Parametric Forecasting and Stochastic Programming Models for Call-Center Workforce Scheduling

Noah Gans
OPIM Department
The Wharton School
U. Pennsylvania

Haipeng Shen
Department of Statistics and OR
UNC Chapel Hill
IIM, U. Hong Kong

Yong-Pin Zhou
ISOM Department
The Foster School
U. Washington

Nikolay Korolev Alan McCord Herbert Ristock
Genesys Telecommunications Laboratories, Inc.

April 17, 2015

Online Supplement

A Remarks

Remark 1 The dimensionality of the vector time series is typically high, for example, $I = 26$ half-hour periods in a 13-hour working day. Our multiplicative model easily captures the two-way (intraday and interday) time dependence that is common to call centers and other large-scale service systems and reduces potentially high-dimensional forecasts to a single dimension, the daily rate, which can be discretized using Gaussian quadrature. Our model is a parametric analogue of the data-driven approach of Shen and Huang (2008) and is an extension of the approach of Whitt (1999), which proposes a similar approach for un-transformed arrival rates and assumes that all days share a common intraday arrival-rate profile.

In terms of forecast accuracy, our model's performance is roughly comparable to that of the models proposed in Weinberg et al. (2007), Shen and Huang (2008), and Aldor-Noiman et al. (2009). While the arrival count data used in our numerical tests do not exhibit monthly or annual seasonal effects, these could be easily incorporated into (1) as well. There are also more complex models of intraday performance, such as those found in Ibrahim and L'Ecuyer (2013) and Oreshkin et al. (2014), that could further reduce within-day forecast variance, but at the cost of higher forecast dimensionality. While the numerical results reported in §2.5 and §4 suggest that our one-dimensional model is sufficiently accurate to provide desired levels of operational performance, future work could consider methods of dimension reduction of these more complex within-day forecasting approaches. □

Remark 2 Our use of the above results implicitly makes two common assumptions. The first is that, in each of the scenarios, the arrival rate is constant over interval i. While common, this assumption is not necessarily innocuous. Nevertheless, effective measures can be taken to account for time-inhomogeneity within intervals.

For a characterization of time-inhomogeneity, see Brown et al. (2005), and for effective responses see Feldman et al. (2007) and Green et al. (2007), as well as the recent article by Kim and Whitt (2014). The second is that, even if the arrival rate were constant, the use of stationary performance measures assumes that the event rate during interval i under scenario k is large enough that transient effects, due to initial conditions at the start of the interval, are not significant. Because our systems include abandonment, relaxation times are similar to those for infinite-server Markovian systems and occur on the order of an expected service time or time to abandonment, whichever is longer (Coolen-Schrijner and Van Doorn 2001). □

Remark 3 While for a specific arrival-rate realization the QoS constraint may be violated, given a correctly forecasted arrival-rate distribution and many *i.i.d.* days, the long-run average abandonment rate should be less than or equal to α^* (Wang and Ahmed 2008). Of course, in our AR(1) setting, arrival-rate distributions need not be *i.i.d.*, and forecast distributions need not be correct. Nevertheless we conjecture that, in our case, the long-run average abandonment rate will still fall at or below α^* and are working to formally prove it is so (Gans et al. 2014). □

Remark 4 Robbins and Harrison (2010) use a variant of (7) in which the QoS constraint becomes $\sum_{i\in\mathcal{I}} \alpha_i \leq \alpha^*\bar{\lambda} + \delta$, and the objective function is augmented to include a penalty for abandonments above the nominal target of α^*: $\min \sum_{j\in\mathcal{J}} c_j x_j + p\delta$. When abandonments carry explicit, rather than implicit costs, for example in certain contractual arrangements used for outsourcing, this alternate form is appropriate. □

Remark 5 The Kaplan-Meier estimator for exponentially-distributed patience divides the total number of abandonments by the sum of the delays of all calls, including those that are served and those that abandon the queue before being served. See Zohar et al. (2002). Our dataset includes records of average delay in queue only for served calls, however, and we include the total delay of served calls in the denominator of our calculation. If we were to include the waiting time of abandoning calls, it would therefore lower the estimate of θ.

Remark 6 Our definition of α' has the virtue of being straightforward to calculate and analyze. Nevertheless, there are other definitions of α' that we can consider. For example, we can use the optimal solution to (7) to define the expected late-period abandonment rate given only the initial forecast (1): $\alpha' = (\sum_{i\in\mathcal{I}_l} \alpha_i)/(\sum_{i\in\mathcal{I}_l} \lambda_i)$. More generally, one may look for mappings $\{\mathbf{y}_{D+1}, \ldots, \mathbf{y}_{D+h_e}\} \mapsto \alpha'$ that satisfy other objectives, such as the stabilization of late-interval abandonment. In §3.2 we present one such scheme. □

Remark 7 There are several extensions to this scheme that we could consider. Rather than setting a fixed update interval, i^*, Mehrotra et al. (2010) use a sequential procedure that looks for the first period, i^*, for which they can reject the null hypothesis that the arrival-rate pattern comes from the initial forecast distribution. An

alternative that follows along the lines of (17) would be to allow for forecast updates and recourse actions after every interval of the scheduling horizon, $i \in \{1, \ldots, I-1\}$. The computation time required to solve such an *a priori* program would grow exponentially with the number of updates, however. A computationally more practical approach would use (13) for repeated *ex post* updates, since computing time would grow only linearly with the number of updates. Even less complex, and perhaps of greatest practical interest, would be an empirical search across days for an optimal static update interval, i^*. Such a test would be straightforward, though time consuming, and we do not pursue in this paper. □

Remark 8 An alternative scheme uses the *ex post* forecast update after i^*, along with the optimal x from RP (17), within UP (13) to recommend the set of recourse actions that are ultimately taken. As in §3.2.4, we identify an appropriate scenario k by rounding up or down. Rather than directly using the associated z_{jgk}s from (17), however, we substitute $\sum_{i \in \mathcal{I}_l} \alpha_{ik}$, the sum of the optimal α_{ik}s from recourse scenario k, for the right-hand side of (13)'s QoS constraint, $\alpha' \bar{\lambda}'$, and solve (13) to find a low-cost set of recourse actions whose abandonment performance nearly matches that of the original set of z_{jgk}s. □

B Discretization of the Arrival-Rate Forecast Using Gaussian Quadrature

Let ω be a normal random variable with mean μ and variance ϕ, and let ω^* be a discrete random variable such that $\omega^* = \omega_k$ with probability p_k, $1 \leq k \leq K$. In this section, we detail the Gaussian quadrature procedure that we use to compute ω_k and p_k, $1 \leq k \leq K$ so that ω^* matches ω for the first $2K-1$ moments.

Without loss of generality, we first consider a standard normal random variable $Z \sim \mathrm{N}(0,1)$. Denote its first $2K-1$ moments by $\mu_k \equiv \mathrm{E}(Z^k), k = 1, \ldots, 2K-1$. Let Z^* be a discrete random variable such that $Z^* = z_k$ with probability p_k, $1 \leq k \leq K$.

Following the work of Miller and Rice (1983), we proceed to compute $\{(z_k, p_k) \mid 1 \leq k \leq K\}$, for Z^* so that its first $2K-1$ moments match $\mu_k, k = 1, \ldots, 2K-1$:

$$\begin{aligned} p_1 + p_2 + \cdots + p_K &= 1, \\ p_1 z_1 + p_2 z_2 + \cdots + p_K z_K &= \mu_1, \\ p_1 z_1^2 + p_2 z_2^2 + \cdots + p_K z_K^2 &= \mu_2, \\ &\vdots \\ p_1 z_1^{2K-1} + p_2 z_2^{2K-1} + \cdots + p_K z_K^{2K-1} &= \mu_{2K-1}. \end{aligned} \tag{1}$$

The above equations can be solved using a standard method. The derivation below slightly differs from the description of Miller and Rice (1983). A polynomial is first defined as

$$q(z) = (z - z_1)(z - z_2) \ldots (z - z_K) \equiv \sum_{k=0}^{K} q_k z^k. \tag{2}$$

It then follows that $q_K = 1$ and $q(z_k) = 0$ for $1 \leq k \leq K$.

Then, consider the first $K+1$ equations in (1), multiply the first one by q_0, the second one by q_1, etc., and sum them up to obtain:

$$\sum_{k=0}^{K} q_k \mu_k = 0.$$

Similarly, take the second through $(K+2)$th equations in (1), and repeat the above multiplication and summation procedure to obtain:

$$\sum_{k=0}^{K} q_k \mu_{k+1} = 0.$$

The above process can be repeated K times to yield the following set of equations:

$$\begin{aligned}
q_0 + \mu_1 q_1 + \mu_2 q_2 \cdots + \mu_{K-1} q_{K-1} &= -\mu_K, \\
\mu_1 q_0 + \mu_2 q_1 + \mu_3 q_2 \cdots + \mu_K q_{K-1} &= -\mu_{K+1}, \\
\mu_2 q_0 + \mu_3 q_1 + \mu_4 q_2 \cdots + \mu_{K+1} q_{K-1} &= -\mu_{K+2}, \\
&\vdots \\
\mu_{K-1} q_0 + \mu_K q_1 + \mu_{K+1} q_2 \cdots + \mu_{2K-2} q_{K-1} &= -\mu_{2K-1}.
\end{aligned} \tag{3}$$

The coefficients of polynomial (2), q_k, can be obtained by solving the equations in (3), and the roots of the polynomial determine the discrete atoms $\{z_k \mid 1 \leq k \leq K\}$. The corresponding probabilities, $\{p_k \mid 1 \leq k \leq K\}$, can then be found from the original equations in (1) with the z_k substituted. The whole procedure is implemented in the R function *gauss.quad.prob*, available within the R package *statmod*. The computation of the Gaussian quadrature approximation is very fast.

Since normal distributions form a location-scale family, our variable of interest ω is related with Z through $\omega = \zeta + \phi Z$. Once the z_k and p_k for the standard normal distribution are obtained, we can easily obtain the discrete approximation ω^* for ω as follows,

$$\omega^* = \omega_k \text{ with probability } r_k, \quad \text{where } \omega_k = \zeta + \phi z_k \text{ and } r_k = p_k, \ \forall 1 \leq k \leq K. \tag{4}$$

It can be easily shown that ω and ω^* share the same first $2K-1$ moments.

As a side note, for the degenerate one-scenario case (i.e. $K = 1$), special care is needed to make sure that Λ_i has the correct mean. In this case, we set $\omega_1 = \sqrt{\zeta^2 + \phi^2}$ instead of the default value ζ. The reason is that Gaussian quadrature only matches the first moment of ω when $K = 1$, which won't guarantee the mean matching for Λ_i since $\Lambda_i = (\omega\theta_i)^2$.

C Abandonment Calculations for the Erlang-A Model

The primitives for the Erlang A model, also called the M/M/n+M model, are Markovian interarrival times with mean $1/\lambda$, Markovian service times with mean $1/\mu$, n servers, and Markovian patience times with mean $1/\theta$. Given these data, we use the exact expressions presented in Mandelbaum and Zeltyn (2007) to calculate the stationary probability that an arriving customer abandons queue before being served, $f(\lambda, \mu, \theta, n)$.

Recall that the incomplete gamma function is defined as $\gamma(x, y) \triangleq \int_0^y t^x e^{-t} dt$ and, for $x > 0, y \geq 0$ and can be calculated recursively using standard procedures. We have used the algorithm described in Bhattacharjee (1970) and is available in MATLAB. In turn, we use system primitives and the incomplete gamma function to define two useful quantities,

$$J \triangleq \left(\frac{e^{\lambda/\theta}}{\theta}\right) \cdot \left(\frac{\theta}{\lambda}\right)^{\frac{n\mu}{\theta}} \cdot \gamma\left(\frac{n\mu}{\theta}, \frac{\lambda}{\theta}\right) \qquad \text{and} \qquad \mathcal{E} \triangleq \frac{\sum_{j=0}^{n-1}(1/j!) \cdot (\lambda/\mu)^j}{1/(n-1)! \cdot (\lambda/\mu)^{n-1}}.$$

The two quantities then allow us to calculate the desired probability, $f(\lambda, \mu, \theta, n) = \frac{1+(\lambda-n\mu)J}{\mathcal{E}+\lambda J}$.

D Arrival Count Data Used in Numerical Tests

Figure 1 displays summary plots of the arrival data used in the European bank tests reported in §2.5 and §4.1. The left plot shows daily counts for each of 176 days, and the right plot shows average within-day count profiles by day of the week.

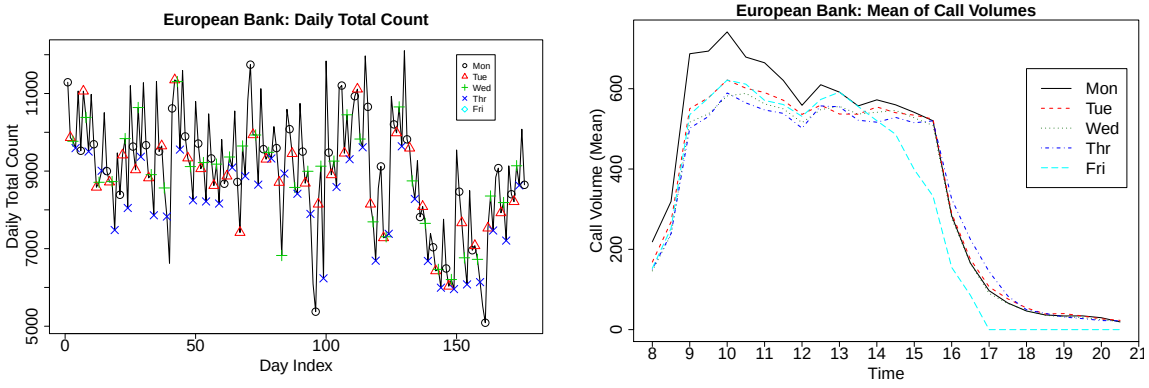

Figure 1: Daily Arrival Counts (left) and Within-Day Average Count Profiles (right) for European Bank

Figure 2 displays summary plots of the arrival data used in the North American bank tests reported in §4.2. The left plot shows daily counts for each of 210 days, and the right plot shows average within-day count profiles by day of the week.

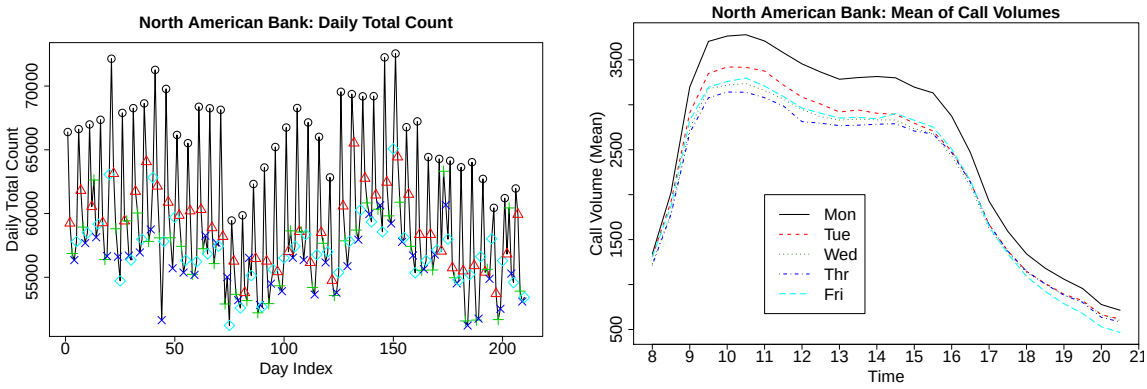

Figure 2: Daily Arrival Counts (left) and Within-Day Average Count Profiles (right) for NA Bank

E Proofs of the Propositions

E.1 Proof of Proposition 3

For notational simplicity, we first drop the subscript d and obtain some generic results.

Suppose we observe y_i, $i = 1, ..., \hat{\imath}$, and denote the corresponding observed vector as $\mathbf{y}_{\hat{\imath}} = (y_1, \ldots, y_{\hat{\imath}})^{\mathsf{T}}$. According to the forecasting model (1) , we then have

$$y_i \text{ is a sample of } Y_i = \sqrt{\Lambda_i} + \epsilon_i = \omega\vartheta_i + \epsilon_i, \quad i = 1, ..., \hat{\imath}, \tag{5}$$

where $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$ and $\omega \sim \mathrm{N}(\zeta, \psi^2)$.

Now consider the posterior density $f(\omega|\mathbf{y}_{\hat{\imath}})$ as follows

$$\begin{aligned} f(\omega|\mathbf{y}_{\hat{\imath}}) \;=\; \frac{f(\omega, \mathbf{y}_{\hat{\imath}})}{f(\mathbf{y}_{\hat{\imath}})} = \frac{\prod_{i=1}^{\hat{\imath}} f(y_i|\omega) \cdot f(\omega)}{f(\mathbf{y}_{\hat{\imath}})} \;&=\; c \cdot e^{-\sum_{i=1}^{\hat{\imath}} \frac{\vartheta_i^2\omega^2 - 2y_i\vartheta_i\omega}{2\sigma^2} - \frac{\omega^2 - 2\zeta\omega}{2\psi^2}} \\ &=\; c \cdot e^{-\frac{\sum_{i=1}^{\hat{\imath}} \vartheta_i^2\omega^2 - 2\sum_{i=1}^{\hat{\imath}} y_i\vartheta_i\omega}{2\sigma^2} - \frac{\omega^2 - 2\zeta\omega}{2\psi^2}}, \\ &=\; c \cdot e^{-\frac{\psi^2\sum_{i=1}^{\hat{\imath}}\vartheta_i^2 + \sigma^2}{2\sigma^2\psi^2}\omega^2 + \frac{\psi^2\sum_{i=1}^{\hat{\imath}} y_i\vartheta_i + \sigma^2\zeta}{\sigma^2\psi^2}\omega}, \end{aligned}$$

where c is a leading constant. Recalling that $a_{\hat{\imath}} = \sum_{i=1}^{\hat{\imath}} \vartheta_i \; y_i$ and $\nu_{\hat{\imath}} = \sum_{i=1}^{\hat{\imath}} \vartheta_i^2$, we can see from the above density function that the posterior of ω, given $\mathbf{y}_{\hat{\imath}}$, is the normal distribution with mean and variance

$$\frac{\psi^2 a_{\hat{\imath}} \;+\; \sigma^2\zeta}{\psi^2\nu_{\hat{\imath}} \;+\; \sigma^2}, \qquad \text{and} \qquad \frac{\sigma^2\psi^2}{\psi^2\nu_{\hat{\imath}} \;+\; \sigma^2}.$$

Now suppose that the prior distribution for day d is $\omega_{d-1,d} \sim \mathrm{N}(\zeta_{d-1,d}, \psi^2_{d-1,d})$, and we observe counts $\{y_{di} \mid i = 1, \ldots, \hat{\imath}\}$ on day d. Let $\nu_{d,\hat{\imath}} = \sum_{i=1}^{\hat{\imath}} \vartheta^2_{l_d,i}$ and $a_{d,\hat{\imath}} \equiv \sum_{i=1}^{\hat{\imath}} \vartheta_{l_d,i} \; y_{d,i}$. It then follows from the above generic results that the posterior distribution $\omega_{d_{\hat{\imath}},d}$ is normal with mean and variance

$$\zeta_{d_{\hat{\imath}},d} = \frac{\psi^2_{d-1,d} a_{d,\hat{\imath}} \;+\; \sigma^2\zeta_{d-1,d}}{\psi^2_{d-1,d}\nu_{d,\hat{\imath}} \;+\; \sigma^2}, \qquad \text{and} \qquad \psi^2_{d_{\hat{\imath}},d} \;=\; \frac{\sigma^2\psi^2_{d-1,d}}{\psi^2_{d-1,d}\nu_{d,\hat{\imath}} \;+\; \sigma^2}.$$

E.2 Lemma 1 and Its Proof

To prove Proposition (5), we first state and prove the following Lemma 1.

Lemma 1 *Suppose* $Y_i = \omega\vartheta_i + \epsilon_i, \quad i = 1, \ldots, \hat{\imath}$, *where* $\omega \sim N(\zeta, \psi^2)$, $\epsilon_i \sim N(0, \sigma^2)$, *and* ω *and* ϵ_i *are independent. Define* $A = \sum_{i=1}^{\hat{\imath}} \vartheta_i Y_i$. *Then* A *is normally distributed with mean* $\nu_{\hat{\imath}}\zeta$ *and variance* $\nu_{\hat{\imath}}^2\psi^2 + \nu_{\hat{\imath}}\sigma^2$, *where* $\nu_{\hat{\imath}} = \sum_{i=1}^{\hat{\imath}} \vartheta_i^2$.

Proof We note that $Y_i|\omega \sim \mathrm{N}(\omega\vartheta_i, \sigma^2)$, and the Y_i's are independent conditional on ω. It then follows that

$$A|\omega \sim \mathrm{N}(\omega\nu_{\hat{\imath}}, \sigma^2\nu_{\hat{\imath}}).$$

Given that $\omega \sim \mathrm{N}(\zeta, \psi^2)$, the density function of A at a can be written as

$$f(a) = \int f(a|\omega)f(\omega)d\omega = \int \frac{1}{\sqrt{2\pi\sigma^2\nu_{\hat{\imath}}}} e^{-\frac{(a-\omega\nu_{\hat{\imath}})^2}{2\sigma^2\nu_{\hat{\imath}}}} \cdot \frac{1}{\sqrt{2\pi\psi^2}} e^{-\frac{(\omega-\zeta)^2}{2\psi^2}} d\omega,$$

which can be reorganized as

$$f(a) = \frac{1}{\sqrt{(2\pi)^2\nu_{\hat{\imath}}\sigma^2\psi^2}} e^{-\frac{a^2}{2\sigma^2\nu_{\hat{\imath}}} - \frac{\zeta^2}{2\psi^2}} \int e^{-\frac{\nu_{\hat{\imath}}^2\omega^2 - 2\nu_{\hat{\imath}}a\omega}{2\sigma^2\nu_{\hat{\imath}}} - \frac{\omega^2 - 2\zeta\omega}{2\psi^2}} d\omega.$$

Note that the exponent of the integrand can be re-expressed as

$$-\frac{\nu_{\hat{\imath}}\psi^2 + \sigma^2}{2\sigma^2\psi^2}(\omega - \frac{a\psi^2 + \zeta\sigma^2}{\nu_{\hat{\imath}}\psi^2 + \sigma^2})^2 + \frac{(a\psi^2 + \zeta\sigma^2)^2}{2\sigma^2\psi^2(\nu_{\hat{\imath}}\psi^2 + \sigma^2)}.$$

Then, the density function $f(a)$ equals to

$$f(a) = \frac{1}{\sqrt{2\pi\nu_{\hat{\imath}}(\nu_{\hat{\imath}}\psi^2 + \sigma^2)}} e^{-\frac{(a-\nu_{\hat{\imath}}\zeta)^2}{2\nu_{\hat{\imath}}(\nu_{\hat{\imath}}\psi^2 + \sigma^2)}},$$

which implies that $A \sim \mathrm{N}(\nu_{\hat{\imath}}\zeta, \nu_{\hat{\imath}}^2\psi^2 + \nu_{\hat{\imath}}\sigma^2)$. □

E.3 Proof of Proposition (5)

Consider the following model:

$$Y_{D+d,i} = \omega_{D+d}\nu_{l_{D+d},i} + \epsilon_{D+d,i}, \qquad i = 1, \ldots, I, \tag{6}$$

where $\omega_{D+d} \sim \mathrm{N}(\zeta_{D,D+d}, \psi^2_{D,D+d})$, $\epsilon_{D+d,i} \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$, and ω_{D+d} and $\epsilon_{D+d,i}$ are independent.

Given (14), it follows from Lemma 1 that

$$A_{D+d,\hat{\imath}} \sim \mathrm{N}(\nu_{D+d,\hat{\imath}}\zeta_{D,D+d}, \ \ \nu^2_{D+d,\hat{\imath}}\psi^2_{D,D+d} + \nu_{D+d,\hat{\imath}}\sigma^2). \tag{7}$$

The definition of $\zeta'_{D,D+d}$ in (15)implies that $\zeta'_{D,D+d}$ is a linear shifted and scaled transformation of $A_{D+d,\hat{\imath}}$; hence $\zeta'_{D,D+d}$ is also normally distributed, and its variance is a scaled version of the variance of $A_{D+d,\hat{\imath}}$, whose expression is given in the right equality of (16).

Below we use induction to prove that $E(\zeta'_{D,D+d}) = \zeta_{D,D+d}$. To begin with, consider $d = 1$. Note that

$$\zeta'_{D,D+1} = \frac{\psi^2_{D,D+1} A_{D+1,\hat{\imath}} \ + \ \sigma^2 \zeta_{D,D+1}}{\psi^2_{D,D+1} \nu_{D+1,\hat{\imath}} \ + \ \sigma^2},$$

which implies that

$$E(\zeta'_{D,D+1}) = \frac{\psi^2_{D,D+1} \nu_{D+1,\hat{\imath}} \zeta_{D,D+1} \ + \ \sigma^2 \zeta_{D,D+1}}{\psi^2_{D,D+1} \nu_{D+1,\hat{\imath}} \ + \ \sigma^2} = \zeta_{D,D+1}.$$

Now suppose that, for $d \geq 2$, we have proven that $E(\zeta'_{D,D+d-1}) = \zeta_{D,D+d-1}$, for any $\hat{\imath} \in \{1, \ldots, I\}$. We then have

$$\zeta_{D,D+d} - E(\zeta'_{D,D+d}) = \frac{\sigma^2}{\psi^2_{D+d-1,D+d} \nu_{D+d,\hat{\imath}} \ + \ \sigma^2} \ (\zeta_{D,D+d} - \zeta_{D+d-1,D+d}) \,. \tag{8}$$

Hence, it suffices to show that $\zeta_{D,D+d} - \zeta_{D+d-1,D+d} = 0$. For that, we note

$$\begin{aligned}
\zeta_{D,D+d} - \zeta_{D+d-1,D+d} &= \zeta_{D,D+d} - E(E(\omega_{D,D+d} | \mathbf{Y}_{D+1}, \ldots, \mathbf{Y}_{D+d-1})) \\
&= \zeta_{D,D+d} - \{\alpha_{l_{D+d}} + \beta[E(E(\omega_{D,D+d-1} | \mathbf{Y}_{D+1}, \ldots, \mathbf{Y}_{D+d-1})) - \alpha_{l_{D+d-1}}]\} \\
&= \zeta_{D,D+d} - \alpha_{l_{D+d}} - \beta[E(\zeta'_{D,D+d-1}) - \alpha_{l_{D+d-1}}] \\
&= \zeta_{D,D+d} - \alpha_{l_{D+d}} - \beta(\zeta_{D,D+d-1} - \alpha_{l_{D+d-1}}) \\
&= \beta^d(\omega_D - \alpha_{l_D}) - \beta^d(\omega_D - \alpha_{l_D}) = 0.
\end{aligned}$$

Thus, we have shown that $E(\zeta'_{D,D+d}) = \zeta_{D,D+d}$ for $d \geq 1$, which concludes the induction proof.

F Results for Wilcoxon Signed Rank Tests

The figures below show results for Wilcoxon signed rank tests that check to see if the median of the difference between two matched sets of values is significantly different from zero. The tests were implemented using the *signrank* function in MATLAB R2013a.

Figure 3 reports the results of one-tailed tests that compare daily abandonment rates across the nine schemes. The result shown in each cell of the figure's table is based on a set of 76 values, where each value is the abandonment rate associated with the column label, less the abandonment rate associated with the row label, for a single day. The one-sided, left-tailed p-value shown in the cell is the probability that the median value is less than zero, and we interpret the p-value to be a measure of the probability that the abandonment rates associated with the column label's scheme are significantly less than those associated with the row label's scheme. Below, we denote a particular cell by its (row,column)-heading coordinates.

Results of 1-Sided Wilcoxon Signed Rank Tests Using European Bank Abandonment Rates
Left-Tail Probabilities for [Colum Values - Row Values]

	SP1	UP1	RP1	SP4	UP4	RP4	SP100	UP100	RP100
SP1	1.000	0.750	0.024	1.000	1.000	1.000	1.000	1.000	1.000
UP1	0.251	1.000	0.032	1.000	1.000	1.000	1.000	1.000	1.000
RP1	0.976	0.968	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SP4	0.000	0.000	0.000	1.000	0.992	0.413	0.868	0.845	0.002
UP4	0.000	0.000	0.000	0.008	1.000	0.323	0.042	0.310	0.000
RP4	0.000	0.000	0.000	0.589	0.678	1.000	0.553	0.639	0.006
SP100	0.000	0.000	0.000	0.134	0.959	0.450	1.000	0.966	0.002
UP100	0.000	0.000	0.000	0.156	0.693	0.363	0.035	1.000	0.001
RP100	0.000	0.000	0.000	0.998	1.000	0.994	0.998	0.999	1.000

Figure 3: Results of One-Tailed Wilcoxon Signed Rank Tests for European Bank Abandonment Data

Results of 1-Sided Wilcoxon Signed Rank Tests Using European Bank Cost per Handled Call
Right-Tail Probabilities for [Column Values - Row Values]

	SP1	UP1	RP1	SP4	UP4	RP4	SP100	UP100	RP100
SP1	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	0.999
UP1	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
RP1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SP4	0.000	0.000	0.000	1.000	0.000	0.000	0.985	0.000	0.000
UP4	0.000	0.000	0.000	1.000	1.000	0.006	1.000	0.371	0.000
RP4	0.000	0.000	0.000	1.000	0.994	1.000	1.000	0.993	0.028
SP100	0.000	0.000	0.000	0.015	0.000	0.000	1.000	0.000	0.000
UP100	0.000	0.000	0.000	1.000	0.631	0.007	1.000	1.000	0.000
RP100	0.001	0.000	0.000	1.000	1.000	0.973	1.000	1.000	1.000

Figure 4: Results of One-Tailed Wilcoxon Signed Rank Tests for European Bank Average Cost Data

For example, cell (SP1,UP1) shows that the p-value associated with the probability that the median of UP1-SP1 differences is less than zero is 0.75. In this case, we cannot reject the null hypothesis that the median value of the differences is less than zero, and we conclude that UP1's abandonment rates do not appear to be significantly higher than SP1's.

In Figure 3, clusters of adjacent (blue) cells with solid borders compare different schemes that use the same number of scenarios. For example, from (SP1,UP1) we see that UP1's daily abandonment rates do not appear to be significantly higher than SP1's. In contrast, from the p-values reported in (SP1,RP1) and (UP1,RP1) we see that RP1's daily abandonment rates appear to be significantly higher than those of the other two 1-scenario schemes. In contrast, results for (SP4,UP4), (SP4,RP4), and (UP4,RP4) show that UP4's daily abandonment rates do not appear to be significantly higher than SP4's, nor do RP4's as compared to SP4's and UP4's. While UP100's rates do not appear to be significantly higher than SP100's, RP100's abandonment rates do appear to be systematically higher than those of the other two schemes. This last fact is visible to the eye in the left panel of Figure 4 in the main paper.

The set of three diagonal (green) cells with dashed borders in Figure 3 reports tests of differences between the results of 4 and 100-scenario versions that use the same scheduling approach. From (SP100,SP4), (UP100,UP4), and (RP100,RP4) we see that the 4-scenario version of these schemes have abandonment-rates

that do not appear to be significantly higher than that of their 100-scenario counterparts. This is evidence that only a small number of scenarios is needed to ensure that the schemes properly compensate for arrival-rate variability.

Figure 4 reports analogous results for daily cost per handled call. Here, we report right-tail probabilities and interpret small p values to suggest that average costs associated with the focal column label's scheme are systematically larger than those associated with the focal row label's scheme. In this figure we see that, for any number of scenarios, the UP scheme appears to have lower average costs than its analogous SP scheme, and the RP scheme appears to have lower average costs than either its SP or UP counterpart. When we compare SP4 to SP100 and UP4 to UP100, we see that the schemes with higher numbers of scenarios do not have significantly lower costs. In contrast RP100 has average costs that appear to be significantly lower than those of RP4, but as Figure 3 indicates, the cost savings also come with significantly higher abandonment rates.

In sum, for the European Bank call center data, the use of UP schemes provides significantly lower costs than SP schemes, without appearing to degrade abandonment rate performance. Furthermore, the 4-scenario versions of the SP and UP schemes appear to perform as well as the 100-scenario versions. While the same observations hold for RP4, they do not hold for RP100.

Results of 1-Sided Wilcoxon Signed Rank Tests Using North American Bank Abandonment Rates
Left-Tail Probabilities for [Colum Values - Row Values]

	SP1	UP1	RP1	SP4	UP4	RP4	SP100	UP100	RP100
SP1	1.000	0.091	0.000	1.000	1.000	1.000	1.000	1.000	1.000
UP1	0.909	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
RP1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SP4	0.000	0.000	0.000	1.000	0.207	0.998	0.621	0.231	0.924
UP4	0.000	0.000	0.000	0.794	1.000	0.997	0.914	0.441	0.978
RP4	0.000	0.000	0.000	0.002	0.003	1.000	0.003	0.004	0.214
SP100	0.000	0.000	0.000	0.380	0.086	0.997	1.000	0.110	0.864
UP100	0.000	0.000	0.000	0.770	0.560	0.996	0.891	1.000	0.965
RP100	0.000	0.000	0.000	0.077	0.022	0.787	0.137	0.036	1.000

Figure 5: Results of One-Tailed Wilcoxon Signed Rank Tests for NA Bank Abandonment Data

Results of 1-Sided Wilcoxon Signed Rank Tests Using North American Bank Cost per Handled Call
Right-Tail Probabilities for [Column Values - Row Values]

	SP1	UP1	RP1	SP4	UP4	RP4	SP100	UP100	RP100
SP1	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
UP1	1.000	1.000	0.000	1.000	1.000	0.000	1.000	1.000	0.000
RP1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SP4	0.000	0.000	0.000	1.000	0.000	0.000	0.766	0.000	0.000
UP4	1.000	0.000	0.000	1.000	1.000	0.000	1.000	0.440	0.000
RP4	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000	0.125
SP100	0.000	0.000	0.000	0.235	0.000	0.000	1.000	0.000	0.000
UP100	1.000	0.000	0.000	1.000	0.562	0.000	1.000	1.000	0.000
RP100	1.000	1.000	0.000	1.000	1.000	0.875	1.000	1.000	1.000

Figure 6: Results of One-Tailed Wilcoxon Signed Rank Tests for NA Bank Average Cost Data

Figures 5 and 6 present the results of analogous Wilcoxon tests for the 110 sets of abandonment rates and

costs per handled call generated with the North American data. From the figures' tables we see that the RP schemes are better behaved in these tests. As we would expect, the RP schemes appear to have systematically lower costs per call than the SP or UP schemes, without having systematically higher abandonment rates. Similarly the 100-scenario versions of SP, UP, and RP all have abandonment rates and costs per call that do not appear to be significantly better than those for their 4-scenario counterparts.

G Computation Times

All the numerical tests were run using OPL-CPLEX Optimization Studio 12.6 on a dedicated PC with a 2.4 GHz Intel Core 2 Quad CPU and 8 GB of RAM.

G.1 Computation Times for Quadrature and Sampling-Based Scenario Generation

In this part, we report computation times for the tests performed in §2.5.1.

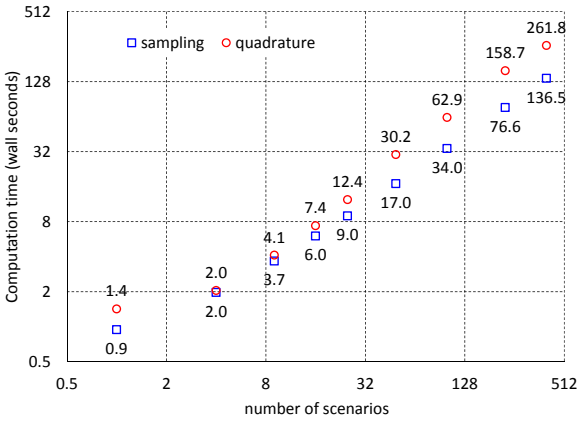

Figure 7: Day 101 Computation Times, by Number of Scenarios

Figure 7 reports so-called "wall" computation times for each method and number of scenarios. That is, the times are the elapsed time recorded by the computer's clock and account for the sometimes significant preprocessing time needed to set up the optimization problems. For each number of scenarios, results for quadrature reflect the solution of a single IP, while those for sampling are the average of solution times of the 100 IP's in each associated set.

The panel's plot shows that, for both methods, computation times appear to grow roughly linearly with the number of scenarios and that the times for quadrature are about double the average computation time (across 100 samples) for larger numbers of scenarios. The longest times, at 400 scenarios, are on average 136 seconds for sampling-based scenarios, and about 260 seconds when using quadrature. (These computation times seem reasonable for practical implementation.) Of course, the longer computation times required by quadrature-

based scenarios are compensated by the fact that IP solutions for quadrature appear to be stable with many fewer scenarios, as little as 4.

G.2 Computation Times for Tests of SPm, UPm, and RPm Schemes

In this part, we report computation times for the tests performed in §4. Figure 8 shows the nine schemes' average wall computation times for the European retail bank call center dataset, in the left panel, and for the North American retail bank call center dataset, in the right panel, respectively

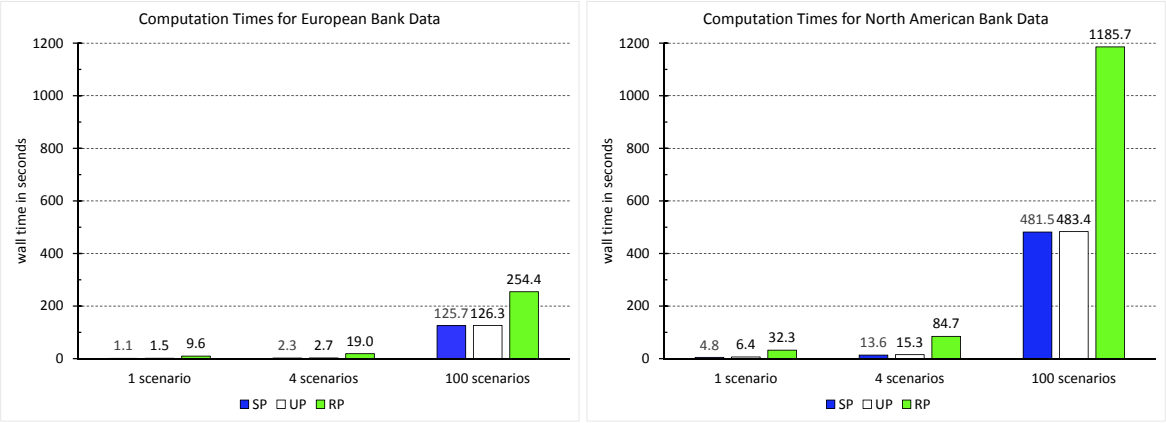

Figure 8: Nine Schemes' Average Computation Times for European (left) and NA (right) Data

Average computation times for the European retail bank call center dataset increase with the number of scenarios and with the complexity of the scheme. All 1 and 4-scenario schemes took less than 20 seconds of wall clock time to set up and solve a given day's mathematical program, on average. RP100 was the most computationally intensive scheme, requiring an average of 254 seconds, about twice the 126-second average required to set up and solve UP100.

Average computation times for the North American retail bank call center dataset were systematically longer than those for the European bank, but with the same relative patterns. For example, UP4 took less than 15 seconds of wall clock time, on average, to set up and solve, while RP4 took a bit less than 85 seconds on average. RP100 was again the most computationally intensive scheme, requiring an average of 1186 seconds – almost 20 minutes – to set up and solve, more than twice the 484-second average required to set up and solve UP100. As noted above, for practical purposes it is sufficient to solve the 4-scenario versions of the recourse programs.

References

S. Aldor-Noiman, P. D. Feigin, and A. Mandelbaum. 2009. Workload Forcasting for a Call Center: Methodology and a Case Study. *Annals of Applied Statistics*, 3(4):1403-1447.

G. P. Bhattacharjee. 1970. Algorithm AS 32: The Incomplete Gamma Integral. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 19(3):285-287.

L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao. 2005. Statistical Analysis of a Telephone Call Center: a Queueing Science Perspective. *J. of the Am. Statistical Association*, 100:36-50.

P. Coolen-Schrijner and E. A. Van Doorn. 2001. On the Convergence to Stationarity of Birth-Death Processes. *Journal of Applied Probability*, 38(3):696-706

Z. Feldman, A. Mandelbaum, W. A. Massey, and W. Whitt. 2007. Staffng of Time-Varying Queues to Achieve Time-Stable Performance. *Management Science*, 54:324-338.

N. Gans, H. Shen, H. Ye, and Y.-P. Zhou. 2014. Long-run Average Optimality of AR(1)-Driven Workforce Scheduling Models. In preparation.

L. V. Green, P. J. Kolesar, and W. Whitt. 2007. Coping with Time-Varying Demand When Setting Staffng Requirements for a Service System. *Production and Operations Management*, 16:13-39.

R. Ibrahim and P. L'Ecuyer. 2013. Forecasting Call Center Arrivals: Fixed-Effects, Mixed-Effects, and Bivariate Models. *Manufacturing & Service Operations Management*, 15(1):72-85.

S.-H. Kim and W. Whitt. 2014. Are Call Center and Hospital Arrivals Well Modeled by Nonhomogeneous Poisson Processes? *Manufacturing & Service Operations Management*, 16(3):464-480.

A. Mandelbaum and S. Zeltyn. 2007. The M/M/n+G Queue: Summary of Performance Measures. Technical Note, Technion, Israel Institute of Technology.

V. Mehrotra, O. Ozluk, and R. Saltzman. 2010. Intelligent Procedures for Intra-Day Updating of Call Center Agent Schedules. *Production and Operations Management*, 19(3):353-367.

A. C. Miller and T. R. Rice. 1983. Discrete Approximations of Probability Distributions. *Management Science*, 29:352-362.

B. N. Oreshkin, N. Regnard, and P. L'Ecuyer. 2014. Rate-Based Daily Arrival Process Models with Application to Call Centers. Working Paper, Université de Montréal.

T. R. Robbins and T. P. Harrison. 2010. A Stochastic Programming Model for Scheduling Call Centers with Global Service Level Agreements. *European Journal of Operational Research*, 207:1608-1617.

H. Shen and J. Z. Huang. 2008. Interday Forecasting and Intraday Updating of Call Center Arrivals. *Manufacturing & Service Operations Management*, 10:391-410.

W. Wang, S. Ahmed. 2008. Sample Average Approximation of Expected Value Constrained Stochastic Programs. *Operations Research Letters*, 36:515-519.

J. Weinberg, L. D. Brown, and J. R. Stroud. 2007. Bayesian Forecasting of an Inhomogeneous Poisson Process with Applications to Call Center Data. *Journal of the American Statistical Association*, 102:1185-1199.

W. Whitt. 1999. Dynamic Staffing in a Telephone Call Center Aiming to Immediately Answer All Calls. 1999. *Operations Research Letters*, 24:205-212.

E. Zohar, A. Mandelbaum, and N. Shimkin. 2002. Adaptive Behavior of Impatient Customers in Tele-Queues: Theory and Empirical Support. *Management Science*, 48(4):566-583.