, to measure the busyness of a kth level employee, then from (18), we have:

$$b_1 : b_2 : \cdots : b_H = w_1^{\frac{1}{2}} : w_2^{\frac{1}{2}} : \cdots : w_H^{\frac{1}{2}} \tag{20}$$

In general, the wage rate of an upper level employee is greater than that of a lower level employee. Thus, the assumption here is:

$$w_H \geq w_{H-1} \geq \cdots \geq w_z \geq w_1 \tag{21}$$

and thus by (20):

$$b_1 \leq b_2 \leq \cdots \leq b_{H-1} \leq b_H \tag{22}$$

Combining this with (13), (14) and (16) gives:

$$C(s) = C_1(s) + C_2(s) \tag{22}$$

where $C_1(s) = \sum_{k=1}^{H} \ w_k t_k$ is the wage spending on employee handle-case time; and $C_2(s) = \sum_{k=1}^{H} \ w_k I_k$ is the wage spent in employee's idle time.

It is well known that for any real numbers $(\alpha_1, \alpha_2, \ldots, \alpha_n)$, $(\beta_1, \beta_2, \ldots, \beta_n)$:

$$\left(\sum_{k=1}^{n} \ \alpha_k \beta_k \right)^2 \leq \left(\sum_{k=1}^{n} \ \alpha_k^2 \right) \left(\sum_{k=1}^{n} \ \beta_k^2 \right) \tag{23}$$

and the equality holds *iff* $(\alpha_1 : \alpha_2 : \cdots : \alpha_n) = (\beta_1 : \beta_2 : \cdots : \beta_n)$

From (16), we have:

$$\begin{aligned}
C_2(s) &= \sum_{k=1}^{H} \ w_k I_k = \frac{\lambda s}{1-s} \left[\sum_{k=1}^{H} \ \frac{p_k}{\mu_k} \right]^{-1} \left[\sum_{k=1}^{H} \left(w_k \ \frac{p_k}{\mu_k} \right)^{\frac{1}{2}} \left(\frac{p_k}{\mu_k} \right)^{\frac{1}{2}} \right]^2; \quad \text{by using (23)} \\
&\leq \frac{\lambda s}{1-s} \left(\sum_{k=1}^{H} \ \frac{p_k}{\mu_k} \right)^{-1} \left(\sum_{k=1}^{H} \ w_k \ \frac{p_k}{\mu_k} \right) \left(\sum_{k=1}^{H} \ \frac{p_k}{\mu_k} \right) \\
&= \frac{\lambda s}{1-s} \left(\sum_{k=1}^{H} \ w_k \ \frac{p_k}{\mu_k} \right); \quad \text{by using (14)} \\
&= \frac{s}{1-s} \ \sum_{k=1}^{H} \ w_k t_k = \frac{s}{1-s} \ C_1(s)
\end{aligned} \tag{24}$$

and:

$$\sum_{k=1}^{H} \ w_k I_k = \frac{\lambda s}{1-s} \ \sum_{k=1}^{H} \ w_k t_k \quad \textit{iff} \quad w_1 = w_2 = \cdots = w_H \tag{25}$$

It is valid from (22), (24) and (25) that:

$$\begin{aligned}
\frac{1}{s} \ C_2(s) = \frac{1}{s} \ \sum_{k=1}^{H} \ w_k I_k \leq C(s) \\
= s \ \left(\frac{1}{s} \ C_2(s)\right) + (1-s)\left(\frac{1}{1-s} \ C_1(s)\right) \leq \frac{1}{1-s} \ C_1(s)
\end{aligned} \tag{26}$$

$$\frac{1}{s} \ \sum_{k=1}^{H} \ w_k I_k = C(s) = \frac{1}{1-s} \ \sum_{k=1}^{H} \ w_k t_k \quad \textit{iff} \quad w_1 = w_2 = \cdots = w_H \tag{27}$$

5. Conclusion

In this interdisciplinary study, the ideas of queuing theory in the field of operations research, along with hierarchical organizational structure, were integrated with the concepts of production and cost functions in microeconomics. This study's major contribution is the translation of the main variable relationships in the problem of organizational hierarchy design, as determined from the case of a case work implementation organization, into organizational hierarchy production and cost functions that can be discussed mathematically.

Given a service level s, we define $C(s)$, the optimal total wage of all employees working within one hour, as the sum of $C_1(s)$ and $C_2(s)$, where $C_1(s)$ denotes wage spending on employee handled-case time and $C_2(s)$ denotes wage spending on employee idle time. The relationship among $C(s)$, $C_1(s)$ and $C_2(s)$ can be formulated as Inequality (26). It follows that $C(s)$ has an upper bound, $(1-s)^{-1} \cdot C_1(s)$, and a lower bound, $s^{-1} \cdot C_2(s)$. When the difference in wage rates at different levels become small, the difference between the upper and lower bounds of $C(s)$ diminishes; ultimately, when the wage rates at all levels are equal, then upper and lower bounds become equal. In particular, when $s = \frac{1}{2}$ and all levels' wage rates are equal, then $C_1\left(\frac{1}{2}\right) = C_2\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)C\left(\frac{1}{2}\right)$. Therefore Inequality (26) is the best possible result in the range of $C(s)$.

Equation (18) indicates that given any two levels i and j, if the quotient of the ith level's wage rate relative to the jth level's wage rate is smaller, or if the quotient of the ith level's handled-case time relative to the jth level's handled-case time is larger, then the quotient of the ith level's idle time relative to the jth level's idle time becomes larger. Equation (17) indicates that given any two levels i and j, the ratio between the idle times of ith and jth levels is independent not only of the service level s but also of the rate of arriving cases λ. Equations (19) and (21) indicate that given an upper level i and lower level j, the rate between the handled-case times of the ith and jth levels is greater than the ratio between the idle times of ith and jth levels. Equation (20) indicates that each level's busy indices are proportional to the square root of each level's wage rates. This implies that busy indices decrease with the hierarchical level. However, when the difference in wage rates between different levels becomes small, the increment of busy indices from one level to the next will diminish; ultimately, when all levels' wage rates are equal, the increment also becomes equal.

The aforementioned cost function can be used only to describe the relationship between the inputs and outputs of single-attribute work. If an organization is required to handle m work tasks with different attributes, and assuming $C_k = C_k(x_1, x_2, \ldots, x^*_{H_k})$ indicates the cost function of the kth-attribute work, then we can use the cost functions $(C_1, C_2, \ldots, C_m)$ to describe the functions of a hierarchy. In this case, employee busyness at different levels can be formulated by computing these cost functions separately and then summing their effects on employee busyness.

Considering the minimisation of total wages, the results of this study addressed relationships between s(organization service level), λ(average case arrival rate), θ_k(proportion of the kth case relative to all cases in the inventory), w_k(wage rate of an employee at the kth level), μ_k(number of cases completed by an employee at the

kth level), and x_{dk}(the optimal total work time of all employees at the kth level per unit time). Observing these relationships enabled exploring the proportions of handled-case time and idle time in the service time of employees at each level. These attributes are useful when an organization is required to adjust the total working hours of individuals at each level of the hierarchy because of environmental changes (e.g., variation of variables such as the case arrival rates, service time of cases, wage rates of employees at different levels, and organization service level). The findings of this study can be applied in the following practical situations: queues for application for passing through customs posts and queues for loan applications (the requirements for a loan application and the position level of the banker responsible for reviewing the application of the loan vary according to the amount of the loan).

In this study, the exogenous variables of the highest level of organizational hierarchy in a case work implementation organization (e.g., wage rates of members at each level of the hierarchy and the case arrival rates) may have varied with changes in the environment (e.g., case arrival rates changed with the organization's off-season and peak-season business cycle). Additionally, the transfer and increase or decrease of personnel at each hierarchy level are often lack the flexibility for adopting instant changes due to established wage systems. These topics are worthy of further discussion in future studies.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Po-Yu Chen http://orcid.org/0000-0001-7769-2752

References

Beckmann, M. J. (1977). Management production functions and the theory of the firm. *Journal of Economic Theory*, *14*(1), 1–18. doi:10.1016/0022-0531(77)90081-3

Beckmann, M. J. (1982). A production function for organizations doing case work. *Management Science*, *28*(10), 1159–1165. doi:10.1287/mnsc.28.10.1159

Bendoly, E., Swink, M., & Simpson, W. P. III. (2014). Prioritizing and monitoring concurrent project work: Effects on switching behavior. *Production and Operations Management*, *23*(5), 847–860. doi:10.1111/poms.12083

Berrio, S. P. R., Ospina, A. F. B., & Martnez, E. J. R. (2014). Organizational design of research results transfer offices: Systematic revision of the literature. *Punto de Vista*, *5*(8), 55–76.

Brockmeyer, E., Halstrom, H. L., & Jensen, A. (1948). *The life and works of A. K. Erlang*. Transactions of the Danish academy of technical sciences. Copenhagen: Copenhagen Telephone.

Chen, M. S., & Chung, J. (1990). Note on the production function for organizations doing case work. *Mathematical Social Sciences*, *19*, 135–141. doi:10.1016/0165-4896(90)90056-D

Delasay, M., Ingolfsson, A., & Kolfal, B. (2016). Modeling load and overwork effects in queueing systems with adaptive service rates. *Operations Research*, *64*(4), 867–885. doi:10.1287/opre.2016.1499

Dowdy, L. W., Almeida, V. A. F., & Menasce, D. A. (2004). *Performance by design: Computer capacity planning by example*. New Jersey: Prentice Hall.

Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing and Service Operations Management*, *5*(2), 79–141.

Gross, D., Shortle, J. F., Thompson, J. M., & Harris, C. M. (2008). *Fundamentals of queueing theory* (4th ed.). New Jersey: John Wiley & Sons.

Kc, D. S., & Terwiesch, C. (2009). Impact of workload on service time and patient safety: An econometric analysis of hospital. *Management Science*, *55*(9), 1486–1498. doi:10.1287/mnsc.1090.1037

Keren, M., & Levhari, D. (1979). The optimum span of control in a pure hierarchy. *Management Science*, *25*(11), 1162–1172. doi:10.1287/mnsc.25.11.1162

Kesavan, S., Staats, B. R., & Gilland, W. (2014). Volume flexibility in services: The costs and benefits of flexible labor resources. *Management Science*, *60*(8), 1884–1906. doi:10.1287/mnsc.2013.1844

Gupta, S., Verma, R., & Victorino, L. (2006). Empirical research published in production and operations management (1992-2005): Trends and future research directions. *Production and Operations Management*, *15*(3), 432–448. doi:10.1111/j.1937-5956.2006.tb00256.x

Malone, T. W. (1987). Modeling coordination in organizations and markets. *Management Science*, *33*(10), 1317–1332. doi:10.1287/mnsc.33.10.1317

Mayhew, L., & Smith, D. (2008). Using queuing theory to analyse the government's 4-h completion time target in accident and emergency departments. *Health Care Management Science*, *11*(1), 11–21. doi:10.1007/s10729-007-9033-8

Scudder, G. D., & Hill, C. A. (1998). A review and classification of empirical research in operations management. *Journal of Operations Management*, *16*(1), 91–101. doi:10.1016/S0272-6963(97)00008-9

Shah, R., Ball, G. P., & Netessine, S. (2017). Plant operations and product recalls in the automotive industry: An empirical investigation. *Management Science*, *63*(8), 2439–2459. doi:10.1287/mnsc.2016.2456

Song, H., Tucker, A. L., & Murrell, K. L. (2015). The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Science*, *61*(12), 3032–3053. doi:10.1287/mnsc.2014.2118

Tan, T. F., & Netessine, S. (2014). When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science*, *60*(6), 1574–1593. doi:10.1287/mnsc.2014.1950

Tarny, M. Y., & Chen, M. S. (1988). Note on the optimum span of control in a pure hierarchy. *European Journal of Operational Research*, *33*(1), 106–113. doi:10.1016/0377-2217(88)90259-7

Thompson, F. M. (1992). Improving the utilization of front-line service delivery system personnel. *Decision Sciences*, *23*(5), 1072–1098. doi:10.1111/j.1540-5915.1992.tb00436.x

Tiffin, S., & Kunc, N. (2011). Measuring the roles universities play in regional innovation systems: A comparative study between Chilean and Canadian natural resource-based regions. *Science and Public Policy*, *38*(1), 55–66. doi:10.3152/016502611X12849792159317

Whitt, W. (2006). The impact of increased employee retention on performance in a customer contact center. *Manufacturing & Service Operations Management*, *8*(3), 235–252. doi:10.1287/msom.1060.0106