“Nursevendor Problem”: Personnel Staffing in the Presence of Endogenous Absenteeism

Linda V. Green
Graduate School of Business, Columbia University, lvg1@columbia.edu

Sergei Savin
The Wharton School, University of Pennsylvania, savin@wharton.upenn.edu

Nicos Savva
London Business School, nsavva@london.edu

The problem of determining nurse staffing levels in a hospital environment is a complex task due to variable patient census levels and uncertain service capacity caused by nurse absenteeism. In this paper, we combine an empirical investigation of the factors affecting nurse absenteeism rates with an analytical treatment of nurse staffing decisions using a novel variant of the newsvendor model. Using data from the emergency department of a large urban hospital, we find that absenteeism rates are correlated with anticipated future nurse workload levels. Using our empirical findings, we analyze a single-period nurse staffing problem considering both the case of constant absenteeism rate (exogenous absenteeism) as well as an absenteeism rate which is a function of the number of scheduled nurses (endogenous absenteeism). We provide characterizations of the optimal staffing levels in both situations and show that the failure to incorporate absenteeism as an endogenous effect results in understaffing.

History: January 6, 2011

1. Introduction and Literature Review

In recent years, hospitals have been faced with ever-increasing pressure from their major payers - federal and state governments, managed care organizations, and large employers - to cut costs. Since nursing personnel accounts for a very large fraction of expenses, the response, in many instances, has been reductions of the nursing staff. Nurse workloads have been further increased by shorter hospital lengths-of-stay (LOS) and increasing use of outpatient procedures, resulting in sicker hospitalized patients who require more nursing care. The adverse impact of these changes has been documented by a number of studies (Needleman et al. (2002), Kovner and Gergen (1998), Aiken et al. (2002), Unruh (2003), Cho et al. (2003)). These effects include increases in medical errors, delays for patients waiting for beds in emergency rooms, and ambulance diversions. In response, a number of state legislatures, e.g. Victoria in Australia (The Victorian Department of Health (2007)) and California in the US (California Department of Health (2004)) have mandated minimum nurse staffing levels.

Establishing the right balance between the quality and cost of patient care is a challenging task. Despite the growing evidence on the influence of nurse staffing levels on the quality of patient care, the question of how such staffing levels should be determined in a particular clinical unit remains largely open. In a hospital environment, nurses are often scheduled to work 8- or 12-hour shifts, and the choice of appropriate nurse staffing levels for a particular shift is complicated by the need to make staffing decisions well in advance (e.g. 6-8 weeks) of that shift when the patient census is unknown. Even if the patient census could be reliably estimated, the prevalence of absenteeism creates a need for the deployment of additional nursing resources beyond what might be considered ideal.

Nurse absenteeism has been widely recognized by practitioners as a serious and growing problem. According to the US Bureau of Labor Statistics, in 2008 nurses exhibited 12.5 incidents of illness or occupational injury per 100 Full Time Employees (FTEs), second only to construction workers, and the highest number of cases involving days away from work, 7.8 per 100 FTEs (US Bureau of Labor Statistics (2008)). These figures are substantially higher than the national average of 4.2 incidents per 100 FTEs, with only half involving time away from work. Similarly, in Canada nurses have one of the highest absenteeism rates (12.2%) of all workers (second only to federal administration, and significantly higher than the Canadian national average of 8.7%), and this absenteeism rate has been increasing over the last 10 years (Statistics Canada (2008)).

Among the numerous potential causes of nurse absenteeism reported in the literature (illness or injury, conflicting family obligations, inflexible staff scheduling, transportation problems, income opportunities outside the workplace, etc.), work-related stress is often identified as one of the leading causes (Jamal (1984), Ho (1997), Shamian et al. (2003)). Numerous studies have emphasized a positive link between the workload nurses are subjected to, work-related stress, and the level of nurse absenteeism (Healy and McKay (1999), Healy and McKay (2000), Demerouti et al. (2000), Bryant et al. (2000), Tummers et al. (2001), McVicar (2003)), pointing to, in the words of Unruh et al. (2007), "a vicious cycle" in which nurse absenteeism leads to higher workloads for nurses who show up, further reinforcing the patterns of absenteeism. The existence of a positive feedback loop between workload and unplanned absences from workplace underscores the endogeneity of nurse absenteeism and alters the nature of the trade-off between having too few and too many nurses scheduled for a particular shift.

While numerous studies have put forward qualitative arguments about the effect of workloads on nurse absenteeism, quantitative studies of this effect as well as analyses of its influence on nurse staffing decisions have been lacking. One reason for this seems to be the lack of a uniformly accepted way of defining and measuring nursing workload, due to a multitude of patient classification schemes used in practice, differences in the nature and intensity of nursing care required in different clinical units, and diverging work environments (Giovannetti and Johnson (1990), Fagerstrom and Rainio (1999), Seago (2002)). Rauhala et al. (2007) is, to the best of our knowledge, the first and only paper which provides quantitative evidence of a positive link between patient-related nurse workload and nurse absenteeism. This work uses a new patient classification system in which the nurse workload created by each patient is recorded as a (self-reported) sum of "nursing intensity" scores from six pre-defined areas of nursing patient-centered activities. For each nurse, this workload measure is expressed as a fraction of the "optimal" workload as determined by a panel of experts. Based on 6 months of data obtained across multiple clinical units and multiple hospitals, the authors report that substantial increases in workload lead to equally substantial increases in absenteeism.

In this paper, we provide an alternative approach to measuring the strength of the link between the nursing workload and the rate of absenteeism using the data from an emergency department (ED) of a large urban hospital. In particular, rather than relying on subjective self-reported data, we use patient census values to calculate nurse-to-patient ratios which are treated as proxies for the workload experienced by nurses working during a particular shift.

The hypotheses about the endogenous nature of absenteeism and its dependence on workload are not limited to nursing. A substantial body of literature on the phenomenon of employee absenteeism exists in the applied psychology and economics domains (see Brown and Sessions (1996) for a comprehensive review of earlier research on the subject). Darr and Johns (2008) summarize the findings of a number of studies which identify work-related stress, often tied to excessive workload, as an important causal factor behind absenteeism. Similar findings have been made in various service settings, for example, among public sector workers (Scott and Wimbush (1991), Voss et al. (2001), Vingard et al. (2005)), and in call centers (Bakker et al. (2003), Aksin et al. (2007)). Despite the growing body of evidence on the importance of explicitly accounting for the endogeneity of employee absenteeism, the existing literature on optimal staffing in service environments either ignores absenteeism or treats it as an exogenous phenomenon (Easton and Goodale (2002), Easton and Goodale (2005), and Whitt (2006) provide examples of call-center staffing with exogenous employee absenteeism). The uncertain supply of service capacity created by nurse absenteeism connects our work with a stream of literature focused on inventory planning in the presence of unreliable supply/stochastic production yield (Yang and Lee (1995), Porteus (2002)), with two important distinctions. First, the overwhelming majority of papers which deal with stochastic supply yields model them as being either additive or multiplicative (Noori and Keller (1986), Ehrhardt and Taube (1987), Henig and Gerchak (1990), Ciarallo et al. (1994), Bollapragada and Morton (1999), Gupta and Cooper (2005), Rekik et al. (2007), Yang et al. (2007)), a justifiable approach in manufacturing settings. The supply uncertainty in our model has binomial structure, a more appropriate choice in personnel staffing settings. Binomial yield models are a relative rarity in the stochastic yield literature, perhaps, due to their limited analytical tractability (Beja (1977), Gerchak and Henig (1994), Grosfeld-Nir and Gerchak (2004), Fadiloglu et al. (2008)). Most importantly, our analysis is the first one to introduce the endogeneity of the yield rates into the binomial model of supply uncertainty. As such, our paper is also related to the growing body of literature (Diwas and Terwiesch (2009), Powell and Schultz (2004), Schultz et al. (1998), and Schultz et al. (1999)) on the effects of workload on various aspects of system productivity.

In summary, our paper makes the following contributions:

1. We conduct an empirical study of nurse absenteeism rates using data collected from the emergency department (ED) of a large urban hospital for each work shift for a period of about 10 months. Using nurse-to-patient ratios as proxies for the workload experienced by nurses, we establish that nurses exhibit a response to the anticipated workload for future shifts. In particular, absenteeism rates per shift show no statistically significant dependence on actual realizations of past shift workload levels, but rather depend on the numbers of nurses scheduled to work during the shift. Specifically, we find that for our data set with an average absenteeism rate of 7.3%, an extra scheduled nurse was associated with an average reduction in the absenteeism rate of 0.6%.

2. We model the nurse staffing decision in the presence of endogenous absenteeism as a novel variant of a single-period newsvendor problem, which combines the uncertainty in the future demand (patient census in a clinical unit) with the uncertainty in the supply of service capacity (number of nurses showing up for work) as well as the possibility of recourse supply adjustment through the use of overtime work and/or external labor supply (e.g., agency nurses). Using our empirical findings, we model nurse absenteeism as a binomial process for which the absenteeism rate is a function of the number of scheduled nurses.

3. We analyze staffing policies for the cases of exogenous and endogenous absenteeism rates. For the case of exogenous absenteeism, we characterize the optimal staffing levels for an arbitrary distribution of the demand for nurses (Proposition 1), and establish their monotonicity properties under a uniformly distributed demand (Corollary 1). In particular, we show that, depending on the cost ratio of regular to excess/overtime capacity, absenteeism could result in either increasing the number of staffed nurses or decreasing it. For the case of endogenous absenteeism we establish a sufficient condition for the absenteeism rate function which guarantees the optimality of the staffing level established by a greedy search approach (Proposition 2). We also compare the optimal staffing decision for a clinical unit that takes into account the endogeneity of absenteeism to that of a "myopic" unit which ignores the endogeneity of absenteeism. In particular, we show that even

allowing the "myopic" unit to repeatedly adjust its staffing decision after observing the actual absenteeism rate, it will staff fewer nurses than optimal (Proposition 3).

The rest of the paper is organized as follows. In Section 2, we describe the results of an empirical study of nurse absenteeism conducted at a large urban hospital ED. The implications of endogenous absenteeism behavior on staffing decisions are analyzed in Section 3 where we introduce and analyze a single-period nurse staffing model. We summarize our findings in Section 4.

2. Endogeneity in Nurse Absenteeism Rates: An Empirical Study

2.1. Clinical Environment and Collected Data

Our study is based on nurse absenteeism and patient census data from the ED of a large New York City hospital. Nurses employed in this unit are full-time employees, each working, on average, 3.25 shifts per week. The unit uses two primary nursing shifts: the "day" shift starts at 8:00am and ends at 8:00pm, while the "night" shift starts at 8:00pm and ends at 8:00am. Another ("evening") shift is also operated from 12:00pm to 12:00am. The evening shift is fundamentally different from the other two shifts. First, the nurses working on this shift are dedicated to this shift and, unlike the other nurses, do not work on the other two shifts. Second, this shift consists of fewer nurses who are more experienced and exhibit less absenteeism than the other two shifts, as shown in Table 1.

In our analysis of absenteeism we limit our attention to the nurses on the day and night shifts. However, we do take into account the evening shift when measuring workload since the evening shift overlaps with both the day shift and the night shift. For each shift, for a period of 10 months starting on July 1, 2008 (304 day shifts, 304 evening shifts and 303 night shifts), we collected the following data: the number of nurses scheduled, the number of nurses absent, and the patient census data recorded every two hours.

The nurse scheduling process starts several weeks before the actual work shift when the initial schedule is established. Often, such an initial schedule undergoes a number of changes and corrections (caused by unexpected events such as family illnesses, as well as planned events that were not communicated to management ahead of schedule such as jury duty and vacation absences) -

Measure	Mean	Standard Deviation	Minimum	Maximum
Day Shift				
Nurses Scheduled	11.4	1.07	8	16
Absenteeism rate	0.0762	0.0799	0	0.4
Patient visits	141	20.1	77	188
Average Census	116	17.1	56.5	158
Maximum Census	136	20.8	64	182
Night Shift				
Nurses Scheduled	10.5	0.849	9	14
Absenteeism rate	0.0707	0.0829	0	0.4
Patient Visits	66.0	9.45	40	95
Average Census	102	14.2	54.3	142
Maximum Census	127	20.4	57	174
Evening Shift				
Nurses Scheduled	3.63	0.756	2	5
Absenteeism rate	0.0589	0.119	0	0.5
Patient visits	137	16.2	75	196
Average Census	125	18.6	58.2	164
Maximum Census	137	20.7	64	182

Table 1 : Descriptive statistics for nurse and patient data.

and those changes continue until the day before the actual shift. In our study, we have used the *final* schedules, i.e., the last schedules in effect before any absenteeism is reported for the shift. The resulting descriptive statistics for three shifts are presented in Table 1.

The average patient census during a shift varies substantially from day to day. After controlling for day and week fixed effects, patient census during day shifts is more predictable (adjusted $R^2 = 52.2\%$) compared to patient census during night shifts (adjusted $R^2 = 32.6\%$). The patient census exhibits significant serial autocorrelation (even after fixed effects are accounted for) with the census during the previous 24 hours, which can be used to improve its predictability (adjusted $R^2 = 67.5\%$ and 50.5% for the day and night shifts respectively). The number of nurses scheduled for a particular type of shift, e.g. Wednesday day, was highly variable. After controlling for day and week fixed effects, the number of nurses scheduled for the day shift is slightly more predictable (with adjusted $R^2 = 27.5\%$ as compared to adjusted $R^2 = 24.1\%$ for the night shift). Also, after controlling for fixed effects, the number of nurses scheduled for a shift shows no statistically significant dependence (at 5% level) on either the patient census during that shift or on the census values for the 14 previous shifts, which correspond to one calendar week.

While we were unable to verify the precise reasons behind significant variations in the numbers of scheduled nurses, we hypothesize that there are two main factors at play here. First of all, the personnel scheduling in hospital environments is often subject to numerous constraints (e.g., union rules). Consequently, a scheduling manager (typically, a senior nurse) may not always be able to schedule the number of nurses he/she considers necessary for a particular shift. In addition, as we mentioned earlier, schedules adopted a few weeks ahead of the actual shift often undergo a series of changes. Both of these factors combine to create a set of "shocks" to the final numbers of scheduled nurses. While such scheduling variations may not be desirable from the point of view of managing the match between the demand for nursing services and the supply of nursing capacity, they provide us with an opportunity to examine how the absenteeism rates are related to the numbers of scheduled nurses.

2.2. Nurse Workload and Absenteeism: Empirical Results

In our study, we model the phenomenon of nurse absenteeism as follows. We treat all nurses as being identical and independent decision makers, and focus on a group of y_t nurses scheduled to work during a particular shift t ($t=1$ for the first shift in the data set, $t=2$ for the second shift, etc., up to $t=607$). For nurse n, $n=1,...,y_t$, the binary variable $Y_{n,t}$ denotes her decision to be absent from work ($Y_{n,t}=1$), or to be present ($Y_{n,t}=0$). We assume that this absenteeism decision is influenced by a number of factors expressed by the vector $\mathbf{x}$ which include parameters related to workload as well as fixed effects such as the day of the week or the shift. Each nurse compares the utility she receives from being absent from work to the utility she receives from going to work. The difference in these utility values is given by $U^*_{n,t}=\mathbf{x}'_t\beta+\epsilon_{n,t}$, where $\epsilon_{n,t}$ are, for each n and t, *i.i.d.* random variables with mean zero. While the utility difference $U^*_{n,t}$ is an unobservable quantity, we can potentially observe each nurse's decision to show up for work. The decision is such that

$$Y_{n,t}=\begin{cases}1, \text{ if } U^*_{n,t}>0,\\ 0, \text{ if } U^*_{n,t}\le 0.\end{cases} \tag{1}$$

Assuming that $\epsilon_{n,t}$ follow the standardized logistic distribution (the standard normal distribution) we obtain the logit (probit) model (Greene (2003)). Note that the assumption of standardized distributions (which sets the variance to 1 for the normal and to $\frac{\pi^2}{3}$ for the logistic distributions) is not a restrictive one. Any other variance would only introduce a multiplicative change to $U^*_{n,t}$ and, since it is only the sign of $U^*_{n,t}$ that matters, nothing is changed by assuming the variance of the standardized distributions. Similarly, the assumption that the decision changes at the utility level of 0 and not some other arbitrary level is not a restrictive one provided that the vector of covariates $\mathbf{x}$ contains a constant term. It is important to keep in mind that our empirical data do not record attendance decisions of individual nurses - rather, we measured the aggregate absenteeism behavior of a group of nurses scheduled for a particular shift. Consequently, we treat all nurses scheduled for a given shift as a homogenous group and build the model for the corresponding group behavior. In particular, we focus on maximum-likelihood-based logit estimation of the probability of absenteeism γ_t during shift t.

Since our goal is to study how the nurse absenteeism rate is affected by workload, we need to measure and quantify nurse workload for each shift. To do so we use the nurse-to-patient ratio as a proxy for the workload nurses experience during a particular shift. The use of the nurse-to-patient ratio as a workload proxy relies on the assumption that each ED patient generates the same amount of nursing workload. In the absence of case mix or patient acuity data, this assumption is consistent with the standard nurse-to-patient ratio approach used in nurse staffing and is justifiable for the purposes of aggregate staff planning presented in Section 3. For shift t, we define the nurse-to-patient ratio variable, denoted as NPR_t, as the ratio of the number of nurses working during a particular shift and the patient census averaged over the duration of that shift. To estimate the number of nurses present, we assume that the number of nurses scheduled in a particular shift *is* the number of nurses actually present. This assumption is consistent with the practice in the medical unit we studied: when a nurse scheduled to work during any given shift fails to appear, the nurse manager fills the gap by either enlisting a nurse from the preceding shift to work overtime, or by using an external supply of nursing capacity (e.g., agency nurses). With this assumption, the number of nurses present during each 24-hour period varies as follows: between 8:00am and

12:00pm it is equal to the number of nurses scheduled for the day shift (y_t), between 12:00pm and 8:00pm it is equal to the number of nurses scheduled for the day shift (y_t) plus the number of nurses scheduled for the evening shift (e_t), between 8:00pm and 12:00am it is equal to the number of nurses scheduled for the evening shift (e_t) plus the number of nurses scheduled for the night shift, while between 12:00am and 8:00am it is equal to the number of nurses scheduled for the night shift (y_t). Thus, we estimate NPR_t as follows

$$\begin{aligned} \mathrm{NPR}_t &= \frac{y_t + \frac{2}{3}e_t}{C_t} \text{ for the day shift,} \\ \mathrm{NPR}_t &= \frac{y + \frac{1}{3}e_t}{C_t} \text{ for the night shift,} \end{aligned} \tag{2}$$

where C_t is the patient census averaged over the duration of shift t.

In making their attendance decisions for shift t, nurses may be influenced by the *past* values of the nurse-to-patient ratios as well as by the value they *anticipate* for shift t. The absenteeism connection to the past workload values reflects a standard "burnout" argument (Healy and McKay (1999), Healy and McKay (2000), Demerouti et al. (2000), Bryant et al. (2000), Tummers et al. (2001), McVicar (2003), Unruh et al. (2007), Rauhala et al. (2007)). However, one has to keep in mind that while such an effect can be expected to be present in individual nurse absenteeism data, it is not clear if it also transpires in the group absenteeism data, since nurses forming the same group, i.e. scheduled for the same shift, did not necessarily experience the same sequence of workload values in the past. The avoidance of *anticipated* high workloads arises because nurses are informed in advance of their schedule and they are aware of how many (and which) other nurses are scheduled to work on the same shift as them. Since nurses anticipate a certain patient census $E[C_t]$, consistent with their past experience of working in the ED, nurses form an expectation about the anticipated workload for that shift. Naturally, if fewer (more) nurses are scheduled on that particular shift than the nurses deem appropriate, they will anticipate a higher (lower) workload than normal. While intuitive, this effect, to the best of our knowledge, has not been studied before. Note that the group attendance data do not present a measurement challenge in this case, since the nurses scheduled for the same shift are to be subjected to the same workload value.

In order to test for the presence of the "burnout" effect, we include in $\mathbf{x}_t$ the values of 14 lagged nurse-to-patient ratios NPR_{t-j}, $j = 1, \ldots, 14$, which correspond to one calendar week (as the number of past shifts we use is rather arbitrary, we conducted our statistical analysis for several different values to make sure the results are not sensitive to the number we choose as long as it is sufficiently large). The "workload avoidance" effect was tested by including in $\mathbf{x}_t$ the anticipated value of nurse-to-patient ratio

$$\begin{aligned} \mathrm{ENPR}_t^1 &= \frac{y_t + \frac{2}{3}e_t}{E\left[C_t\right]} \text{ for the day shift,} \\ \mathrm{ENPR}_t^1 &= \frac{y_t + \frac{1}{3}e_t}{E\left[C_t\right]} \text{ for the night shift,} \end{aligned} \tag{3}$$

where e_t is the number of nurses scheduled in the evening shift which overlaps with $\frac{2}{3}$ $(\frac{1}{3})$ of the duration of the day (night) shift in question. While day and night shift nurses are informed in detail about the schedule for their shifts, it is not clear that they would be as familiar with the details of the schedule of the evening shift staffed by a different pool of nurses. Motivated by this observation, we estimate two models based on alternative definitions of the expected nurse-to-patient ratio. In the first definition (ENPR_t^1 of equation (3)) we use the exact number of evening nurses scheduled (e_t), while in the second definition (ENPR_t^2) we use the average value of e_t (averaged over all evening shifts in our sample). The latter formulation reflects the situation where day- and night-shift nurses do not know precisely how many evening nurses will be present but they form a rational expectation about this value. In other words, in the second model day- and night-shift nurses behave as if they ignore any variation in the number of nurses scheduled for the evening shift that overlaps with their own shift. $E\left[C_t\right]$ is set to the expectation of patient census values computed over *all* shifts in our sample. This formulation reflects an assumption that nurses, when making their attendance decisions, use a mental model which captures any potential difference occurring on different days/shifts with a fixed effect and, therefore, focus on expected patient census value. We also assume that the nurses form rational expectations about the patients census which are consistent with empirically observed patient census data. In addition to the models based on (3), we have also estimated several alternative variants in which, when estimating the expected patient census during the upcoming shift, nurses a) distinguish between day and night shifts, but not between days of the week, b) distinguish between days of the week, but not between day shifts or night shifts, and c) distinguish between both the shift type and the day of the week. The estimation results for these alternative models are presented in the Appendix and are quite similar to those based on equation (3).

Besides the 14 lagged values of the nurse-to-patient ratio (NPR) and the anticipated nurse-to-patient ratio ($\text{ENPR}_t^i, i = 1, 2$), the vector of covariates $\mathbf{x_t}$ includes a number of dummy variables that capture fixed effects. We include a day-of-the-week dummy variable to capture any systematic variation in absenteeism across days, a day/night-shift fixed effect to capture variations between day and night shifts, and a week fixed effect to capture any systematic variations which remain constant over a period of one week and affect absenteeism but are otherwise unobservable. Finally, we include a holiday fixed effect which takes the value of 1 on US public holidays and zero on any other day. This last variable is designed to deal with a potential endogeneity problem, since nurses may be inherently reluctant to work on some select days, such as public holidays. These days are known to the management of the clinical unit which tries to accommodate nurses' aversion by staffing fewer nurses on such days. Nevertheless, the nurses that are scheduled to work on these "undesirable" days are still more likely to be absent, irrespective of the chosen staffing levels. By including the holiday variable we are trying to explicitly account for this effect. It is possible that there exist other correlated variables that we omit, but to the extent that they do not vary drastically over a period of one week, the week fixed effect should be able to capture the influence of those variables.

Specifically, the models we estimate are:

$$\begin{aligned}\text{logit}(\gamma_t) = {} & \beta_0^i + \beta_{\text{ENPR}}^i \times \text{ENPR}_t^i + \sum_{j=1}^{14} \beta_{\text{NPR},j}^i \times \text{NPR}_{t-j} + \sum_{d=2}^{7} \beta_{\text{DAY},d}^i \times \text{DAY}_{d,t} \\ & + \sum_{f=2}^{44} \beta_{W,f}^i \times W_{f,t} + \beta_{\text{DAYSHIFT}}^i \times \text{DAYSHIFT}_t + \beta_{\text{HOLIDAY}}^i \times \text{HOLIDAY}_t, \end{aligned} \tag{4}$$

where γ_t is the probability that a nurse is absent in shift t, $i = 1, 2$ refers to the definition of ENPR^i used, $\text{DAY}_{d,t}$ and $W_{f,t}$ are the day and week fixed effects, $DAYSHIFT_t$ and HOLIDAY_t are the shift and holiday fixed effects. The estimation results for equation (4) are presented in Table 2. Model I uses the first definition of the anticipated nurse-to-patient ratio (ENPR_t^1), while Model III uses the second definition (ENPR_t^2).

As we have argued earlier, the effect of the lagged nurse-to-patient ratios as proxies for workload experienced by the nurses, might not be detectable in the data which is aggregated across the nurse pool. The effect of the anticipated nurse-to-patient ratio, however, does not suffer from such data aggregation. In order to further test whether the anticipated nurse-to-patient ratio has any effect on absenteeism, we also estimated the restricted versions of Models I and III (which we denote as Models II and IV), where we omit the 14 lagged nurse-to-patient ratio variables. If the lagged nurse-to-patient ratios are not related to absenteeism (i.e. if $\beta^i_{\text{NPR},j} = 0$ for all $j = 1, \ldots, 14$), omitting these variables will not introduce any bias even if the lagged nurse-to-patient variables are correlated with the variables included in the model. Model II uses the first definition of the anticipated nurse-to-patient ratio (ENPR_t^1), while Model IV uses the second definition (ENPR_t^2).

As can be seen from Table 2, the anticipated nurse-to-patient ratio has a significant effect (at the 5% or 10% level) on absenteeism rates in Models I, III and IV. In Model II the p-value of the anticipated nurse-to-patient ratio is 10.7%. The more nurses that are scheduled for a particular shift, the less likely each nurse is to be absent. In particular, according to the first model we estimate, the marginal effect of staffing an extra nurse (calculated at the mean values of all remaining independent variables, using the expected patient census value of 109) on the individual absenteeism rate is around $0.575\% = 0.626/109$. In other words, the absenteeism rate would decrease from its average value of 7.34% to about 6.78% when an extra nurse is added to the schedule. The coefficient of the anticipated nurse-to-patient ratio is statistically more significant in Models III and IV, where the variation in the number of scheduled evening nurses is ignored. This might suggest that when nurses decide whether to show up for work, they place greater emphasis on the number of nurses

	MODEL I		MODEL II		MODEL III		MODEL IV	
Variable	Coefficient (robust error)	Marginal Effect+ (robust error)	Coefficient (robust error)	Marginal Effect+ (robust error)	Coefficient (robust error)	Marginal Effect+ (robust error)	Coefficient (robust error)	Marginal Effect+ (robust error)
ENPR[1]	-10.01*	-0.626*	-8.534	-0.536				
	(5.669)	(0.355)	(5.291)	(0.333)				
ENPR[2]					-15.24**	-0.952**	-13.25**	-0.832**
					(6.355)	(0.398)	(5.892)	(0.372)
NPR1	-0.248	-0.0155			-0.474	-0.0296		
	(3.061)	(0.191)			(3.071)	(0.192)		
NPR2	2.463	0.154			3.024	0.189		
	(3.438)	(0.215)			(3.459)	(0.216)		
NPR3	-0.492	-0.0308			-0.712	-0.0445		
	(2.989)	(0.187)			(2.982)	(0.186)		
NPR4	3.314	0.207			3.673	0.229		
	(3.265)	(0.204)			(3.275)	(0.204)		
NPR5	-2.620	-0.164			-2.917	-0.182		
	(3.020)	(0.189)			(3.023)	(0.189)		
NPR6	5.912*	0.369*			5.837*	0.365*		
	(3.249)	(0.202)			(3.251)	(0.202)		
NPR7	0.518	0.0324			0.615	0.0384		
	(3.155)	(0.197)			(3.157)	(0.197)		
NPR8	1.556	0.0972			1.570	0.0981		
	(3.245)	(0.203)			(3.254)	(0.203)		
NPR9	-4.112	-0.257			-4.252	-0.266		
	(2.998)	(0.187)			(3.002)	(0.187)		
NPR10	5.656*	0.353*			5.746*	0.359*		
	(3.205)	(0.200)			(3.207)	(0.200)		
NPR11	-3.320	-0.207			-3.263	-0.204		
	(2.980)	(0.186)			(2.991)	(0.187)		
NPR12	2.462	0.154			2.707	0.169		
	(3.271)	(0.205)			(3.276)	(0.205)		
NPR13	-2.934	-0.183			-2.724	-0.170		
	(3.016)	(0.189)			(3.018)	(0.189)		
NPR14	4.934*	0.308*			4.823	0.301		
	(2.987)	(0.186)			(2.978)	(0.186)		
Dayshift	0.112	0.00698	0.233	0.0146	0.211	0.0132	0.327**	0.0206**
	(0.151)	(0.00945)	(0.143)	(0.00905)	(0.163)	(0.0102)	(0.155)	(0.00980)
Holiday	-0.853*	-0.0384***	-0.787**	-0.0365***	-0.848*	-0.0382***	-0.783*	-0.0364***
	(0.440)	(0.0135)	(0.401)	(0.0131)	(0.440)	(0.0135)	(0.401)	(0.0131)
Tuesday	-0.0567	-0.00348	0.0202	0.00128	-0.0557	-0.00342	0.0134	0.000844
	(0.227)	(0.0137)	(0.192)	(0.0122)	(0.227)	(0.0137)	(0.192)	(0.0121)
Wednesday	0.0484	0.00307	0.0372	0.00236	0.0545	0.00346	0.0326	0.00207
	(0.238)	(0.0153)	(0.192)	(0.0123)	(0.236)	(0.0153)	(0.191)	(0.0122)
Thursday	0.169	0.0111	0.133	0.00870	0.180	0.0119	0.137	0.00898
	(0.229)	(0.0159)	(0.189)	(0.0129)	(0.226)	(0.0158)	(0.188)	(0.0129)
Friday	0.285	0.0195	0.240	0.0162	0.278	0.0190	0.239	0.0162
	(0.235)	(0.0175)	(0.186)	(0.0136)	(0.233)	(0.0173)	(0.186)	(0.0135)
Saturday	0.302	0.0208	0.342*	0.0239*	0.309	0.0213	0.345*	0.0242*
	(0.241)	(0.0181)	(0.182)	(0.0141)	(0.241)	(0.0182)	(0.181)	(0.0141)
Sunday	0.226	0.0152	0.291	0.0200	0.247	0.0167	0.304*	0.0210
	(0.218)	(0.0156)	(0.184)	(0.0139)	(0.217)	(0.0158)	(0.184)	(0.0139)
Constant	-2.324		-2.192***		-1.778		-1.740**	
	(1.420)		(0.753)		(1.443)		(0.783)	
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6481		6631		6481		6631	
Log Likelihood	-1654		-1690		-1653		-1688	
LR-test	102.7		92.35		105.3		95.13	

***, **, * denote statistical significance at 1% 5% and 10% confidence levels, respectively. Robust errors are shown in parentheses.
+Marginal effects are calculated at the mean values of the independent variables. For the dummy variables the marginal effect shows the difference caused by changing the variable from 0 to 1.

Table 2: Estimation results for logit models.

working in their shift rather than the number of nurses working in the evening shift that overlaps with their own.

It is interesting to note that the lag 6, lag 10 and lag 14 (lag 6 and lag 10) of the nurse-to-patient ratio variables have positive coefficients in Model I (Model III) which are individually significant at the 10% level. This seems to imply that the probability of a nurse being absent on any shift would increase if the shift 3, 5 or 7 days ago had a higher nurse-to-patient ratio. Although these lags are individually significant, the Wald test statistic and the Likelihood Ratio test statistic for joint significance of all 14 lagged workload variables in Model I as well as Model III reject the hypothesis (even at the 10% level) that lagged nurse-to-patient ratios have any joint explanatory power. In the absence of any plausible explanation as to why a lighter workload on a similar shift 3, 5 or 7 days ago might increase absenteeism, and in light of the weakness of this statistical relationship, we are inclined to treat this result as spurious. One possible reason for our inability to find a statistically significant connection between past workload values and the absenteeism rates may be related to our use of cumulative, unit-level, rather than individual-nurse level absenteeism data.

Interestingly, the holiday variable's effect is significant (at 10% confidence level) and negative, thus suggesting that nurses are about 3.8% less likely to be absent on public holidays. In addition, as indicated by the models that do not include the lagged variables (Models II and IV), there is some (weak) evidence that nurses are more likely to be absent on weekends and they are also more likely to be absent during a day shift, when conflicting family obligations often cited as an important reason behind nurse absenteeism (Gillies (1994), Erickson et al. (2000), Nevidjon and Erickson (2001)) are likely to be more prevalent.

Week fixed effects are jointly significant (at the 1% level) in all four models. One of the effects that week dummies seem to pick up reasonably well is the impact of weather (in particular, heavy snow conditions) on absenteeism. For example, week 35 of our data set includes March 2-4, 2009. During six day and night shifts corresponding to these dates snow on the ground in New York City was recorded to be more than 5 inches,[1] the level identified by the New York Metropolitan Transportation Authority as the one at which the public transportation disruptions are likely to set in.[2] The week-35 fixed effect is positive and significant (at the 5% level) in Models II and IV

[1] www.accuweather.com

[2] www.mta.info/lirr/WinterWeatherTravelTips/

with a marginal effect equal to 8.75% in Model I and 9.80% in Model IV. Only 3 other days in our dataset had as much snow on the ground (December 3, December 21 and January 20).

2.3. Verification Tests and Study Limitations

In order to check the validity of our model estimation procedure, we have also conducted several verification tests described below.

2.3.1. Alternative specifications To ensure the robustness of our results, we estimated a number of alternative modeling specifications. Namely, we estimated the models of equation (4) under the probit specification. The results were almost identical in terms of variable significance, model significance and magnitude of marginal effects. We also estimated variants of our models which use month fixed effect variables instead of week fixed effect variables, and the estimation results were similar to those for the models described by (4). Finally, we estimated models with alternative definitions of the expected nurse-to-patient ratio, as reported in the Appendix. Our main finding that higher (lower) anticipated nurse-to-patient ratios decrease (increase) nurse absenteeism appears to be robust to these alternative modeling specifications.

2.3.2. Comparison Between Best-Fit Coefficients for Two Nursing Shifts: Joint versus Separate Models The model and its analysis described in Section 2.2 assume that any difference between the night and the day shifts is completely captured by the dummy variable "Dayshift". However, it is possible that the two shifts are inherently different, and so are their best-fit coefficients. To test the hypothesis that the best-fit coefficients for two shifts differ, we fit an unrestricted binary choice model to the data for each of the two shifts separately, and compared the fit with the restricted model fitted to both shifts by constructing a Likelihood Ratio test. The test provides support for the hypothesis that the values of the best-fit coefficients for the two shifts are identical; the hypothesis that these coefficients are not the same is rejected at the 5% confidence level in all models.

2.3.3. Data limitations The main data limitations of our study can be outlined as follows. First, we have only collected data for a duration of ten months - and this limitation does not allow us to study seasonal variations in the absenteeism rates. Second, our data comes from a single, specialized clinical unit (Emergency Department) of a particular hospital. Confirming our results using data from other types of clinical units or from a similar department of a different hospital would strengthen our empirical findings.

Finally, our data is aggregated at the working shift level. Individual-nurse-level data would allow us to better measure the effects of both the workload and the staffing decisions on absenteeism. It would also enable us to observe any heterogeneity in the behavior of nurses and capture any correlations in the absenteeism behavior of individual nurses.

3. Endogenous Nurse Absenteeism: Implications for Nurse Staffing

To study the implications of absenteeism, and, in particular, of its endogenous nature on staffing level decisions, we construct a stylized model of nurse staffing. The aim of such a model is not to produce a decision support tool for the clinical unit in question, but, rather, to generate managerial insights regarding the impact of nurse absenteeism, in general, and the endogenous nature of absenteeism, in particular, on a decision of how many nurses to staff.

We assume that a clinical unit uses the primary nursing care (PNC) mode of nursing care delivery (Seago (2001)) which was employed in the ED we studied. Under the PNC mode, the nursing staff includes only registered nurses (as opposed to licensed practical nurses or unlicensed nursing personnel) who provide all direct patient care throughout the patient's stay in the clinical unit. The nurse staffing process starts several weeks in advance of the actual shift for which planning is performed, and it is then that a hospital staff planner needs to decide how many nurses (y) to schedule for that particular shift. Due to the phenomenon of absenteeism, the actual number of nurses who show up for work on that shift, N, is uncertain. We model N as a binomial random variable $B(y, 1-\gamma(y))$, where $\gamma(y)$ is the probability that any scheduled nurse will be absent from work:

$$\text{Prob}(N=k|y,\gamma(y)) = p(k;y,\gamma(y)) = \begin{cases} \frac{y!}{k!(y-k)!}(\gamma(y))^{(y-k)}(1-\gamma(y))^k, & \text{for } 0 \le k \le y, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

A simpler, "exogenous" approach to modeling the absenteeism would treat γ as a constant, independent of any other problem parameters, including the number of scheduled nurses y. In a more complex, "endogenous" setting described in Section 2, the absenteeism probability γ depends on the anticipated nurse-to-patient ratio, which, for given expected patient census value, is a function of y.

We assume that the clinical unit follows a policy of specifying, for each value of the average patient census during a shift, C, a target integer number of nurses $T = R(C)$ required to provide an adequate patient care during a particular shift. We assume that C takes on discrete values and that $R(C)$ is a monotone increasing function with $R(0) = 0$. A simple example of $R(C)$ is provided by a "ratio" approach, under which $R(C) = \lceil \alpha C \rceil$, with $\alpha \in [0,1]$ representing a mandated nurse-to-patient ratio. Alternately, if a clinical unit is modeled as a queueing system in which patients generate service requests and nurses play the role of servers, as was done in Yankovic and Green (2010), $R(C)$ can take a more complex form to ensure that certain patient service performance measures, such as the expected time patients wait to be served, conform to pre-specified constraints. At the time of the nurse staffing decision, we assume that the decision maker uses a known probability density function of the average patient census C during the shift for which the personnel planning is conducted:

$$\begin{aligned} \text{Prob}\,(C=n) &= p_C(n), n \in \mathcal{N}^+, \\ \sum_{n=0}^{\infty} p_C(n) &= 1. \end{aligned} \tag{6}$$

We treat the uncertain factors in our model (the demand uncertainty expressed by the patient census C and the supply uncertainty expressed by N) as being independent, and assume that the realized values of C and N become known shortly before the beginning of the shift. Any nursing shortage $(R(C)-N)^+$ is covered by either hiring agency nurses or asking nurses who have just completed their shift to stay overtime. We further assume that, once scheduled, all nurses are paid

\$$w_r$ per shift, whether they actually show up for work or not. In addition, if more nurses show up for work than required ($N > R(C)$) they all have to be paid and cannot be "sent home". The per-shift cost of extra/overtime nurses is \$$w_e$, which, we assume, is greater than \$$w_r$. The goal of the decision maker is choose a nurse staffing level y which minimizes the expected cost $W(y)$ of meeting the target $R(C)$:

$$W(y) = w_r y + w_e \mathbf{E}_{C,N}\left[(R(C) - N)^+\right], \tag{7}$$

where $\mathbf{E}_{C,N}$ denotes expectation taken with respect to both the number of patients and the number of nurses that show up for work. Note that since there is a one-to-one correspondence in our model between the patient demand C and the number of required nurses T, we can re-cast the calculation of the expectation with respect to the demand value in terms of an equivalent calculation over the distribution of T, using the corresponding probability distribution function. In particular, let $\mathcal{S}_n$ be the set of average patient census values, all corresponding to the same number of required nurses n:

$$\mathcal{S}_n = \left(C \in \mathcal{N}^+ | R(C) = n\right). \tag{8}$$

Then, the probability distribution for T is given by

$$\text{Prob}\,(T = n) = p_T(n), n \in \mathcal{N}^+, \tag{9}$$

where

$$\begin{aligned} p_T(n) &= \sum_{l \in \mathcal{S}_n} p_C(l), \\ \sum_{n=0}^{\infty} p_T(n) &= 1. \end{aligned} \tag{10}$$

In turn, the cost minimization based on (7) becomes

$$\min_{y \in \mathcal{N}^+} \left(w_r y + w_e \mathbf{E}_{T,N}\left[(T - N)^+\right]\right). \tag{11}$$

Note that with no absenteeism ($\gamma(y)=0$), the number of nurses showing up for work N is equal to the number of scheduled nurses y, and the nurse staffing problem reduces to a standard newsvendor model, with the optimal staffing level given by

$$y_0^* = \min\left(y \in \mathcal{N}^+ | F_T(y) \geq 1 - \frac{w_r}{w_e}\right), \tag{12}$$

with

$$F_T(y) = \sum_{n=0}^{y} p_T(n) \tag{13}$$

being the cumulative density function of the demand function evaluated at y, and the value $1-\frac{w_r}{w_e}$ playing the role of the critical newsvendor fractile.

Below we present an analysis of the staffing decision (11), starting with the case of exogenous absenteeism which we will use as a benchmark.

3.1. Optimal Nurse Staffing Under Exogenous Absenteeism Rate

Consider a clinical unit which experiences an endogenous nurses' absenteeism rate $\gamma(y)$, but treats it as exogenous. For example, the schedule planner uses the average value of all previously observed daily absenteeism rates, γ_{ave}. The cost function to be minimized under this approach is given by

$$\begin{aligned} W_{\text{ave}}(y) &= w_r y + w_e \sum_{k=0}^{y}\sum_{n=0}^{\infty}(n-k)^+ p_T(n) p(k; y, \gamma_{\text{ave}}) \\ &= w_r y + w_e \sum_{k=0}^{y} q(k) p(k; y, \gamma_{\text{ave}}), \end{aligned} \tag{14}$$

where

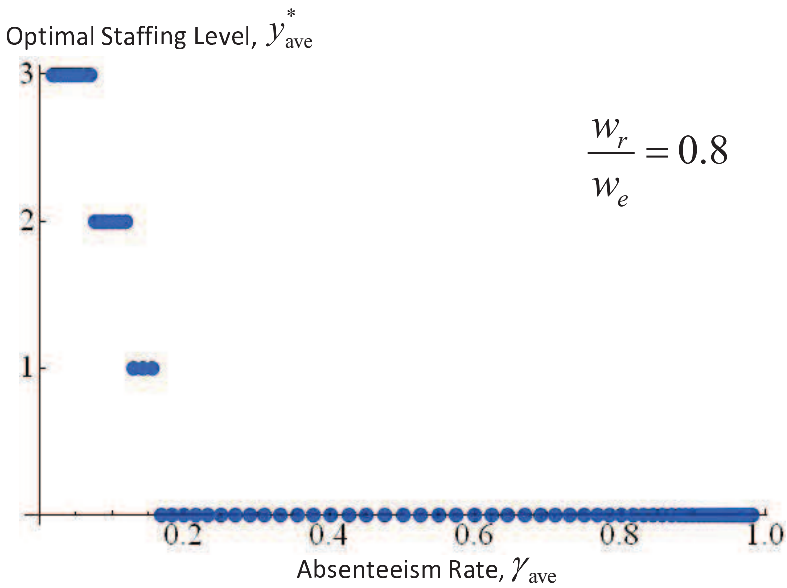
$$q(k) = \sum_{n=0}^{\infty}(n-k)^+ p_T(n). \tag{15}$$

The optimal staffing level in this case is expressed by the following result.

Proposition 1. *a) The minimizer of (14) is given by*

$$y_{\text{ave}}^* = \min\left(y \in \mathcal{N}^+ | \sum_{k=0}^{y} F_T(k) p(k; y, \gamma_{\text{ave}}) \geq 1 - \frac{w_r}{w_e(1-\gamma_{\text{ave}})}\right), \tag{16}$$

and is a non-increasing function of $\frac{w_r}{w_e}$.



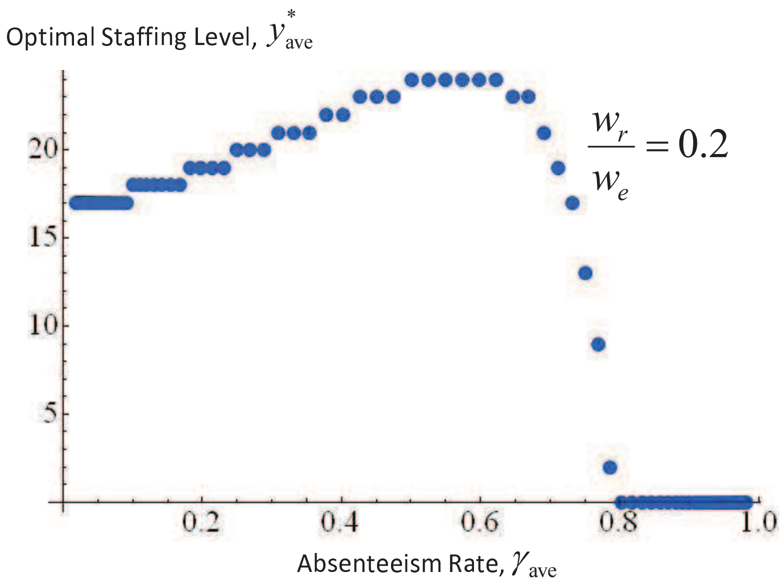


Figure 1: Optimal staffing level as a function of the absenteeism rate for different values of the cost ratio $\frac{w_r}{w_e}$ and a discrete uniform target nursing level distribution with support on $[0, T_{\max}]$ with $T_{\max} = 10$.

b) Consider two cumulative distribution functions for the required number of nurses T, $F_T^1(k)$ and $F_T^2(k)$ such that $F_T^1(k) \geq F_T^2(k)$ for all $k \in \mathcal{N}^+$, and let $y^{*,i}_{\text{ave}}$ be the optimal staffing levels corresponding to $F_T^i(k)$, $i = 1, 2$. Then, $y^{*,1}_{\text{ave}} \leq y^{*,2}_{\text{ave}}$.

We relegate all the proofs to the Appendix. Note that (16) represents a generalization of the expression for the optimal staffing levels without absenteeism (12). As in the no-absenteeism setting, it is never optimal to decrease staffing levels when the target nursing level increases or when the cost advantage associated with earlier staffing becomes more pronounced. While this behavior of the optimal policy is intuitive, the dependence of the optimal staffing levels on the value of the absenteeism rate is not as straightforward. In particular, depending on the interplay between the ratio of the cost parameters $\frac{w_r}{w_e}$, the characteristics of the target nursing level distribution, and the absenteeism rate, the increase in the absenteeism rate can increase or decrease the optimal staffing level. A detailed characterization can be obtained for some target nursing level distributions, for example, for a discrete uniform distribution:

Corollary 1. *Let*

$$F_T(k) = \begin{cases} \frac{k+1}{T_{\max}+1}, & \textit{for } 0 \le k \le T_{\max}, \\ 1, & k \ge T_{\max}. \end{cases} \tag{17}$$

Then, for $\frac{w_r}{w_e} \ge \frac{1}{4}$, *the optimal nurse staffing level is given by*

$$y^*_{\text{ave}} = \left\lceil \left(\frac{T_{\max}}{1-\gamma_{\text{ave}}} - \frac{T_{\max}+1}{(1-\gamma_{\text{ave}})^2} \frac{w_r}{w_e} \right) \right\rceil, \tag{18}$$

and is a non-decreasing (non-increasing) function of γ_{ave} *for* $\gamma_{\text{ave}} \le \gamma^{\text{u}}_{\text{ave}}$ ($\gamma_{\text{ave}} > \gamma^{\text{u}}_{\text{ave}}$), *where*

$$\gamma^{\text{u}}_{\text{ave}} = \max\left(0, 1 - 2\left(1 + \frac{1}{T_{\max}}\right) \frac{w_r}{w_e}\right). \tag{19}$$

The monotonicity properties of the optimal staffing levels formalized in Corollary 1 are illustrated in Figure 1: for a given value of the cost ratio $\frac{w_r}{w_e}$, there exists a critical value of the absenteeism rate $\gamma^{\text{u}}_{\text{ave}}$ for which the optimal response to an increase in absenteeism switches from staffing more nurses to staffing fewer. Note that, irrespective of the distribution for targeted nursing level, for high value of the absenteeism rate or high value of the cost ratio $\frac{w_r}{w_e}$ (to be precise, for $\gamma_{\text{ave}} \ge 1 - \frac{w_r}{w_e}$), it is more cost-effective not to staff any nurses in advance and to rely exclusively on the extra/overtime mechanisms of supplying the nursing capacity. For low values of the absenteeism rate and low values of the cost ratio $\frac{w_r}{w_e}$, higher absenteeism can induce an increase in staffing level, as it is cheaper to counter the increased absenteeism by staffing more nurses. However, as the cost ratio $\frac{w_r}{w_e}$ increases, it becomes more cost-effective to staff fewer nurses, relying increasingly on the extra/overtime supply mechanism. The fact that the critical rate $\gamma^{\mathrm{u}}_{\mathrm{ave}}$ is a non-decreasing function of the expected demand and the non-increasing function of the cost ratio $\frac{w_r}{w_e}$ is intuitive: the higher are the expected demand for nurses and the cost advantage of earlier staffing, the wider is the range of absenteeism rates which justify a staffing increase as a response to an increase in the threat of absenteeism.

3.2. Endogenous Absenteeism: Optimal Staffing

We now explicitly take into account the endogenous nature of nurse absenteeism. Consistent with our empirical findings, in particular, with the logit model specification, we use

$$\gamma(y) = \frac{1}{1 + e^{\alpha + \beta y}}, \tag{20}$$

where both α and β are positive constants. The assumption about positive values for these absenteeism rate parameters is plausible in a wide range of settings: $\beta > 0$ implies that the absenteeism rate declines with the number of scheduled nurses, while $\alpha > 0$ ensures that the absenteeism rate is not too high even when the number of scheduled nurses is low and the anticipated workload is high. In particular, evaluating the best-fit logit model in (4) using the estimates reported in Table 2 we obtain

$$\beta = -\frac{\beta^1_{\mathrm{ENPR}}}{E[C_t]} = 0.092, \tag{21}$$

with $\beta^1_{\mathrm{ENPR}} = -10.01$, $E[C_t] = 109.0$. In order for the average absenteeism rate to match our sample average of 7.34% we set

$$\alpha = 1.533. \tag{22}$$

In the endogenous absenteeism setting, the expected staffing cost (11) becomes

$$W(y) = w_r y + w_e L(y, \gamma(y)), \tag{23}$$

where

$$L(y,\gamma(y)) = \sum_{k=0}^{y} q(k)p(k;y,\gamma(y)), \tag{24}$$

with $q(k)$ defined by (15). Note that for general absenteeism rate function $\gamma(y)$ the increasing marginal property of the "exogenous" staffing cost function (14) with respect to the number of scheduled nurses may not hold. Below we formulate a sufficient condition for this property to be preserved under endogenous absenteeism. First, for a given distribution of the targeted nursing level $p_T(k), k \geq 0$, we introduce the following quantity:

$$\gamma_T(y) = 1 - \min\left(1, \left(\frac{\sum_{k=y-2}^{\infty} p_T(k)}{yp_T(y-1) + p_T(y-2)}\right)^{\frac{1}{y-1}}\right), \; y \in \mathcal{N}^+, y \geq 2. \tag{25}$$

As shown below, (25) represents one of the bounds on the absenteeism rate function which ensure the optimality of the greedy-search approach to finding the optimal nurse staffing level.

Proposition 2. *Consider an endogenous absenteeism setting characterized by (20) with $\alpha, \beta \geq 0$. Then, the optimal staffing level is given by*

$$y^* = \min\left(y \in \mathcal{N}^+ | L(y+1,\gamma(y+1)) - L(y,\gamma(y)) \geq -\frac{w_r}{w_e}\right), \tag{26}$$

and is a non-increasing function of $\frac{w_r}{w_e}$, provided that

$$\gamma(y) \leq \min\left(\frac{2}{y}, \gamma_T(y)\right) \tag{27}$$

for any $y \geq y^*$. In addition, consider two cumulative distribution functions for the required number of nurses T, $F_T^1(k)$ and $F_T^2(k)$ such that $F_T^1(k) \geq F_T^2(k)$ for all $k \in \mathcal{N}^+$, and let $y^{*,i}$ be the optimal staffing levels corresponding to $F_T^i(k)$, $i = 1,2$. Then, $y^{*,1} \leq y^{*,2}$, provided that (27) holds for any $y \geq y^{*,2}$.

The sufficient condition (27) states, intuitively, that the increasing marginal shape of the staffing cost function with respect to the number of scheduled nurses is preserved under endogenous absenteeism if the absenteeism rate is not too high, so that (23) is not too different from the cost function in (11). In particular, this sufficient condition requires that the absenteeism rate function is limited from above by two separate bounds. The first bound implies that the expected number of absent nurses does not exceed 2 irrespective of the number of nurses actually scheduled for work. For the absenteeism rate function (20) this is ensured by the following restriction on the values of α and β:

LEMMA 1. *For $\gamma(y) = \frac{1}{1+e^{\alpha+\beta y}}$, with $\alpha, \beta \geq 0$,*

$$2\beta e^{1+\alpha+2\beta} \geq 1 \Rightarrow y\gamma(y) \leq 2. \tag{28}$$

In the ED we studied, the estimated values of $\alpha = 1.533$ and $\beta = 0.092$ are high enough to satisfy (28). In particular, the maximum value of the product of number of scheduled nurses y and the estimated absenteeism rate calculated using these values is equal to 0.804, well below 2.

The second bound on the right-hand side of (27) takes the form of an effective absenteeism rate function which depends exclusively on the distribution of the targeted nursing level. Note that $\gamma_T(y) \geq 0$ if and only if

$$\frac{p_T(y-1)}{\sum_{k=y-1}^{\infty} p_T(k)} \geq \frac{1}{y}. \tag{29}$$

The expression of the left-hand side of (29) is the hazard rate function for the distribution of the targeted nursing level. Thus, (29) stipulates that the bound described by (27) is meaningful only in settings where such hazard rate evaluated at y exceeds $\frac{1}{y+1}$: for example, for the discrete uniform distribution on $[0, T_{\max}]$, (29) is satisfied if and only if $y \geq \frac{T_{\max}}{2} + 1$, i.e., when the number of scheduled nurses exceeds the expected need for nurses. Note that the constraint $\gamma(y) \leq \gamma_T(y)$ implies, in the same spirit as (28), the lower-bound restriction on the values of α and β:

$$\gamma(y) \leq \gamma_T(y) \Leftrightarrow \alpha + \beta y \geq \log\left(\frac{1-\gamma_T(y)}{\gamma_T(y)}\right). \tag{30}$$

For the case when the targeted nursing level is given by the discrete uniform distribution on $[0, T_{\max}]$, $\gamma_T(y) = 1 - \min\left(1, \left(\frac{T_{\max}-y+3}{y+1}\right)^{\frac{1}{y-1}}\right)$, and is an increasing function of y. Thus, for given values of α and β, (27) is satisfied provided that the number of scheduled nurses is high enough: there exists a critical threshold $y_U(T_{\max})$ such that $\gamma(y) \leq \gamma_T(y)$ for any $y \geq y_U(T_{\max})$. Figure 2

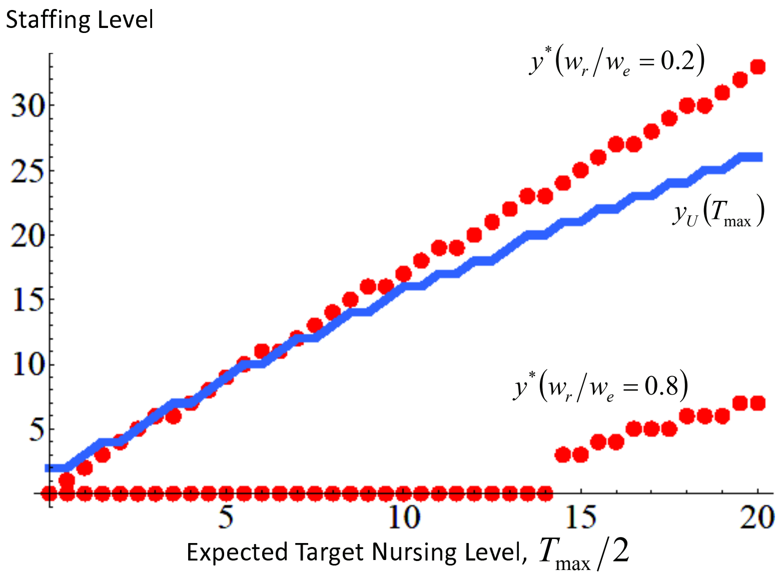


Figure 2: $y_U(T_{\max})$ and the optimal staffing levels for $\frac{w_r}{w_e}=0.2$ and $\frac{w_r}{w_e}=0.8$ under discrete uniform distribution for the target nursing level as functions of the expected target nursing level $\frac{T_{\max}}{2}$ ($\alpha=1.533, \beta=0.092$).

shows the bound $y_U(T_{\max})$ as well as the optimal staffing levels for two values of the cost ratio $\frac{w_r}{w_e}=0.2$ and $\frac{w_r}{w_e}=0.8$ as functions of $T_{\max}$ for $\alpha=1.533$ and $\beta=0.092$. Note that in settings with a substantial cost premium for extra/overtime nursing capacity the sufficient conditions of Proposition 2 are satisfied even for moderate values of expected demand for nurses (in particular, for $\frac{w_r}{w_e}=0.2$, (27) holds when $\frac{T_{\max}}{2}\geq 4$). On the other hand, in settings where the costs of regular and extra/overtime staffing are comparable, (27) may turn out to be overly restrictive.

3.3. Endogenous Absenteeism: Myopic Staffing with Learning vs. Optimal Staffing

In this section, we compare the optimal nurse staffing levels with those made by a clinical unit which treats the absenteeism rate as being exogenous and employs a trial-and-error procedure we label "myopic staffing with learning". In this latter case, the clinical unit selects staffing level y^{ML} such that

$$y^{\mathrm{ML}}=\min\left(y\in\mathcal{N}^+\Big|\sum_{k=0}^{y}F_T(k)p(k;y,\gamma(y^{\mathrm{ML}}))\geq 1-\frac{w_r}{w_e(1-\gamma(y^{\mathrm{ML}}))}\right), \tag{31}$$

with $\gamma(y)$ defined by (20). Note that (31) reflects a self-consistent way of selecting the staffing level: y^{ML} is the best staffing decision in the setting where the absenteeism rate is exogenous and determined by $\gamma(y^{\mathrm{ML}})$. In other words, a clinical unit assuming that the absenteeism rate is given by constant value $\gamma(y^{\mathrm{ML}})$ will respond by scheduling y^{ML} nurses and, as a result of this decision, will observe exactly the same value of the absenteeism rate, even if the true absenteeism process is endogenous and described by $\gamma(y)$. An intuitive way of rationalizing the choice of y^{ML} is to consider a sequence of "exogenous" staffing levels $y_n, n \in \mathcal{N}^+$, such that

$$y_{n+1} = \min\left(y \in \mathcal{N}^+ | \sum_{k=0}^{y} F_T(k) p(k; y, \gamma(y_n)) \geq 1 - \frac{w_r}{w_e(1-\gamma(y_n))} \right), n \in \mathcal{N}^+. \tag{32}$$

Equation (32) reflects a sequence of repeated adjustments of staffing levels, starting with some y_0, each based on the value of the absenteeism rate observed after the previously chosen staffing level is implemented. In this updating scheme, y^{ML} can be thought of as the limit $\lim_{n\to\infty} y_n$, if such limit exists. It is important to note that for a general demand distribution $F_T(k)$ and a general absenteeism rate function $\gamma(y)$, the set of staffing levels E satisfying (31) may be empty or may contain multiple elements. The analysis of existence and uniqueness of y^{ML} is further complicated by the discrete nature of staffing levels. In the following discussion, we by-pass this analysis and assume that there exists at least one staffing level satisfying (31). As the following result shows, even if E contains multiple elements, each of them is bounded from above by the optimal "endogenous" staffing level.

Proposition 3. *Consider y^* defined by (26) and suppose that a) (27) holds, and b) the set of staffing levels satisfying (31), E, is non-empty. Then,*

$$y^{\mathrm{ML}} \leq y^* \tag{33}$$

for any $y^{\mathrm{ML}} \in E$.

Proposition 3 implies that ignoring the endogenous nature of absenteeism can lead to understaffing in settings where the endogenous absenteeism rate declines with the number of scheduled nurses. Figure 3 illustrates the degree of such understaffing as a function of the cost ratio $\frac{w_r}{w_e}$ assuming a discrete uniform targeted nursing level distribution on $[0, T_{\max}]$, $T_{\max} = 20$, for different combinations of absenteeism rate parameters α and β. Cost ratio values were varied from $\frac{w_r}{w_e} = 0.01$ to

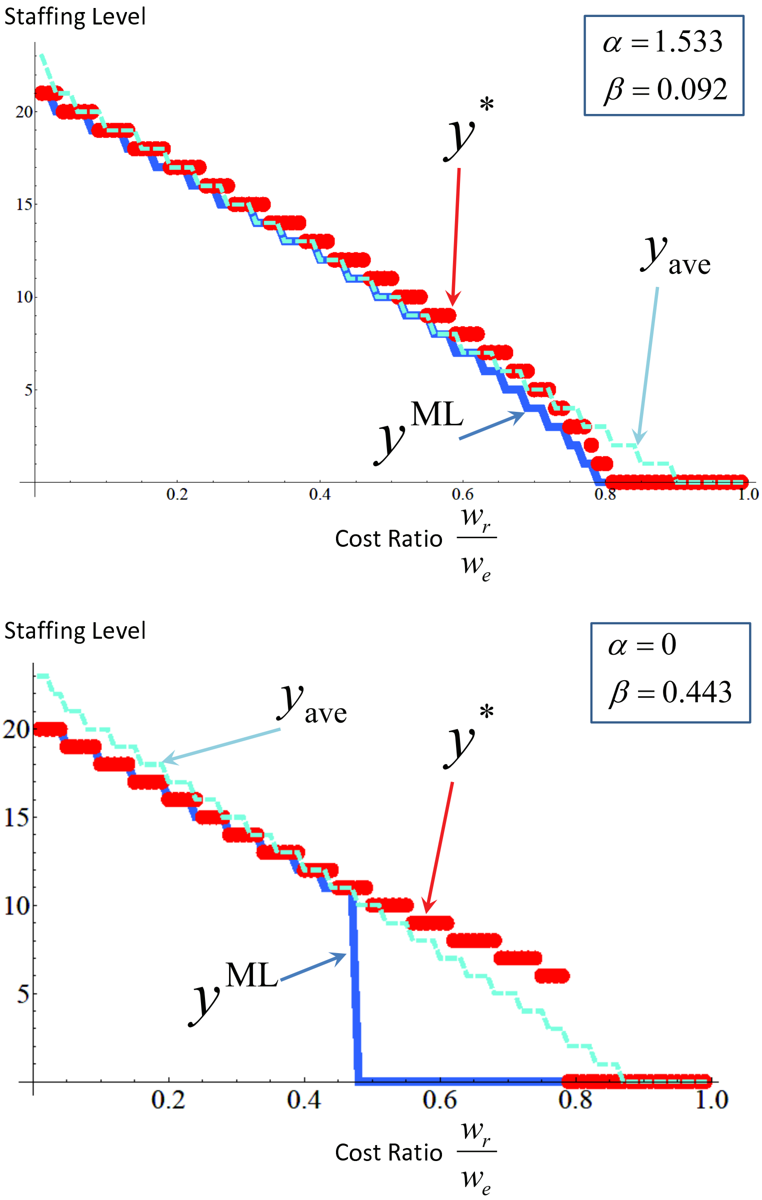


Figure 3 Optimal staffing levels y^*, self-consistent staffing levels y^{ML}, and "exogenous" staffing levels y_{ave} as functions of the cost ratio $\frac{w_r}{w_e}$ under discrete uniform distribution for the targeted nursing level with the support on $[0, T_{\max}]$, $T_{\max} = 20$, for different combinations of absenteeism rate parameters.

$\frac{w_r}{w_e} = 0.99$ in the steps of 0.01, and the search for the optimal number of nurses was conducted on the interval of integers between 0 and 40. Figure 3 also shows nurse staffing levels y_{ave} obtained under the exogenous absenteeism rate γ_{ave} computed by averaging the estimated absenteeism rate function $\gamma(y) = \frac{1}{1+e^{\alpha+\beta y}}$ over the nurse staffing levels observed in our empirical study:

$$\gamma_{\text{ave}} = \frac{1}{y_{\max} - y_{\min} + 1} \left(\sum_{y=y_{\min}}^{y_{\max}} \frac{1}{1 + e^{\alpha + \beta y}} \right), \tag{34}$$

where $y_{\min} = 0$ and $y_{\max} = 20$ reflect the smallest and largest possible targeted nursing level realizations. Note that for the values estimated from our data, $\alpha = 1.533$ and $\beta = 0.092$, $\gamma_{\text{ave}} = 0.09$. As Figure 3 shows, both the ML and the "exogenous" approaches may be off by one or two nurses as compared to the optimal staffing levels. This staffing gap may be significant in cases where the cost ratio $\frac{w_r}{w_e}$ is relatively high and the overall number of staffed nurses is low. Note that in the absenteeism settings described by relatively high values of α and relatively low values of β the magnitude of derivative of the absenteeism rate over the number of scheduled nurses, $\beta\gamma(y)(1-\gamma(y))$ is made insignificant by low values of both β and $\gamma(y)$ (the latter one being a result of high α). Thus, in such settings, the absenteeism rate function is nearly constant. Figure 3 also shows the staffing decisions in an alternative absenteeism environment: for $\alpha = 0$ and $\beta = 0.432$, the value of the average absenteeism rate computed using (34) is still 0.09, and yet the shape of the absenteeism rate function cannot be well approximated by any constant absenteeism rate. As a result, large staffing gaps may open up: the maximum staffing gap is 5 under the exogenous policy, and 11 under the ML policy. Figure 4 shows the relative cost gaps for the exogenous and the ML policies for the same combinations of model parameters: while for $\alpha = 1.533$ and $\beta = 0.092$ maximum cost gaps were 0.5% and 6.6% for the ML and the exogenous policies, respectively, a shift to a higher degree of absenteeism endogeneity ($\alpha = 0$ and $\beta = 0.432$) results in a potential performance deterioration for both policies (maximum performance gaps are now 26% for the ML policy and 13% for the endogenous policy).

4. Discussion

In our empirical study, we have used observations from a large urban hospital ED to study nurse absenteeism behavior at the shift level. We find that nurse absenteeism is exacerbated when fewer nurses are scheduled for a particular shirt. This is consistent with nurses exhibiting aversion to anticipated workload. We do not find any evidence to support the hypothesis that unusually high

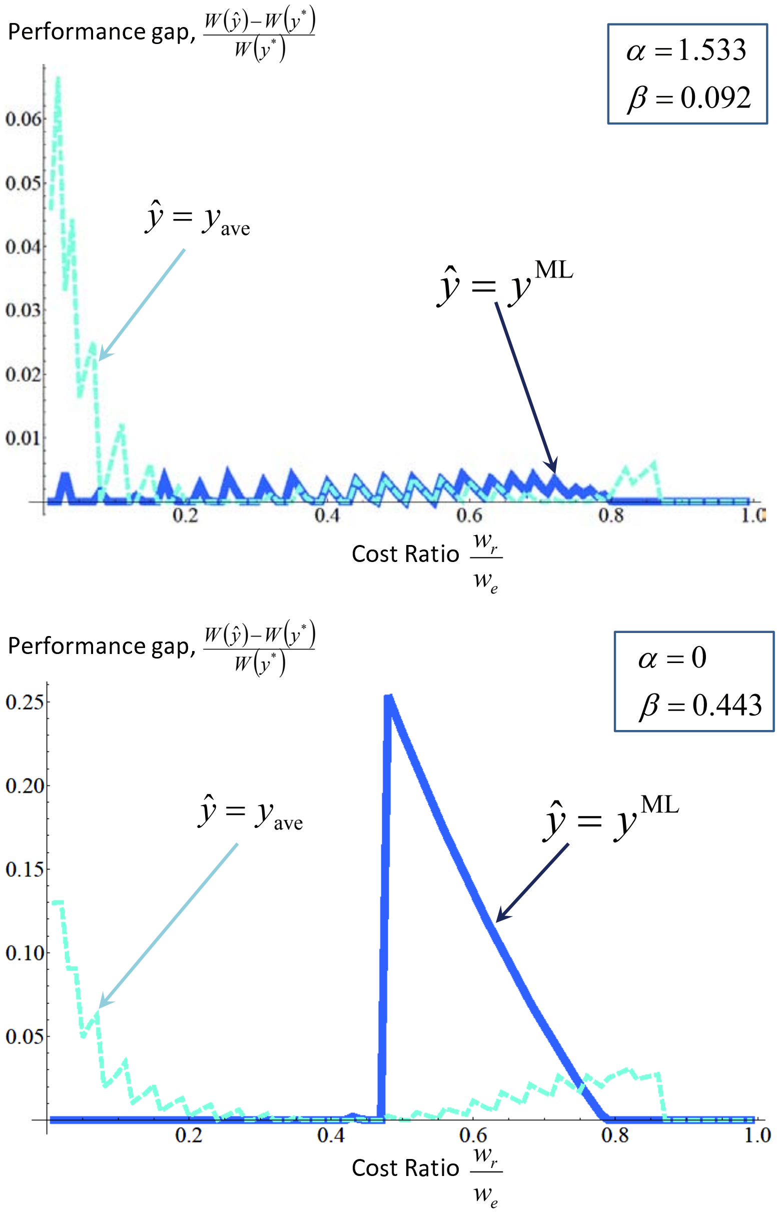


Figure 4: Performance gaps between the optimal staffing policy and the "myopic-with-learning" (ML) and "exogenous" policies as functions of the cost ratio $\frac{w_r}{w_e}$ under discrete uniform distribution for the target nursing level with the support on $[0, T_{\max}]$, $T_{\max} = 20$, for different combinations of absenteeism rate parameters.

workloads in previous shifts affect absenteeism. This might be due to the fact that the shift level data we use does not allow us to measure individual nurse workloads.

On the analytical front, we demonstrated that the failure to properly account for the endogenous nature of nurse absenteeism can lead to significant deviations from optimal staffing decisions. In particular, in settings where higher numbers of scheduled nurses result in lower absenteeism rates, a clinical unit which treats absenteeism as an exogenous phenomenon, will often under-supply nursing staff capacity even if allowed to repeatedly adjust its staffing decisions in response to observed absenteeism rates. Such understaffing may result in substantial cost increases even in settings with low absenteeism rates, as long as absenteeism exhibits a substantial degree of endogeneity.

It is important to note that our assumption about the unlimited availability of extra/overtime nursing capacity may not be valid in some clinical environments. In such environments, it may be impossible to replace absent nurses at a reasonable cost or in reasonable time, and the endogeneity of absenteeism can lead to significant understaffing with the possibility of serious deterioration of service quality and longer ED delays. In other clinical units, the use of agency nurses who may be less familiar with the unit can lead to similar declines in quality of patient care and in an increase in the rate of medical errors. Thus, the accurate understanding of the nature of nurse absenteeism and the use of a model that accurately incorporates this phenomenon in determining appropriate staffing levels is imperative to assuring high quality patient care.

References

Aiken, L. H., S.P. Clarke, D.M. Sloane, J. Sochalski, J.H. Silber. 2002. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *Journal of American Medical Association* **288** 1987–1993.

Aksin, Z., M. Armony, V. Mehrotra. 2007. The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management* **16**(6) 665–688.

Bakker, A. B., E. Demerouti, W. B. Schaufeli. 2003. Dual processes at work in a call center: An application of the job-demands-resources model. *European Journal of Work and Organizational Psychology* **12**(4) 393–428.

Beja, A. 1977. Optimal reject allowance with constant marginal production efficiency. *Naval Research Logistics Quarterly* **24**(1) 21–33.

Bollapragada, S., T. E. Morton. 1999. Myopic heuristics for the random yield problem. *Operations Research* **47**(5) 713–722.

Brown, S., J. Sessions. 1996. The economics of absence: Theory and evidence. *Journal of Economic Surveys* **10**(1) 23–31.

Bryant, C., G. Fairbrother, P. Fenton. 2000. Professional issues: The relative influence of personal and workplace descriptors on stress. *British Journal of Nursing* **9** 876–880.

California Department of Health. 2004. Nurse-to-patient staffing ratio regulations. http://www.cdph.ca.gov/services/DPOPP/regs/Documents/R-37-01_Regulation_Text.pdf, last checked on May 24, 2010.

Cho, S.H., S. Ketefian, V.H. Barkauskas. 2003. The effects of nurse staffing on adverse outcomes, morbidity, mortality, and medical costs. *Nursing Research* **52**(2) 71–79.

Ciarallo, F.W., R. Akella, T. Morton. 1994. A periodic review, production planning model with uncertain capacity and uncertain demand- optimality of extended myopic policies. *Management Science* **40** 320–332.

Darr, W., G. Johns. 2008. Work strain, health, and absenteeism: A meta-analysis. *Journal of Occupational Health Psychology* **13**(4) 293–318.

Demerouti, E., A. Bakker, F. Nachreiner, W.B. Schaufeli. 2000. A model of burnout and life satisfaction amongst nurses. *Journal of Advanced Nursing* **32** 454–464.

Diwas, KC, C. Terwiesch. 2009. Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science* **55**(9) 1486–1498.

Easton, F.F., J.C. Goodale. 2002. Labor scheduling with employee turnover and absenteeism. *Working Paper, Syracuse University* .

Easton, F.F., J.C. Goodale. 2005. Schedule recovery: Unplanned absences in service operations. *Decision Sciences* **36**(3) 459–488.

Ehrhardt, R., L. Taube. 1987. An inventory model with random replenishment quantities. *Int. J. Prod. Res.* **25**(12) 1795–1804.

Erickson, R.J., L. Nichols, C. Ritter. 2000. Family influences on absenteeism: Testing an expanded process model. *Journal of Vocational Behavior* **57** 246–272.

Fadiloglu, M., E. Berk, M. C. Gurbuzc. 2008. Supplier diversification under binomial yield. *Operations Research Letters* **36** 539–542.

Fagerstrom, L., A.-K. Rainio. 1999. Professional assessment of optimal nursing care intensity – a new method for assessment of staffing levels for nursing care. *Journal of Clinical Nursing* **8** 369–379.

Gerchak, Y., M. I. Henig. 1994. A flexible conceptualization of random yield and its implications for source selection. *Proceedings of the First Conference of ORSA Technical Section on Manufacturing Management, Carnegie Mellon University* 133–139.

Gillies, D.A. 1994. *Nursing Management: A Systems Approach.* W.B. Saunders.

Giovannetti, P., J.M. Johnson. 1990. A new generation patient classification system. *Journal of Nursing Administration* **20**(5) 33–40.

Greene, W.H. 2003. *Econometric analysis.* Pearson Education.

Grosfeld-Nir, A., Y. Gerchak. 2004. Multiple lotsizing in production to order with random yields: Review of recent advances. *Annals of Operations Research* **126** 43–69.

Gupta, D., W. L. Cooper. 2005. Stochastic comparisons in production yield management. *Operations Research* **53**(2) 377–384.

Healy, C., M.F. McKay. 1999. Identifying sources of stress and job satisfaction in the nursing environment. *Australian Journal of Advanced Nursing* **17** 30–35.

Healy, C., M.F. McKay. 2000. Nursing stress: the effect of coping strategies and job satisfaction in a sample of Australian nurses. *Journal of Advanced Nursing* **31** 681–688.

Henig, M., Y. Gerchak. 1990. The structure of periodic review policies in the presence of variable yield. *Operations Research* **38** 634–643.

Ho, J.T.S. 1997. Corporate wellness programmes in Singapore: Effect on stress, satisfaction and absenteeism. *Journal of Managerial Psychology* **12** 177–189.

Jamal, M. 1984. Job stress and job performance controversy: An empirical assessment. *Organizational Behavior and Human Performance* **33** 1–21.

Kovner, C., P.J. Gergen. 1998. Nurse staffing levels and adverse outcomes following surgery in U.S. hospitals. *Journal of Nursing Scholarship* **30**(4) 315–321.

McVicar, A. 2003. Workplace stress in nursing: a literature review. *Journal of Advanced Nursing* **44** 633–642.

Needleman, J., P. Buerhaus, S. Mattke. 2002. Nurse-staffing levels and patient outcomes in hospitals. *New England Journal of Medicine* **346**(22) 1715–1722.

Nevidjon, B., J. Ives Erickson. 2001. The nursing shortage: Solutions for the short and long term. *Online Journal of Issues in Nursing* .

Noori, A.H., G. Keller. 1986. One-period order quantity strategy with uncertain match between the amount received and quantity requisitioned. *INFOR* **24**(1) 1–11.

Porteus, E.L. 2002. *Foundations of Stochastic Inventory Theory*. Stanford Business Books.

Powell, S. G., K.L. Schultz. 2004. Throughput in serial lines with state-dependent behavior. *Management Science* **50** 1095–1105.

Rauhala, A., M. Kivimäki, L. Fagerström, M. Elovainio, M. Virtanen, J. Vahtera, A.-K. Rainio, K. Ojaniemi, J. Kinninen. 2007. What degree of work overload is likely to cause increased sickness absenteeism among nurses? Evidence from the RAFAELA patient classification system. *Journal of Advanced Nursing* **57**(3) 286–295.

Rekik, R., E. Sahin, Y. Dallery. 2007. A comprehensive analysis of the newsvendor model with unreliable supply. *OR Spectrum* **29**(2) 207–233.

Schultz, K.L., D.C. Juran, J.W. Boudreau. 1999. The effects of low inventory on the development of productivity norms. *Management Science* **45** 1664–1678.

Schultz, K.L., D.C. Juran, J.W. Boudreau, L.J. Thomas J.O. McClain. 1998. Modeling and worker motivation in JIT production systems. *Management Science* **44** 1595–1607.

Scott, K.D., J.C. Wimbush. 1991. Teacher absenteeism in secondary education. *Educational Administration Quarterly* **27**(4) 506–529.

Seago, J.A. 2001. Chapter 39 in "Making health care safer: A critical analysis of patient safety practices". *Evidence Report/Technology Assessment Number 43, AHRQ Publication No. 01-E058, Agency for Healthcare Research and Quality* .

Seago, J.A. 2002. A comparison of two patient classification instruments in an acute care hospital. *Journal of Nursing Administration* **32**(5) 243–249.

Shamian, J., L. O'Brien-Pallas, D. Thomson, C. Alksnis, M.S. Kerr. 2003. Nurse absenteeism, stress and workplace injury: What are the contributing factors and what can/should be done about it? *The International Journal of Sociology and Social Policy* **23**(8/9) 81–103.

Statistics Canada. 2008. Work absence rates. http://www.statcan.gc.ca/pub/71-211-x/71-211-x2009000-eng.pdf, last checked on May 24, 2010.

The Victorian Department of Health. 2007. Multiple business agreement 2007-2011. http://www.health.vic.gov.au/__data/assets/pdf_file/0008/356696/Nurses-Public-Sector-EBA-2004-2007-_AG840794-2.pdf, last checked on May 24, 2010.

Tummers, G.E.R., P.P.M. Janssen, A. Landeweerd, I. Houkes. 2001. A comparative study of work characteristics and reactions between general and mental health nurses: a multi-sample analysis. *Journal of Advanced Nursing* **36**(1) 151–162.

Unruh, L. 2003. Licensed nurse staffing and adverse outcomes in hospitals. *Medical Care* **41**(1) 142–152.

Unruh, L., L. Joseph, M. Strickland. 2007. Nurse absenteeism and workload: Negative effect on restraint use, incident reports and mortality. *Journal of Advanced Nursing* **60**(6) 673–681.

US Bureau of Labor Statistics. 2008. Industry injury and illness data. http://www.bls.gov/iif/oshwc/osh/os/osnr0032.pdf, last checked on May 24, 2010.

Vingard, E., P. Lindberg, M. Josephson, M. Voss, B. Heijbel, L. Alfredsson, S. Stark S, A. Nygren. 2005. Long-term sick-listing among women in the public sector and its associations with age, social situation, lifestyle, and work factors: a three-year follow-up study. *Scandinavian Journal of Public Health* **33**(5) 370–375.

Voss, M., B. Floderus, F. Diderichsen. 2001. Physical, psychosocial, and organisational factors relative to sickness absence: a study based on Sweden post. *Occupational and Environmental Medicine* **58** 178–184.

Whitt, W. 2006. Staffing a call center with uncertain arrival rate and absenteeism. *Production and Operations Management* **15**(1) 88–102.

Yang, C.A., H.A. Lee. 1995. Lot sizing with random yields: a review. *Operations Research* **43**(2) 311–334.

Yang, S., J. Yang, L. Abdel-Malek. 2007. Sourcing with random yields and stochastic demand: A newsvendor approach. *Computers and Operations Research* **34** 3682–3690.

Yankovic, N., L.V. Green. 2010. Identifying good nursing levels: A queueing approach. *Operations Research* forthcoming.

Appendix

Estimating Absenteeism Rates Using Alternative Workload Anticipation Models

In addition to the models estimated in Section 2 which involved two definitions of the anticipated workload variable (ENPR_t^1 and ENPR_t^2), we also test six alternative definitions of anticipated workload.

- Under the alternatives $i=3$ and $i=4$, the denominator of ENPR_t^i is set to the expectation of patient census values computed only over shifts of the same type as t (i.e., over all day shifts if t corresponds to a day shift, or over all night shifts, if t corresponds to a night shift). Furthermore, ENPR_t^3 uses the actual number of nurses scheduled in the evening shift while ENPR_t^4 uses the expected number of nurses scheduled in the evening shift, where this expectation is also computed only over shifts of the same type. Under this alternative, nurses distinguish explicitly between day and night shifts, but not between days of the week, when calculating the expected nurse-to-patient ratio for the upcoming shift. Any differences between days of the week are captured by the weekday dummy.

- Under the alternative $i=5$ and $i=6$, the denominator of ENPR_t^i is set to the expectation of patient census values computed only over shifts that belong to the same type of day as t (i.e., over all Monday shifts if t corresponds to a Monday shift, etc.). Furthermore, ENPR_t^5 uses the actual number of nurses scheduled in the evening shift while ENPR_t^6 uses the expected number of nurses scheduled in the evening shift, where this expectation is also computed only over shifts that belong to the same day. Under this alternative, when calculating the expected nurse-to-patient ratio for the upcoming shift, nurses distinguish explicitly between days of the week, but not between day shifts or night shifts. Any differences between the day and the night shift are captured by the shift dummy.

- Under the alternatives $i=7$ and $i=8$, the denominator of ENPR_t^i is set to the expectation of patient census values computed only over shifts that belong to the same type and the same type of day as t (i.e., over all Monday day shifts if t corresponds to a day shift on Monday, etc.).

Furthermore, ENPR_t^7 uses the actual number of nurses scheduled in the evening shift while ENPR_t^8 uses the expected number of nurses scheduled in the evening shift, where this expectation is also computed only over shifts that belong to the same day and type. Under this alternative, for a given value of y_t, scheduled nurses anticipate different nurse-to-patient ratios, depending on both day of the week and the type of the shift (day or night) that t corresponds to.

These six "workload anticipation" alternatives along with the alternatives $i = 1, 2$ estimated in Section 2 correspond to eight mental models nurses may potentially use for estimating workload during an upcoming shift, with the least sophisticated among them being alternative 2, followed by 1 then 4, 3, 6, 5, 8, and 7 which is the most sophisticated alternative. Besides workload related variables, we also include the fixed effects of Section 2. Namely, we include dayshift, holiday and day of the week dummies as well as week fixed effects. As in Section 2, for each of the 6 new definitions of the anticipated workload, we will use our data to estimate two different logit models. The first model includes the 14 lag workload variables while the second does not. Specifically, further to the four models estimated in Section 2, the 12 models we will estimate are:

$$\begin{aligned}\text{logit}(\gamma_t) = {} & \beta_0^i + \beta_{\text{ENPR}}^i \times \text{ENPR}_t^i + \sum_{j=1}^{14} \beta_{\text{NPR},j}^i \times \text{NPR}_{t-j} + \sum_{d=2}^{7} \beta_{\text{DAY},d}^i \times \text{DAY}_{d,t} \\ & + \sum_{f=2}^{44} \beta_{\text{W},f}^i \times W_{f,t} + \beta_{\text{DAYSHIFT}}^i \times \text{DAYSHIFT}_t + \beta_{\text{HOLIDAY}}^i \times \text{HOLIDAY}_t\end{aligned}$$

where $i \in \{3, ..., 8\}$ refers to the definition of ENPR^i used, $\text{DAY}_{d,t}$ and $W_{f,t}$ are the day and week fixed effects, $DAYSHIFT_t$ and HOLIDAY_t are shift and holiday fixed effects.

We report the estimation results for these 12 models in the enclosed Table 3. We only report the marginal effects of the estimation on the probability of absenteeism. In all of the models estimated, the lagged nurse-to-patient variables are not jointly significant at the 10% confidence level, although individually for some models the lag 6, 10 and 14 variable appear to be significant. In 11 out of the 12 models, the anticipated workload is significant at the 5% or the 10% level with a negative coefficient with a statistically similar magnitude to the coefficient of the anticipated workload variable reported in Section 2. We also find that the holiday dummy variable is significant at the 1% level and negative, suggesting that nurses are less likely to be absent on a holiday.

Interestingly, in Models VI, VIII, XIV and XVI, where the anticipated workload is adjusted according to the shift, we find that the shift fixed effect is no longer significant. A possible explanation for this is that the difference in absenteeism between the day and the night shift is completely explained by the difference in anticipated workload between the two shifts. In contrast, in all models where we omit the lagged workload variables (i.e. all "even-numbered" models), including Models X, XII, XIV and XVI, where the anticipated workload variable is adjusted according to the day of the week, we still find that there is a significant difference between weekends and weekdays. Nurses are more likely to be absent on the weekend, and this increase in the probability of absenteeism is not explained by the differences in workloads between the weekdays and the weekends.

Proof of Proposition 1

Note that $W_{\text{ave}}(y+1)-W_{\text{ave}}(y)=w_r+w_e\left(L(y+1,\gamma_{\text{ave}})-L(y,\gamma_{\text{ave}})\right)=w_r+w_e\Delta L(y,\gamma_{\text{ave}})$. We will establish the result of the proposition by showing that $\Delta L(y,\gamma_{\text{ave}})\le 0$ and

$\Delta^2 L(y,\gamma_{\text{ave}})=L(y+2,\gamma_{\text{ave}})-2L(y+1,\gamma_{\text{ave}})+L(y,\gamma_{\text{ave}})\ge 0$. First, note that

$$q(k+1)-q(k)=\sum_{n=k+1}^{\infty}(n-k-1)p_T(n)-\sum_{n=k}^{\infty}(n-k)p_T(n)=-\sum_{n=k+1}^{\infty}p_T(n)\le 0, \tag{A1}$$

and

$$q(k+2)-2q(k+1)+q(k)=-\sum_{n=k+2}^{\infty}p_T(n)+\sum_{n=k+1}^{\infty}p_T(n)=p_T(k+1)\ge 0. \tag{A2}$$

Then,

$$\begin{aligned}\Delta L(y,\gamma_{\text{ave}}) &= L(y+1,\gamma_{\text{ave}})-L(y,\gamma_{\text{ave}})=\sum_{k=0}^{y+1}q(k)p(k;y+1,\gamma_{\text{ave}})-\sum_{k=0}^{y}q(k)p(k;y,\gamma_{\text{ave}})\\ &=\sum_{k=0}^{y+1}q(k)(p(k;y+1,\gamma_{\text{ave}})-p(k;y,\gamma_{\text{ave}})),\end{aligned} \tag{A3}$$

where we have used $p(y+1;y,\gamma_{\text{ave}}))\equiv 0$. Next, using the following property of the binomial distribution

$$p(k;y+1,\gamma_{\text{ave}})=(1-\gamma_{\text{ave}})p(k-1;y,\gamma_{\text{ave}})+\gamma_{\text{ave}}p(k;y,\gamma_{\text{ave}}),k=0,\dots,y \tag{A4}$$

Variable	MODEL V	MODEL VI	MODEL VII	MODEL VIII	MODEL IX	MODEL X	MODEL XI	MODEL XII	MODEL XIII	MODEL XIV	MODEL XV
	Marginal Effect$^+$ (robust error)	Marginal Effect$^+$ (robust error)	Marginal Effect$^+$ (robust error)	Marginal Effect$^+$ (robust error)	Marginal Effect$^+$ (robust error)	Marginal Effect$^+$ (robust error)	Marginal Effect$^+$ (robust error)	Marginal Effect$^+$ (robust error)	Marginal Effect$^+$ (robust error)	Marginal Effect$^+$ (robust error)	Marginal Effect$^+$ (robust error)
ENPR^3	-0.630*	-0.531									
	(0.361)	(0.340)									
ENPR^4			-0.928**	-0.799**							
			(0.402)	(0.376)							
ENPR^5					-0.623*	-0.547*					
					(0.345)	(0.321)					
ENPR^6							-0.933**	-0.830**			
							(0.388)	(0.359)			
ENPR^7									-0.665**	-0.602*	
									(0.338)	(0.311)	
ENPR^8											-0.917**
											(0.374)
NPR1	-0.0126		-0.0266		-0.0305		-0.0519		-0.0477		-0.0767
	(0.191)		(0.192)		(0.192)		(0.193)		(0.193)		(0.194)
NPR2	0.151		0.183		0.166		0.205		0.182		0.224
	(0.215)		(0.216)		(0.216)		(0.217)		(0.217)		(0.218)
NPR3	-0.0289		-0.0410		-0.0345		-0.0494		-0.0492		-0.0663
	(0.187)		(0.186)		(0.187)		(0.187)		(0.188)		(0.188)
NPR4	0.206		0.228		0.199		0.219		0.193		0.209
	(0.204)		(0.204)		(0.204)		(0.204)		(0.203)		(0.203)
NPR5	-0.163		-0.181		-0.158		-0.172		-0.143		-0.150
	(0.189)		(0.189)		(0.188)		(0.188)		(0.188)		(0.188)
NPR6	0.371*		0.368*		0.365*		0.358*		0.358*		0.347*
	(0.202)		(0.202)		(0.202)		(0.203)		(0.203)		(0.203)
NPR7	0.0300		0.0341		0.0332		0.0403		0.0333		0.0404
	(0.197)		(0.197)		(0.197)		(0.197)		(0.197)		(0.197)
NPR8	0.0989		0.101		0.0945		0.0936		0.0927		0.0907
	(0.203)		(0.203)		(0.203)		(0.203)		(0.203)		(0.203)
NPR9	-0.260		-0.268		-0.254		-0.261		-0.253		-0.256
	(0.187)		(0.187)		(0.187)		(0.187)		(0.187)		(0.187)
NPR10	0.355*		0.360*		0.351*		0.354*		0.344*		0.340*
	(0.200)		(0.200)		(0.199)		(0.199)		(0.199)		(0.199)
NPR11	-0.209		-0.207		-0.199		-0.191		-0.187		-0.172
	(0.186)		(0.187)		(0.187)		(0.187)		(0.187)		(0.188)
NPR12	0.156		0.172		0.152		0.166		0.155		0.167
	(0.205)		(0.205)		(0.204)		(0.204)		(0.205)		(0.205)
NPR13	-0.186		-0.175		-0.183		-0.170		-0.188		-0.180
	(0.189)		(0.189)		(0.189)		(0.189)		(0.189)		(0.189)
NPR14	0.308*		0.302		0.315*		0.310*		0.324*		0.322*
	(0.186)		(0.186)		(0.186)		(0.186)		(0.186)		(0.185)
Dayshift	-0.00273	0.00626	-0.00188	0.00735	0.00721	0.0151*	0.0132	0.0210**	-0.00194	0.00725	-0.000735
	(0.00697)	(0.00620)	(0.00702)	(0.00627)	(0.00945)	(0.00896)	(0.0102)	(0.00972)	(0.00702)	(0.00625)	(0.00710)
Holiday	-0.0383***	-0.0365***	-0.0382***	-0.0363***	-0.0383***	-0.0365***	-0.0381***	-0.0363***	-0.0381***	-0.0364***	-0.0379***
	(0.0135)	(0.0131)	(0.0136)	(0.0131)	(0.0135)	(0.0131)	(0.0136)	(0.0131)	(0.0136)	(0.0131)	(0.0137)
Tuesday	-0.00336	0.00145	-0.00318	0.00115	-0.000365	0.00403	0.00125	0.00503	-0.000143	0.00422	0.000343
	(0.0137)	(0.0122)	(0.0137)	(0.0122)	(0.0139)	(0.0124)	(0.0141)	(0.0124)	(0.0140)	(0.0124)	(0.0140)
Wednesday	0.00325	0.00256	0.00380	0.00241	0.00846	0.00704	0.0117	0.00924	0.00871	0.00739	0.0102
	(0.0153)	(0.0123)	(0.0153)	(0.0123)	(0.0158)	(0.0127)	(0.0163)	(0.0129)	(0.0158)	(0.0127)	(0.0160)
Thursday	0.0113	0.00886	0.0122	0.00921	0.0189	0.0155	0.0238	0.0195	0.0193	0.0162	0.0224
	(0.0159)	(0.0129)	(0.0159)	(0.0129)	(0.0168)	(0.0137)	(0.0176)	(0.0143)	(0.0168)	(0.0138)	(0.0173)
Friday	0.0196	0.0164	0.0193	0.0164	0.0279	0.0236	0.0319*	0.0275*	0.0282	0.0243*	0.0308*
	(0.0175)	(0.0136)	(0.0173)	(0.0136)	(0.0183)	(0.0145)	(0.0189)	(0.0150)	(0.0183)	(0.0146)	(0.0187)
Saturday	0.0207	0.0239*	0.0211	0.0241*	0.0451*	0.0456**	0.0593**	0.0581**	0.0469*	0.0480**	0.0596**
	(0.0181)	(0.0141)	(0.0182)	(0.0141)	(0.0252)	(0.0213)	(0.0288)	(0.0239)	(0.0253)	(0.0212)	(0.0288)
Sunday	0.0151	0.0200	0.0166	0.0209	0.0352*	0.0382**	0.0484**	0.0498**	0.0366*	0.0401**	0.0467**
	(0.0156)	(0.0138)	(0.0158)	(0.0139)	(0.0209)	(0.0187)	(0.0241)	(0.0210)	(0.0210)	(0.0186)	(0.0234)
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6481	6631	6481	6631	6481	6631	6481	6631	6481	6631	6481
Log Likelihood	-1654	-1690	-1653	-1689	-1654	-1689	-1653	-1688	-1654	-1689	-1653
LR-test	102.6	92.20	104.8	94.47	103.0	92.73	105.6	95.59	103.5	93.64	105.7

***, **, * denote statistical significance at 1% 5% and 10% confidence levels, respectively. Robust errors are shown in parentheses.
+Marginal effects are calculated at the mean values of the independent variables. For the dummy variables the marginal effect shows the difference caused by changing the variable from 0 to 1.

Table 3: Estimation results for alternative ENPR models.

which assumes that $p(k-1;y,\gamma_{\text{ave}}) \equiv 0$, we get for (A3)

$$\sum_{k=0}^{y+1} q(k)(p(k;y+1,\gamma_{\text{ave}}) - p(k;y,\gamma_{\text{ave}}))$$

$$= (1-\gamma_{\text{ave}}) \sum_{k=0}^{y+1} q(k)\,(p(k-1;y,\gamma_{\text{ave}}) - p(k;y,\gamma_{\text{ave}}))$$

$$= (1-\gamma_{\text{ave}}) \sum_{k=0}^{y+1} \left(q(k+1)-q(k)\right) p(k;y,\gamma_{\text{ave}})$$

$$= (1-\gamma_{\text{ave}}) \sum_{k=0}^{y} \left(q(k+1)-q(k)\right) p(k;y,\gamma_{\text{ave}}) \le 0, \tag{A5}$$

where in the second line we have used

$$\begin{aligned}\sum_{k=0}^{y+1} q(k)p(k-1;y,\gamma_{\text{ave}}) &= \sum_{k=1}^{y+1} q(k)p(k-1;y,\gamma_{\text{ave}}) = \sum_{k=0}^{y} q(k+1)p(k;y,\gamma_{\text{ave}}) \\ &= \sum_{k=0}^{y+1} q(k+1)p(k;y,\gamma_{\text{ave}}).\end{aligned} \tag{A6}$$

Similarly,

$$\begin{aligned}\Delta^2 L(y,\gamma_{\text{ave}}) &= L(y+2,\gamma_{\text{ave}}) - 2L(y+1,\gamma_{\text{ave}}) + L(y,\gamma_{\text{ave}}) \\ &= (1-\gamma_{\text{ave}}) \left(\sum_{k=0}^{y+2} \left(q(k+1)-q(k)\right) p(k;y+1,\gamma_{\text{ave}}) - \sum_{k=0}^{y+1} \left(q(k+1)-q(k)\right) p(k;y,\gamma_{\text{ave}}) \right) \\ &= (1-\gamma_{\text{ave}}) \left(\sum_{k=0}^{y+2} \left(q(k+1)-q(k)\right) \left(p(k;y+1,\gamma_{\text{ave}}) - p(k;y,\gamma_{\text{ave}})\right) \right),\end{aligned} \tag{A7}$$

where we have used $p(y+2;y,\gamma_{\text{ave}}) \equiv 0$. Further, using (A4) in (A7) we get

$$\begin{aligned}&= (1-\gamma_{\text{ave}}) \left(\sum_{k=0}^{y+2} \left(q(k+1)-q(k)\right) \left(p(k;y+1,\gamma_{\text{ave}}) - p(k;y,\gamma_{\text{ave}})\right) \right) \\ &= (1-\gamma_{\text{ave}})^2 \left(\sum_{k=0}^{y+2} \left(q(k+1)-q(k)\right) \left(p(k-1;y,\gamma_{\text{ave}}) - p(k;y,\gamma_{\text{ave}})\right) \right).\end{aligned} \tag{A8}$$

Note that

$$\begin{aligned}&= \left(\sum_{k=0}^{y+2} \left(q(k+1)-q(k)\right) p(k-1;y,\gamma_{\text{ave}}) \right) = \left(\sum_{k=1}^{y+2} \left(q(k+1)-q(k)\right) p(k-1;y,\gamma_{\text{ave}}) \right) \\ &= \left(\sum_{k=0}^{y+1} \left(q(k+2)-q(k+1)\right) p(k;y,\gamma_{\text{ave}}) \right) = \left(\sum_{k=0}^{y+2} \left(q(k+2)-q(k+1)\right) p(k;y,\gamma_{\text{ave}}) \right),\end{aligned} \tag{A9}$$

where we have used $p(-1;y,\gamma_{\text{ave}}) \equiv p(y+1;y,\gamma_{\text{ave}}) \equiv p(y+2;y,\gamma_{\text{ave}}) \equiv 0$. Thus, using (A9), we obtain for (A8)

$$\Delta^2 L(y,\gamma_{\text{ave}}) = (1-\gamma_{\text{ave}})^2 \left(\sum_{k=0}^{y} \left(q(k+2)-2q(k+1)+q(k)\right) p(k;y,\gamma_{\text{ave}}) \right) \ge 0. \tag{A10}$$

Thus, since $\Delta L(y,\gamma_{\text{ave}})$ is non-positive and non-decreasing in y, we conclude that the smallest y for which $\Delta L(y,\gamma_{\text{ave}}) \ge -\frac{w_r}{w_e}$ is the global minimum of $W_{\text{ave}}(y)$. Note that because of the monotonicity of $\Delta L(y,\gamma_{\text{ave}})$ with respect to y, larger values of $\frac{w_r}{w_e}$ translate into equal or smaller values of y^*_{ave}. In addition, note that, combining (A1) and (A5), we get

$$\Delta L(y,\gamma_{\text{ave}}) = -\left((1-\gamma_{\text{ave}})\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma_{\text{ave}})\right), \tag{A11}$$

so that

$$\Delta L(y,\gamma_{\text{ave}}) \geq -\frac{w_r}{w_e} \tag{A12}$$

is equivalent to

$$\sum_{k=0}^{y} F_T(k)p(k;y,\gamma_{\text{ave}}) \geq 1 - \frac{w_r}{w_e(1-\gamma_{\text{ave}})}. \tag{A13}$$

Now, consider $F_T^1(k)$ and $F_T^2(k)$ such that $F_T^1(k) \geq F_T^2(k)$ for any $k \in \mathcal{N}^+$ and let $\Delta L^1(y,\gamma_{\text{ave}})$ and $\Delta L^2(y,\gamma_{\text{ave}})$ be the respective expressions on the left-hand side of (A11). Then, $\Delta L^1(y,\gamma_{\text{ave}}) \geq \Delta L^2(y,\gamma_{\text{ave}})$ for any $y \in \mathcal{N}^+$ and, respectively,

$$\begin{aligned} y^{*,1}_{\text{ave}} &= \min\left(y \in \mathcal{N}^+ | \Delta L^1(y,\gamma_{\text{ave}}) \geq -\frac{w_r}{w_e}\right) \\ &\leq \min\left(y \in \mathcal{N}^+ | \Delta L^2(y,\gamma_{\text{ave}}) \geq -\frac{w_r}{w_e}\right) = y^{*,2}_{\text{ave}}. \end{aligned} \tag{A14}$$

□

Proof of Corollary 1

Under the discrete uniform demand distribution specified by (17), the sum in the expression for the optimal staffing level (16) becomes

$$\sum_{k=0}^{y} F_T(k)p(k;y,\gamma_{\text{ave}}) = \begin{cases} \frac{y(1-\gamma_{\text{ave}})+1}{T_{\max}+1}, & \text{for } y \leq T_{\max}, \\ \frac{1}{T_{\max}+1}\sum_{k=0}^{T_{\max}}(k+1)p(k;y,\gamma_{\text{ave}}) + \sum_{k=T_{\max}+1}^{y} p(k;y,\gamma_{\text{ave}}), & y \geq T_{\max}+1. \end{cases} \tag{A15}$$

Note that for $y = T_{\max}$, (A15) becomes

$$\frac{T_{\max}(1-\gamma_{\text{ave}})+1}{T_{\max}+1} = 1 - \frac{\gamma_{\text{ave}}T_{\max}}{T_{\max}+1} \geq 1 - \frac{w_r}{w_e(1-\gamma_{\text{ave}})}, \tag{A16}$$

as long as

$$\frac{w_r}{w_e} \geq \gamma_{\text{ave}}(1-\gamma_{\text{ave}})\frac{T_{\max}}{T_{\max}+1}. \tag{A17}$$

The supremum of the right-hand side of (A17) is $\frac{1}{4}$ (for $\gamma_{\text{ave}} = 0.5$ and $T_{\max} \to \infty$), so that (A17) is implied by $\frac{w_r}{w_e} \geq \frac{1}{4}$. Thus, under this condition, the optimal staffing level does not exceed $T_{\max}$ and, consequently,

$$y^*_{\text{ave}} = \min\left(y \in \mathcal{N}^+ \mid \frac{y(1-\gamma_{\text{ave}})+1}{T_{\max}+1} \geq 1 - \frac{w_r}{w_e(1-\gamma_{\text{ave}})}\right) = \left\lceil \left(\frac{T_{\max}}{1-\gamma_{\text{ave}}} - \frac{T_{\max}+1}{(1-\gamma_{\text{ave}})^2}\frac{w_r}{w_e}\right)\right\rceil. \quad \text{(A18)}$$

Further, differentiating the expression under the "ceiling" function on the right-hand side of (A18) with respect to γ_{ave}, we get

$$\frac{1}{(1-\gamma_{\text{ave}})^2}\left(T_{\max} - 2\,(T_{\max}+1)\,\frac{w_r}{w_e(1-\gamma_{\text{ave}})}\right), \quad \text{(A19)}$$

which is non-negative (non-positive) if and only if $\gamma_{\text{ave}} \leq \gamma^{\text{u}}_{\text{ave}}$. $\square$

Proof of Proposition 2

Using

$$L(y,\gamma(y)) = \sum_{k=0}^{y} q(k)p(k;y,\gamma(y)), \quad \text{(A20)}$$

we have

$$W(y+1) - W(y) = w_r + w_e \Delta L(y,\gamma(y)), \quad \text{(A21)}$$

where

$$\Delta L(y,\gamma(y)) = L(y+1,\gamma(y+1)) - L(y,\gamma(y)). \quad \text{(A22)}$$

Now,

$$\begin{aligned}\Delta L(y,\gamma(y)) &= \sum_{k=0}^{y+1} q(k)p(k;y+1,\gamma(y+1)) - \sum_{k=0}^{y} q(k)p(k;y,\gamma(y+1)) \\ &+ \sum_{k=0}^{y} q(k)p(k;y,\gamma(y+1)) - \sum_{k=0}^{y} q(k)p(k;y,\gamma(y)). \end{aligned} \quad \text{(A23)}$$

Using (A1) and (A5), we have

$$\begin{aligned}&\sum_{k=0}^{y+1} q(k)p(k;y+1,\gamma(y+1)) - \sum_{k=0}^{y} q(k)p(k;y,\gamma(y+1)) \\ &= -(1-\gamma(y+1))\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(y+1)) \leq 0. \end{aligned} \quad \text{(A24)}$$

Next,

$$\sum_{k=0}^{y} q(k)(p(k;y,\gamma(y+1)) - p(k;y,\gamma(y))) = \sum_{k=0}^{y} q(k) \int_{y}^{y+1} \frac{\partial p(k;y,\gamma(s))}{\partial s} ds$$
$$= \sum_{k=0}^{y} q(k) \int_{y}^{y+1} \frac{\partial p(k;y,\gamma(s))}{\partial \gamma(s)} \frac{d\gamma(s)}{ds} ds.$$
$$= \int_{y}^{y+1} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^{y} q(k) \frac{\partial p(k;y,\gamma(s))}{\partial \gamma(s)}. \tag{A25}$$

Note that

$$\sum_{k=0}^{y} q(k) \frac{\partial p(k;y,\gamma(s))}{\partial \gamma(s)}$$
$$= \sum_{k=0}^{y} q(k) \left(\left(\frac{y!}{k!(y-k)!} \right) (y-k)\gamma(s)^{y-k-1}(1-\gamma(s))^k - k\gamma(s)^{y-k}(1-\gamma(s))^{k-1} \right)$$
$$= \sum_{k=0}^{y-1} q(k) \left(\frac{y!}{k!(y-k-1)!} \right) \gamma(s)^{y-k-1}(1-\gamma(s))^k - \sum_{k=1}^{y} q(k) \left(\frac{y!}{(k-1)!(y-k)!} \right) \gamma(s)^{y-k}(1-\gamma(s))^{k-1}$$
$$= y \sum_{k=0}^{y-1} q(k) p(k;y-1,\gamma(s)) - y \sum_{k=1}^{y} q(k) p(k-1;y-1,\gamma(s))$$
$$= y \sum_{k=0}^{y-1} (1-F_T(k))\, p(k;y-1,\gamma(s)), \tag{A26}$$

so that (A25) becomes

$$\sum_{k=0}^{y} q(k)(p(k;y,\gamma(y+1)) - p(k;y,\gamma(y))) = y \int_{y}^{y+1} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^{y-1} (1-F_T(k))\, p(k;y-1,\gamma(s)) \leq 0, \tag{A27}$$

since $\frac{d\gamma(s)}{ds} < 0$. Thus, combining (A24) and (A27), we get $\Delta L(y,\gamma(y)) \leq 0$. Further, consider

$$\Delta^2 L(y,\gamma(y)) = \Delta L(y+1,\gamma(y+1)) - \Delta L(y,\gamma(y))$$
$$= -(1-\gamma(y+2)) \sum_{k=0}^{y+1} (1-F_T(k))\, p(k;y+1,\gamma(y+2)) + (1-\gamma(y+1)) \sum_{k=0}^{y} (1-F_T(k))\, p(k;y,\gamma(y+1))$$
$$+ (y+1) \int_{y+1}^{y+2} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^{y} (1-F_T(k))\, p(k;y,\gamma(s))$$
$$- y \int_{y}^{y+1} ds \frac{d\gamma(s)}{ds} \sum_{k=0}^{y-1} (1-F_T(k))\, p(k;y-1,\gamma(s)). \tag{A28}$$

Focusing on the second line in (A28), we get

$$-(1-\gamma(y+2)) \sum_{k=0}^{y+1} (1-F_T(k))\, p(k;y+1,\gamma(y+2)) + (1-\gamma(y+1)) \sum_{k=0}^{y} (1-F_T(k))\, p(k;y,\gamma(y+1))$$

$$
\begin{aligned}
&= -(1-\gamma(y+2))\sum_{k=0}^{y+1}(1-F_T(k))\,p(k;y+1,\gamma(y+2)) + (1-\gamma(y+2))\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(y+2))\\
&- (1-\gamma(y+2))\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(y+2)) + (1-\gamma(y+1))\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(y+1))\\
&= (1-\gamma(y+2))\left(\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(y+2)) - \sum_{k=0}^{y+1}(1-F_T(k))\,p(k;y+1,\gamma(y+2))\right)\\
&+ \sum_{k=0}^{y}(1-F_T(k))\left((1-\gamma(y+1))p(k;y,\gamma(y+1)) - (1-\gamma(y+2))p(k;y,\gamma(y+2))\right). \qquad \text{(A29)}
\end{aligned}
$$

Note that, using (A7)-(A10), we obtain

$$
\begin{aligned}
&(1-\gamma(y+2))\left(\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(y+2)) - \sum_{k=0}^{y+1}(1-F_T(k))\,p(k;y+1,\gamma(y+2))\right)\\
&= (1-\gamma(y+2))^2\left(\sum_{k=0}^{y}p_T(k+1)p(k;y,\gamma(y+2))\right) \geq 0. \qquad \text{(A30)}
\end{aligned}
$$

Further,

$$
\begin{aligned}
&\sum_{k=0}^{y}(1-F_T(k))\left((1-\gamma(y+1))p(k;y,\gamma(y+1)) - (1-\gamma(y+2))p(k;y,\gamma(y+2))\right)\\
&= -\sum_{k=0}^{y}(1-F_T(k))\int_{y+1}^{y+2}\frac{\partial((1-\gamma(s))p(k;y,\gamma(s)))}{\partial s}ds\\
&= -\sum_{k=0}^{y}(1-F_T(k))\int_{y+1}^{y+2}\frac{\partial((1-\gamma(s))p(k;y,\gamma(s)))}{\partial \gamma(s)}\frac{d\gamma(s)}{ds}ds. \qquad \text{(A31)}
\end{aligned}
$$

Focusing on the third and the fourth lines in (A28), we obtain

$$
\begin{aligned}
&(y+1)\int_{y+1}^{y+2}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(s)) - y\int_{y}^{y+1}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y-1}(1-F_T(k))\,p(k;y-1,\gamma(s))\\
&= (y+1)\int_{y+1}^{y+2}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(s)) - y\int_{y+1}^{y+2}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y-1}(1-F_T(k))\,p(k;y-1,\gamma(s))\\
&+ y\int_{y+1}^{y+2}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y-1}(1-F_T(k))\,p(k;y-1,\gamma(s)) - y\int_{y}^{y+1}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y-1}(1-F_T(k))\,p(k;y-1,\gamma(s))\\
&= (y+1)\int_{y+1}^{y+2}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(s)) - y\int_{y+1}^{y+2}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y-1}(1-F_T(k))\,p(k;y-1,\gamma(s))\\
&+ y\int_{y+1}^{y+2}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y-1}(1-F_T(k))\,p(k;y-1,\gamma(s)) - y\int_{y+1}^{y+2}ds\frac{d\gamma(s-1)}{ds}\sum_{k=0}^{y-1}(1-F_T(k))\,p(k;y-1,\gamma(s-1))\\
&= (y+1)\int_{y+1}^{y+2}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(s)) - y\int_{y+1}^{y+2}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y-1}(1-F_T(k))\,p(k;y-1,\gamma(s))
\end{aligned}
$$

$$
\begin{aligned}
&-\beta y\int_{y+1}^{y+2}ds\left(\sum_{k=0}^{y-1}(1-F_T(k))\left(\gamma(s)(1-\gamma(s))p(k;y-1,\gamma(s))-\gamma(s-1)(1-\gamma(s-1))p(k;y-1,\gamma(s-1))\right)\right)\\
&=\int_{y+1}^{y+2}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y}(1-F_T(k))\left((y+1)p(k;y,\gamma(s))-yp(k;y-1,\gamma(s))\right)\\
&-\beta y\int_{y+1}^{y+2}ds\int_{s-1}^{s}d\xi\frac{d\gamma(\xi)}{d\xi}\sum_{k=0}^{y-1}(1-F_T(k))\frac{\partial(\gamma(\xi)(1-\gamma(\xi))p(k;y-1,\gamma(\xi)))}{\partial\gamma(\xi)},
\end{aligned}
\tag{A32}
$$

where we have used $\frac{d\gamma(s)}{ds}=-\beta\gamma(s)(1-\gamma(s))$. Combining (A31) and (A32), we get

$$
\begin{aligned}
&\int_{y+1}^{y+2}ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y}(1-F_T(k))\left(-\frac{\partial(1-\gamma(s))p(k;y,\gamma(s))}{\partial\gamma(s)}+(y+1)p(k;y,\gamma(s))-yp(k;y-1,\gamma(s))\right)\\
&-\beta y\int_{y+1}^{y+2}ds\int_{s-1}^{s}d\xi\frac{d\gamma(\xi)}{d\xi}\sum_{k=0}^{y-1}(1-F_T(k))\frac{\partial(\gamma(\xi)(1-\gamma(\xi))p(k;y-1,\gamma(\xi)))}{\partial\gamma(\xi)}\\
&=\int_{y+1}^{y+2}\frac{ds}{\gamma(s)}\frac{d\gamma(s)}{ds}\sum_{k=0}^{y}(1-F_T(k))\,2p(k;y,\gamma(s))\left(y(\gamma(s)-1)+k+\gamma(s)\right)\\
&-\beta y\int_{y+1}^{y+2}ds\int_{s-1}^{s}d\xi\frac{d\gamma(\xi)}{d\xi}\sum_{k=0}^{y-1}(1-F_T(k))\frac{\partial(\gamma(\xi)(1-\gamma(\xi))p(k;y-1,\gamma(\xi)))}{\partial\gamma(\xi)},\\
&=-\beta\int_{y+1}^{y+2}ds(1-\gamma(s))\sum_{k=0}^{y}(1-F_T(k))\,2p(k;y,\gamma(s))\left(y(\gamma(s)-1)+k+\gamma(s)\right)\\
&-\beta y\int_{y+1}^{y+2}ds\int_{s-1}^{s}d\xi\frac{d\gamma(\xi)}{d\xi}\sum_{k=0}^{y-1}(1-F_T(k))\,p(k;y-1,\gamma(\xi))\left(y(1-\gamma(\xi))-k-\gamma(\xi)\right),\\
&=2\beta\int_{y+1}^{y+2}ds(1-\gamma(s))\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(s))\left(y(1-\gamma(s))-k-\gamma(s)\right)\\
&+\beta^2y\int_{y+1}^{y+2}ds\int_{s-1}^{s}d\xi\gamma(\xi)(1-\gamma(\xi))\sum_{k=0}^{y-1}(1-F_T(k))\,p(k;y-1,\gamma(\xi))\left(y(1-\gamma(\xi))-k-\gamma(\xi)\right).
\end{aligned}
\tag{A33}
$$

Sufficient condition for (A33) to be non-negative is that, for given y,

$$
\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma)\left(y(1-\gamma)-k-\gamma\right)\geq 0,
\tag{A34}
$$

for any $\gamma\in[\gamma(y+2),\gamma(y+1)]$. Note that

$$
\begin{aligned}
&\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma)\left(y(1-\gamma)-k-\gamma\right)\\
&=\sum_{k=0}^{y-2}(1-F_T(k))\,p(k;y,\gamma)\left(y(1-\gamma)-k-\gamma\right)\\
&+(1-F_T(y-1))\,\gamma y(1-\gamma)^{y-1}\left(1-(y+1)\gamma\right)+(1-F_T(y))\,(1-\gamma)^y(-(y+1)\gamma),
\end{aligned}
\tag{A35}
$$

so that, for $(y+1)\gamma \le 2$, $y(1-\gamma)-k-\gamma \ge 0$ for all $k=0,...,y-2$, and

$$\begin{aligned}
&\sum_{k=0}^{y-2}(1-F_T(k))\,p(k;y,\gamma)\,(y(1-\gamma)-k-\gamma) \\
&\ge (1-F_T(y-2))\sum_{k=0}^{y-2}p(k;y,\gamma)\,(y(1-\gamma)-k-\gamma) \\
&= (1-F_T(y-2))\left(\sum_{k=0}^{y}p(k;y,\gamma)\,(y(1-\gamma)-k-\gamma)-\gamma y(1-\gamma)^{y-1}\,(1-(y+1)\gamma)-(1-\gamma)^y(-(y+1)\gamma)\right) \\
&= (1-F_T(y-2))\left(-\gamma-\gamma y(1-\gamma)^{y-1}\,(1-(y+1)\gamma)-(1-\gamma)^y(-(y+1)\gamma)\right). \qquad \text{(A36)}
\end{aligned}$$

Thus, the expression in (A35) is non-negative if

$$\begin{aligned}
&(1-F_T(y-2))\left(-\gamma-\gamma y(1-\gamma)^{y-1}\,(1-(y+1)\gamma)-(1-\gamma)^y(-(y+1)\gamma)\right) \\
&\ge (1-F_T(y))\,\gamma(1-\gamma)^y(y+1)+(1-F_T(y-1))\,\gamma y(1-\gamma)^{y-1}\,((y+1)\gamma-1), \qquad \text{(A37)}
\end{aligned}$$

or, equivalently,

$$(F_T(y)-F_T(y-2))\,(1-\gamma)^y(y+1)+(F_T(y-1)-F_T(y-2))\,y(1-\gamma)^{y-1}((y+1)\gamma-1)\ge 1-F_T(y-2). \qquad \text{(A38)}$$

The left-hand side of (A38) can be re-arranged as

$$\begin{aligned}
&(F_T(y)-F_T(y-2))\,(1-\gamma)^y(y+1)+(F_T(y-1)-F_T(y-2))\,y(1-\gamma)^{y-1}((y+1)\gamma-1) \\
&= (p_T(y)+p_T(y-1))\,(1-\gamma)^y(y+1)+p_T(y-1)y(1-\gamma)^{y-1}((y+1)\gamma-1) \\
&= p_T(y)(1-\gamma)^y(y+1)+p_T(y-1)(1-\gamma)^y\left(1+\frac{y^2\gamma}{1-\gamma}\right) \\
&\ge (1-\gamma)^y\,(p_T(y)(y+1)+p_T(y-1)). \qquad \text{(A39)}
\end{aligned}$$

Thus, (A38) is satisfied as long as

$$(1-\gamma)^y\,(p_T(y)(y+1)+p_T(y-1))\ge 1-F_T(y-2), \qquad \text{(A40)}$$

which can be re-arranged as

$$\gamma \le 1-\left(\frac{1-F_T(y-2)}{(y+1)p_T(y)+p_T(y-1)}\right)^{1/y}. \qquad \text{(A41)}$$

Combining (A41) with $(y+1)\gamma \leq 2$ and $\gamma \in [\gamma(y+2), \gamma(y+1)]$, and noting that for $\gamma = 0$ the monotonicity of $\Delta L(y, \gamma(y))$ is assured, we obtain the final sufficient condition

$$\gamma(y+1) \leq \min\left(\frac{2}{y+1}, 1 - \min\left(1, \left(\frac{1 - F_T(y-2)}{(y+1)p_T(y) + p_T(y-1)}\right)^{1/y}\right)\right) = \min\left(\frac{2}{y+1}, \gamma_T(y+1)\right). \tag{A42}$$

Given that $\Delta L(y, \gamma(y))$ is a monotone function of y if (A42) is satisfied, we establish the monotonicity of the optimal staffing level y^* with respect to changes in $\frac{w_r}{w_e}$ and $F_T(k)$ following the same arguments used in the proof of Proposition 1. □

Proof of Lemma 1

Note that for $\alpha, \beta \geq 0$, the function

$$x\gamma(x) = \frac{x}{1 + e^{\alpha + \beta x}} \tag{A43}$$

defined on continuous set $x \geq 0$ has a unique global maximum x^* which satisfies the first-order optimality condition

$$e^{-\alpha} = e^{\beta x^*}\left(\beta x^* - 1\right). \tag{A44}$$

Thus, the maximum value for (A43) can be expressed as

$$x^*\gamma(x^*) = \frac{1}{\beta}\frac{\beta x^*}{1 + e^{\alpha + \beta x^*}} = \frac{\beta x^* - 1}{\beta}. \tag{A45}$$

Now, (A45) does not exceed 2 if and only if $\beta x^* \leq 2\beta + 1$, which, from (A44) is equivalent to

$$e^{-\alpha} \leq 2\beta e^{2\beta+1} \Leftrightarrow 2\beta e^{\alpha + 2\beta + 1} \geq 1. \tag{A46}$$

Thus, (A46) ensures that the maximum of $y\gamma(y)$ cannot exceed 2 for all integer values of y as well. □

Proof of Proposition 3

Under (27), the optimal staffing level y^* satisfies (26), which can be re-expressed as

$$y^* = \min\left(y \in \mathcal{N}^+ | w_r + w_e\left(L(y+1, \gamma(y+1)) - L(y, \gamma(y))\right) \geq 0\right), \tag{A47}$$

where, using (A23),(A24) and (A27), we have

$$L(y+1,\gamma(y+1)) - L(y,\gamma(y)) = -(1-\gamma(y+1))\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(y+1)) + y\int_{y}^{y+1} ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y-1}(1-F_T(k))\,p(k;y-1,\gamma(s)). \quad \text{(A48)}$$

Now, consider an element of E, y^{ML}, which satisfies

$$y^{\mathrm{ML}} = \min\left(y\in\mathcal{N}^+ | w_r + w_e\left(-(1-\gamma(y^{\mathrm{ML}}+1))\right)\sum_{k=0}^{y}(1-F_T(k))\,p(k;y,\gamma(y^{\mathrm{ML}}+1)) \geq 0\right), \quad \text{(A49)}$$

so that

$$w_r + w_e\left(-(1-\gamma(y^{\mathrm{ML}}+1))\right)\sum_{k=0}^{y^{\mathrm{ML}}}(1-F_T(k))\,p(k;y^{\mathrm{ML}},\gamma(y^{\mathrm{ML}}+1)) \geq 0, \quad \text{(A50)}$$

Suppose that $y^{\mathrm{ML}} > y^* \Leftrightarrow y^{\mathrm{ML}} \geq y^*+1$. This, in turn, implies that

$$w_r + w_e\left(-(1-\gamma(y^*+1))\right)\sum_{k=0}^{y^*}(1-F_T(k))\,p(k;y^*,\gamma(y^*+1)) < 0, \quad \text{(A51)}$$

and also, since $\frac{d\gamma(y)}{dy} \leq 0$, that

$$w_r + w_e\left(-(1-\gamma(y^*+1))\right)\sum_{k=0}^{y^*}(1-F_T(k))\,p(k;y^*,\gamma(y^*+1)) + w_e\left(y^*\int_{y^*}^{y^*+1} ds\frac{d\gamma(s)}{ds}\sum_{k=0}^{y^*-1}(1-F_T(k))\,p(k;y^*-1,\gamma(s))\right) < 0, \quad \text{(A52)}$$

which contradicts (A47). Thus, $y^{\mathrm{ML}} \leq y^*$. $\square$