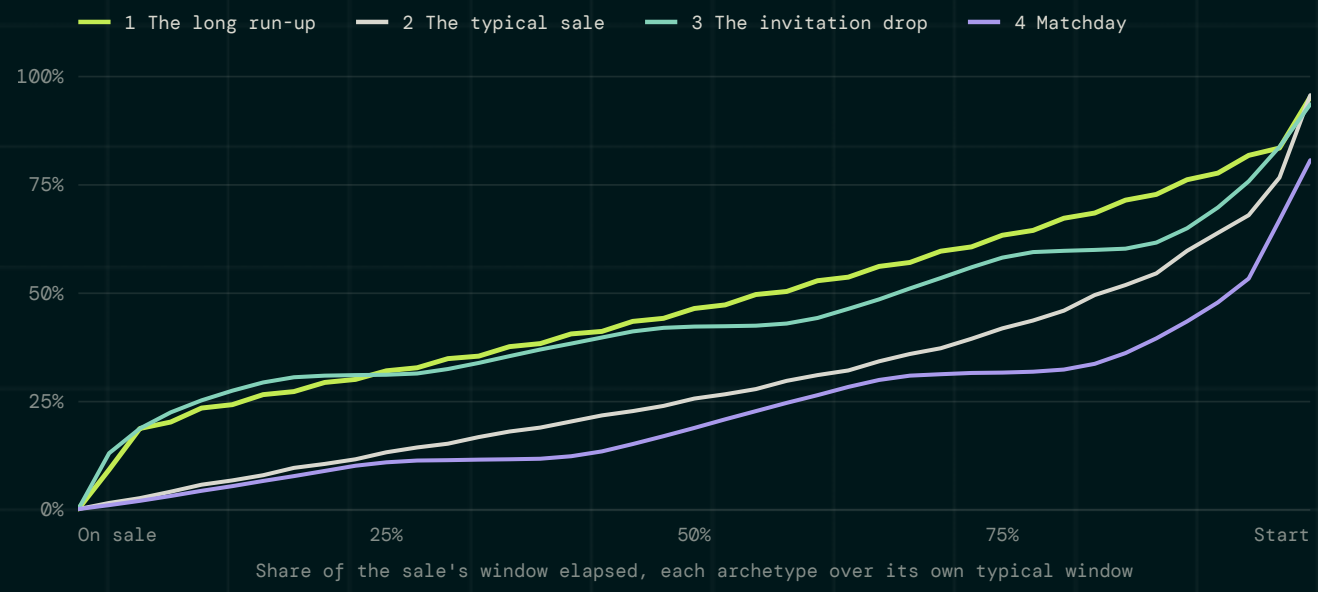
loft

PAPER

APPLIED STATISTICS

Forecasting sell-outs in advance ticket sales with a Bayesian mixture model

Sales curves, the effect of capacity, and a limit on the precision of sell-out timing.



PREPARED BY

Tendo Michael and Rodgers Tenyigwa

PUBLISHED

September 2026

READ ONLINE

loft.ug/research/sellout-model

ABOUT THIS PAPER

PREPARED BY	Tendo Michael and Rodgers Tenyigwa
ORGANISATION	Loft Technologies, September 2026
SERIES	Paper · Applied statistics
PUBLISHED	September 2026

CONTENTS

1 Introduction 3
2 The model 4
3 What the model implies 8
4 Consequences for an organiser 9
5 Evaluation 10
6 Scope 10
7 Conclusions and research directions 10
References 12

HOW TO CITE

Tendo, M. and Tenyigwa, R. (2026). Forecasting Sell-outs in Advance Ticket Sales with a Bayesian Mixture Model. Paper. Loft Technologies.

Read online and download: loft.ug/research/sellout-model

0.92
AUC of the sell-out forecast once a tenth of capacity is sold, out of sample

6%
Median error in the forecast of final sales at half sold, as a share of capacity

×1.4
Buyers drawn by an allocation twice as large: demand grows as capacity to the power 0.50

1/√k
The lowest relative uncertainty in a sell-out time when k orders are still to come

ABSTRACT

We propose a model of how demand for a ticket type arrives over an advance sale, and use it to forecast whether the ticket type will sell out before the event and, if so, when. The cumulative share of demand is modelled as the sum of three components: purchases following the announcement, purchases building up to the event date, and purchases in the hours around the start. Time is measured on a clock weighted by the daily pattern of purchasing. The weights and timescales of the components are drawn from a small number of archetypes, selected by the attributes of the sale. Demand is assumed to scale with capacity as c^{α} with $\alpha < 1$, and we estimate $\alpha \approx 0.50$ across past sales; purchases arrive as discrete orders. The model has three implications. First, the probability that a ticket type sells out is largely determined by its capacity before any purchase is made, and decreases steeply with it. Second, the time of a sell-out cannot be forecast with a relative uncertainty below $1/\sqrt{k}$ when k orders remain, so it must be forecast as an interval. Third, the share of buyers who switch to the next, more expensive ticket type after a sell-out is not distinguishable from zero in past sales; the model therefore reports the share that ticket type would need in order to sell out, rather than a probability. Forecasts are obtained by Bayesian inference over the archetypes from the orders observed so far. Out of sample, the sell-out probability has an AUC of 0.92 when 10% of capacity has been sold, and the median absolute error of the forecast of final sales is 15% of capacity at that stage and 6% at half sold. The model is more accurate than naive rate extrapolation and than the pickup method of revenue management, whose forecasts are biased downward by the capacity limits of the past sales from which they are derived.

KEYWORDS

ticket sales forecasting; sell-out; Bayesian mixture model; censored demand; revenue management; pickup method; event ticketing

§ 1 Introduction

During a sale, an organiser needs to know whether a ticket type will sell out before the event, when this will happen, and, if it will not, how many tickets will remain. The answers inform pricing, promotion and the release of further ticket types.

Advance sales have been forecast for half a century in **revenue management**, from Littlewood's work on airline bookings[1] to the booking-curve and *pickup* methods that remain its standard: project the final total from the bookings on hand and the growth that similar flights or nights showed from the same point[2]. That literature also identified the difficulty that shapes this problem. A flight that fills records its capacity, not its demand; recorded sales are *censored*, and demand must be unconstrained before it can be forecast[3][4][5]. **Marketing science** models the timing of purchases of new products as mixtures of timing distributions, of eager buyers and followers[6], or of a small number of trajectory classes whose scale, but not shape, can be predicted before release[7]. Event tickets specifically have received little attention[8][9]. **Operations research** on strategic customers explains why the timing of a purchase responds to the risk of stock running out[10][11], and collective attention is known to arrive in bursts, triggered from outside or propagated from person to person[12].

Censoring is the central difficulty here. A ticket type that sells out records its capacity rather than its demand, and the sales that matter most for the question, those that sell out, are the ones whose demand is not fully observed. Forecasts that extrapolate from past sales therefore inherit their capacity limits. We instead specify

a model of the sale process whose parameters can be estimated from the observed part of each sale, and derive forecasts from it.

We proceed as follows. § 2 states the model as six assumptions and gives the inference that turns a live sale's orders into a forecast. § 3 derives three implications: the prior probability of a sell-out as a function of capacity, a limit on the precision with which the time of a sell-out can be forecast, and a threshold for the next ticket type when a cheaper one sells out. § 4 discusses their use by organisers. § 5 evaluates the forecasts out of sample and compares them with alternative methods. The companion paper on crowd arrival[13] derives its arrival curve from individual choices; the curve here is descriptive, and we discuss a behavioural derivation in § 7.

§ 2 The model

The unit of analysis is a ticket type, with a capacity c, a time at which sales open, and a close, the start of the event. A ticket type is *sold out* when 95% of its capacity has been sold, since organisers commonly hold back a small buffer, or when its sales stop abruptly at a point where its recent rate would have exhausted the remainder. Demand after a sell-out is not observed. Demand consists of purchase decisions: paid orders, and registrations for free ticket types. Tickets issued by the organiser (complimentary tickets, invitations, allocations to sponsors) use capacity but are not counted as demand.

(A1) Purchasing follows an activity clock. The rate of purchasing varies strongly with the time of day. Time is therefore measured in activity hours: each clock hour is weighted by the share of purchases made in it, with separate profiles for the final 36 hours before the close and for earlier. Between 1 and 5 a.m. the clock runs at 7% of its afternoon rate. All timescales below are in activity hours.

(A2) Demand has a launch component, a deadline component and a door component. The share of a ticket type's eventual demand bought by time t is

$$F_\theta(t) = w_H\, L(t) + w_D\, R(t) + w_0\, D(t), \qquad w_D = 1 - w_H - w_0 \tag{1}$$

where, in activity time $a(t)$, with $Y = a_{\text{close}} - a_{\text{on}}$ the length of the window, the launch and the deadline ramp are Weibull curves normalised to run from zero to one over it,

$$L(t) = \frac{1 - e^{-(x/\mu)^{\kappa_\ell}}}{1 - e^{-(Y/\mu)^{\kappa_\ell}}}, \quad x = a(t) - a_{\text{on}} \tag{2}$$

$$R(t) = \frac{e^{-(y/\lambda)^{\kappa_d}} - e^{-(Y/\lambda)^{\kappa_d}}}{1 - e^{-(Y/\lambda)^{\kappa_d}}}, \quad y = a_{\text{close}} - a(t) \tag{3}$$

and the door is spread evenly over the three hours either side of the close, in clock time,

$$D(t) = \min\!\Big(1,\ \max\!\big(0,\ \tfrac{t - t_{\text{close}} + 3\,\text{h}}{6\,\text{h}}\big)\Big) \tag{4}$$

(the launch also runs through those three hours). The launch share w_H is bought on a timescale μ after selling opens (steepness $\kappa_\ell = 1$ decays from the start; $\kappa_\ell = 2$ builds first, as news spreads); the deadline share w_D arrives on a timescale λ before the close; the door share is w_0. The launch and deadline components

correspond to demand released by the announcement and by the approach of the event. They change the rate of sales in opposite directions: the first decays over time and the second grows. A forecast that extrapolates the current rate therefore overestimates the remaining sales when the launch component dominates, and underestimates them when the deadline component does.

(A3) Sales follow a small number of archetypes, selected by their attributes. Each archetype k is a distribution over the curve's parameters,

$$p_k(\theta) \;\propto\; a_k(w_H)\, d_k(w_0)\, g_k(\mu, \kappa_\ell)\, g'_k(\lambda, \kappa_d) \tag{5}$$

$$g_k(\mu, \kappa_\ell) = \mathcal{N}\big(\log \mu;\; m_k, s_k^2\big)\, c_k(\kappa_\ell) \tag{6}$$

with g'_k of the same form for the deadline timescale and steepness, and a_k, d_k and c_k probabilities over the few values each share and steepness takes. The probability that a sale follows each archetype depends on its attributes x: the length of the sale; whether the event is sport, music or cinema; whether admission is free; and whether capacity is set for the event as a whole rather than for the ticket type,

$$\pi_k(x) = \frac{e^{\gamma_k^\top x}}{\sum_{j=1}^{K} e^{\gamma_j^\top x}} \tag{7}$$

(A4) Demand grows sublinearly with capacity. Total demand A relative to capacity, $\rho = A/c$, is log-normal within each archetype, and its centre moves with the sale's attributes z, capacity first among them:

$$p(\theta, \rho \mid x, z) = \sum_{k=1}^{K} \pi_k(x)\, p_k(\theta)\, \mathcal{N}\big(\log \rho;\; \nu_k + \beta^\top z,\; \sigma_k^2\big) \tag{8}$$

with the capacity term $\beta_c \log c$, $\beta_c = \alpha - 1$, so that demand grows as c^α. If organisers sized allocations in proportion to their audience, α would be one and ρ would not depend on capacity. We postulate $\alpha < 1$: larger allocations are set further above their expected demand. Across past sales the logarithmic slope of ρ on capacity is about −0.50, so we take

$$\rho \;\propto\; c^{\alpha - 1}, \qquad \alpha \approx 0.50 \tag{9}$$

so an allocation twice as large attracts about 1.4 times as many buyers. This is a comparison across sales, not a causal statement about a single sale: it does not imply that enlarging an allocation would attract more buyers. It describes what an allocation of a given size can expect before sales begin.

(A5) Purchases arrive as orders. A group booking is a single decision, so the arrivals are orders of about q tickets each, forming a Poisson process given the curve. Orders are over-dispersed relative to the Poisson distribution, since a shared post or a group conversation can produce several at once.

(A6) Some buyers turned away by a sell-out may switch to the next ticket type. When a ticket type sells out, a share η of the buyers still arriving for it may buy the next, more expensive ticket type. Estimated from that ticket type's sales in excess of its own forecast after the cheaper one sold out, η is 0.09 (95% interval −0.19 to

0.33), which is consistent with no switching. The model therefore sets $\eta = 0$ in its forecasts and reports the share that would be required instead (§ 3.3).

2.1 The archetypes

We fit (A2)–(A4) by expectation–maximisation[14] to past sales with at least 100 buyer orders; smaller sales contain too few orders to determine a curve. Each sale's orders are used only up to the time it sold out. Scored on later sales, the held-out log-likelihood per sale varies by less than 0.01 between four and eight archetypes, and the forecasts are equivalent, so we use the smallest number, $K = 4$. Replacing the selection rule (A3) with fixed archetype probabilities lowers the held-out log-likelihood by 0.07 per sale. The four archetypes differ mainly in how purchases are distributed over the sale (Table 1, Figure 1). **The typical sale**, three quarters of all sales, is bought mostly in its final week. **The long run-up** is bought steadily over the preceding four weeks, as with festival releases and theatre runs, and about half of these sales sell out. **The invitation drop**, consisting of free and invitation-only releases, has nearly half of its tickets claimed when it opens. **Matchday**, covering league football and large concerts in big venues, sells over a third of its tickets in the three hours either side of the start and rarely sells out. The fitted model has 114 parameters.

Table 1. The four archetypes as calibrated. Window and capacity are medians over the sales each archetype explains best; under each name, how it sells over its own typical window.

#	ARCHETYPE	TYPICAL WINDOW	TYPICAL CAPACITY	SHARE OF SALES	SOLD OUT
1	**The long run-up**. Bought steadily over the four weeks before the event, as with festival releases and theatre runs. 19% bought in the first 5% of its window, 46% by half-way, 5% after the start.	19.8 days	450	16%	54%
2	**The typical sale**. The most common kind: bought mostly in the final week. 2% bought in the first 5% of its window, 26% by half-way, 4% after the start.	27.7 days	600	77%	26%
3	**The invitation drop**. Free and invitation-only releases: nearly half claimed when they open. 19% bought in the first 5% of its window, 42% by half-way, 7% after the start.	3.4 days	600	1%	9%
4	**Matchday**. League football and large concerts in big venues: over a third bought in the hours around the start. 2% bought in the first 5% of its window, 19% by half-way, 20% after the start.	2.4 days	3,000	6%	4%

The four ways a ticket type sells

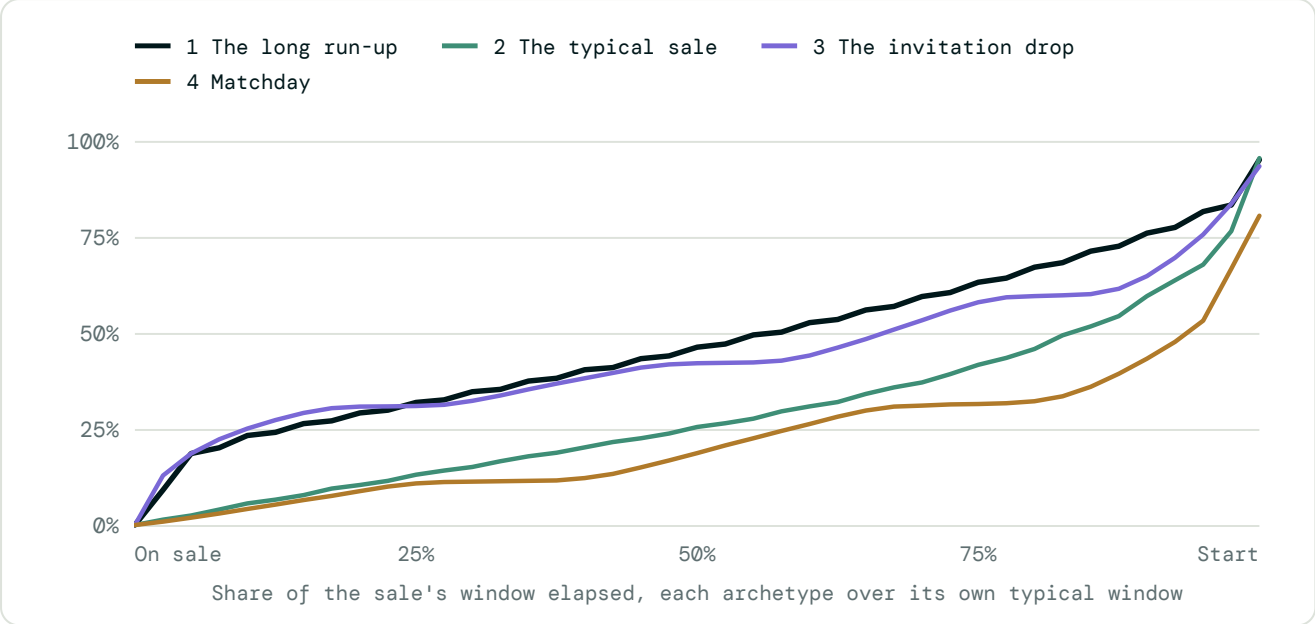


Figure 1. The four archetypes as curves, each over its own typical window: the expected share of demand bought as the window elapses, from going on sale to the start. The flat steps are nights.

2.2 Forecasting a live sale

For a sale in progress with N orders so far, grouped into intervals b holding n_b orders each, the curve puts a share $\Delta F_{\theta,b}$ of demand in each interval. By (A5) the orders' likelihood, tempered by the sale's over-dispersion ϕ, is

$$\ell(\theta, A) = \frac{1}{\phi}\Big[N \log \tfrac{A}{q} + \sum_b n_b \log \Delta F_{\theta,b} - \tfrac{A}{q} F_\theta(t_{\text{now}})\Big] \tag{10}$$

The tempering is measured on the sale itself, as the robust Pearson dispersion of its orders around the best-fitting curve, bounded below by a base $\phi_0 = 1.5$ and above by a rule that grows with the orders per interval:

$$\phi = \min\Big(\phi_0\big(1 + \tfrac{N}{n_{\text{int}}}\big),\ \max\big(\phi_0,\ \tfrac{\operatorname{median}_b\, r_b^2}{0.455}\big)\Big), \qquad r_b = \frac{n_b - \hat{n}_b}{\sqrt{\hat{n}_b}} \tag{11}$$

where $\hat{n}_b$ are the orders the best-fitting curve expects in interval b, n_{int} is the number of intervals, and 0.455 is the median of χ^2_1, its expected value for Poisson counts. A regular series of orders is therefore given more weight than an irregular one. Combined with the prior (A3)–(A4) the likelihood gives the posterior, evaluated on a grid of 5,088 curves and 32 levels of demand. The rest of the sale is then simulated: a curve and a total are drawn from the posterior, the demand still to come is multiplied by a random shock representing events not yet reflected in the orders, and orders arrive at the curve's rate,

$$A_{\text{future}} = A\,(1 - F_{\text{now}})\, e^{\sqrt{1 - F_{\text{now}}}\, Z}, \qquad Z \sim \mathcal{N}(0,1), \quad F_{\text{now}} = F_\theta(t_{\text{now}}) \tag{12}$$

$$\text{orders in } [t, t+\delta] \ \sim\ \text{Poisson}\Big(\frac{A_{\text{future}}}{q} \cdot \frac{F_\theta(t+\delta) - F_\theta(t)}{1 - F_{\text{now}}}\Big) \tag{13}$$

so the shock is larger the more of the sale remains, and leaves the median forecast unchanged. The forecast is the distribution of these simulated outcomes: the number sold by the close and, derived from it, the probability that the ticket type sells out, the time at which it does so, and otherwise the number left. Forecasts are produced once a ticket type has 10 buyer orders. With fewer, none of the methods we evaluated improved materially on the base rate of sell-outs, and a rate estimated from k orders has a relative uncertainty of about $1/\sqrt{k}$ under any model.

§ 3 What the model implies

3.1 Capacity largely determines the probability of a sell-out

Before the first order, the forecast is given by the prior alone. By (A4) the chance that a ticket type sells out is the chance that its demand reaches 95% of capacity,

$$P_0(c) = \sum_{k=1}^{K} \pi_k(x)\ \Phi\Big(\frac{\nu_k + \beta^\top z - \log 0.95}{\sigma_k}\Big) \tag{14}$$

and capacity enters only through $\beta_c \log c = (\alpha - 1)\log c$. With $\alpha = 1$ it would not enter at all, and the probability of a sell-out would not depend on capacity. With $\alpha \approx 0.50$ every doubling of capacity multiplies the expected fill by $2^{\alpha-1} \approx 0.71$, and the probability of a sell-out falls steeply: for a typical paid ticket type on sale for ten days it is 97% at 20 seats, 59% at 300 and 14% at 2,000 (Figure 2). The choice of capacity therefore largely determines the probability of a sell-out before sales begin.

The bigger the allocation, the less likely it sells out

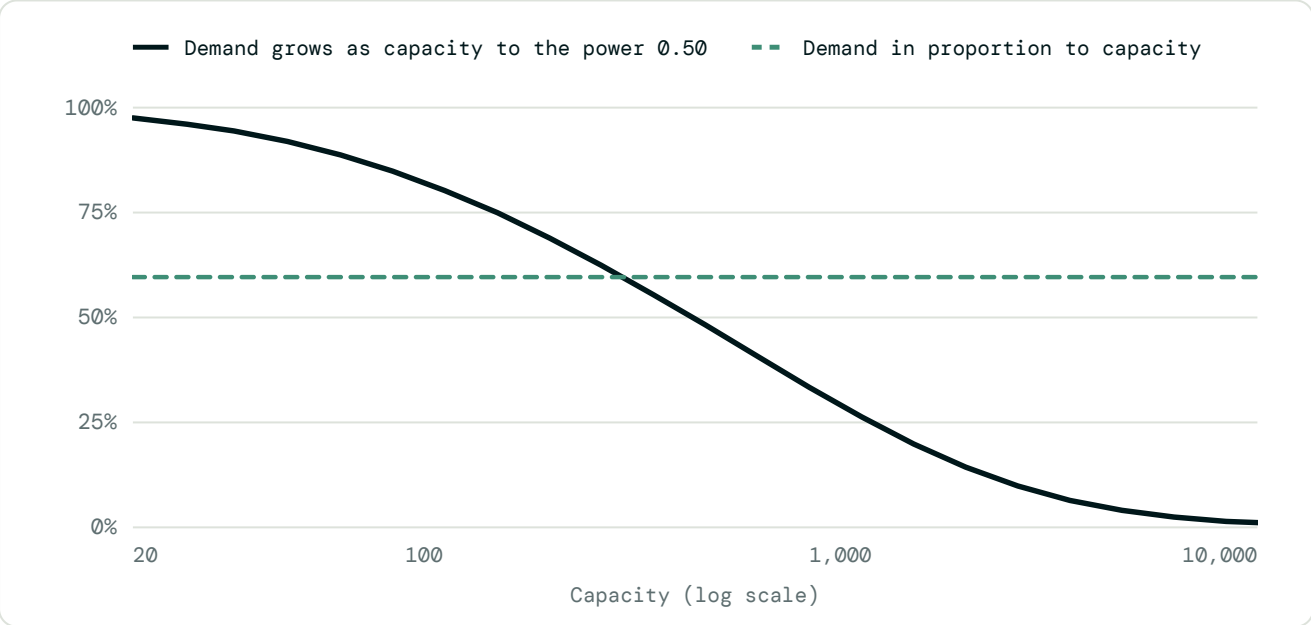


Figure 2. The model's prior probability that a typical paid ticket type on sale for ten days sells out, before any order, against its capacity: with demand growing as capacity to the power 0.50, and with demand in proportion to capacity. The two curves coincide at a capacity of 300.

3.2 A limit on the precision of the time of a sell-out

Near the end of a sale, few orders remain. If k orders remain and, by (A5), they arrive at rate r, the time until the last of them is a sum of k exponential waits:

$$T_k \sim \mathrm{Gamma}(k, r), \qquad \frac{\mathrm{sd}(T_k)}{\mathbb{E}\,T_k} = \frac{1}{\sqrt{k}} \tag{15}$$

This is a lower bound, attained only if the rate is known exactly:

Table 2. The precision limit, $1/\sqrt{k}$.

ORDERS STILL TO COME, K	1	3	10	30	100
Lowest possible relative uncertainty in the time remaining	100%	58%	32%	18%	10%

The bound matters most when the question is most often asked. Organisers usually ask when a ticket type will sell out once it is nearly sold out, and at that point few orders remain; for an allocation of a few hundred tickets, the final hours of a sale may contain only a handful. The time of such a sell-out cannot be determined precisely from the sale's data, by this or any other model, and should be forecast as an interval. For this reason § 2.2 simulates the remainder of a sale order by order rather than along a smooth curve, so that the reported intervals include this uncertainty.

3.3 A threshold for the next ticket type

When a cheaper ticket type is about to sell out, whether the next, more expensive one also sells out depends on three quantities the model estimates: the buyers expected to be turned away from the cheaper ticket type, O; the sales the next ticket type still needs, M; and its own buyers still to come, B. By (A6) it sells out when its own buyers and those who switch together cover what it needs, $B + \eta\, O \geq M$, that is when

$$\eta \;\geq\; s^{*} = \frac{M - B}{O} \tag{16}$$

If $s^{*} \leq 0$ it sells out on its own buyers; if $s^{*} > 1$ it cannot, whatever share switches. Between the two, the outcome depends entirely on η, which is not distinguishable from zero in past sales (A6). A probability computed from an assumed value of η would conceal this dependence; the threshold states it. An organiser cannot observe η, but can often judge whether, for example, one in three of the buyers turned away would pay the higher price.

§ 4 Consequences for an organiser

Choosing a capacity. Given the attributes of a sale, $P_0(c)$ gives the probability of a sell-out implied by a proposed capacity before release. Read the other way, it gives the capacity at which a ticket type has even odds of selling out: about 440 seats for a typical paid ticket type on sale for ten days. A first release of 50 tickets has a prior probability of selling out of about 91%; one of 2,000, about 14%.

Interpreting a sell-out time. The model reports a sell-out time when a sell-out is the more likely outcome, and otherwise the expected number of unsold tickets. Near the end of a small sale the interval for the time cannot be narrow (§ 3.2), and plans that depend on the time of a sell-out should allow for the whole interval.

Pricing the next ticket type. When a cheaper ticket type is about to sell out, § 3.3 gives the share of its turned-away buyers that the next ticket type needs. If the share is small, the next ticket type is likely to sell out at its current price; if it is large, it is likely to be left with unsold tickets unless conditions change. Between the two,

whether the required share is plausible is a judgement for the organiser, and the threshold presents it as such rather than as a probability.

§ 5 Evaluation

The model was fitted once, with all modelling choices fixed, and applied without refitting to a hold-out set of ticket types[15][16], each forecast when 10%, 25%, 50% and 75% of its capacity had been sold. Table 3 reports, for the forecast of final sales, the mean quantile loss at the 10th, 50th and 90th percentiles[17][18], the median absolute error as a share of capacity, and the coverage of the central 80% interval; for the sell-out probability, the AUC[19] and the Brier skill score relative to the base rate at each stage[20]; and, for ticket types that sold out, the median ratio error $\max(\hat{\tau}/\tau,\ \tau/\hat{\tau})$ of the forecast time remaining. The benchmarks are pickup[2] and naive extrapolation of the current sales rate.

Table 3. Forecast accuracy on the hold-out set, all stages pooled. The best value in each column is in bold (for coverage, the value closest to 80%). Naive extrapolation provides no interval.

METHOD	QUANTILE LOSS	MEDIAN ABS. ERROR	80% COVERAGE	AUC	BRIER SKILL	TIME RATIO ERROR
Proposed model	**0.047**	**9.7%**	**79%**	**0.91**	**0.39**	**1.45**
Pickup	0.077	16.1%	76%	0.85	−0.10	—
Naive rate extrapolation	—	—	—	0.81	0.09	1.99

On the hold-out set the sell-out probability discriminates well from the first stage (AUC 0.92 at 10% sold), and the median absolute error of the forecast of final sales falls from 15% of capacity at that stage to 6% at half sold. Pickup has a higher quantile loss, by 0.030 (95% CI 0.015 to 0.045; cluster bootstrap over ticket types[21]), and is biased downward early in a sale: at 10% sold its median error for ticket types that sold out is −72.5% of capacity, against −0.2% for the model, as expected from censoring (§ 1). For the time to sell-out, the median ratio error of naive extrapolation is 1.99, against 1.45 for the model.

§ 6 Scope

The model has no information about a price change, a line-up announcement or a competing event until its effect appears in the orders. Demand beyond capacity is never observed, so how far demand exceeds supply is an extrapolation of the curve (A2). The capacity law (A4) compares sales and says nothing causal about changing one sale's allocation. Its exponent, and the finding that the switching share (A6) is not measurably above zero, are estimated in a single market in which tickets are bought almost entirely by mobile phone, and both may differ elsewhere.

§ 7 Conclusions and research directions

The central difficulty in forecasting sell-outs is censoring: the sales that matter most record their capacity, not their demand. Modelling the sale process, rather than extrapolating past totals, avoids it, and out of sample the model's forecasts are close to unbiased where those of pickup are not. Its limits are structural rather than statistical: the time of a sell-out cannot be known more precisely than $1/\sqrt{k}$ with k orders remaining, and the fate of the next ticket type turns on a switching share that past sales do not identify.

We see two directions for further work. First, the curve should be derived rather than described: purchase timing as a choice between the cost of committing early and the risk of missing out, which grows as stock runs down, in the manner of the arrival model[13], would explain the archetypes with fewer free parameters, and would predict the quickening of a sale as its last tickets go that the strategic-customer literature describes[10]. Second, the switching share η should be measured more precisely, since the model's answers for the next ticket type depend on it more than on any other parameter.

References

1 K. Littlewood, “Forecasting and control of passenger bookings”, *AGIFORS Symposium Proceedings* 12, 1972, 95–117.

2 L. R. Weatherford and S. E. Kimes, “A comparison of forecasting methods for hotel revenue management”, *International Journal of Forecasting* 19, 2003, 401–415.

3 K. T. Talluri and G. J. van Ryzin, *The Theory and Practice of Revenue Management*, Springer, 2004.

4 L. R. Weatherford and S. Pölt, “Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues”, *Journal of Revenue and Pricing Management* 1(3), 2002, 234–254.

5 P. Guo, B. Xiao and J. Li, “Unconstraining methods in revenue management systems: research overview and prospects”, *Advances in Operations Research*, 2012, 270910.

6 W. W. Moe and P. S. Fader, “Using advance purchase orders to forecast new product sales”, *Marketing Science* 21(3), 2002, 347–364.

7 M. S. Sawhney and J. Eliashberg, “A parsimonious model for forecasting gross box-office revenues of motion pictures”, *Marketing Science* 15(2), 1996, 113–131.

8 D. S. Putler and S. Lele, “An easily implemented framework for forecasting ticket sales to performing arts events”, *Marketing Letters* 14(4), 2003, 307–320.

9 J. A. Suher, “Forecasting event ticket sales”, University of Pennsylvania, 2008.

10 Q. Liu and G. J. van Ryzin, “Strategic capacity rationing to induce early purchases”, *Management Science* 54(6), 2008, 1115–1131.

11 X. Su, “Intertemporal pricing with strategic customer behavior”, *Management Science* 53(5), 2007, 726–741.

12 R. Crane and D. Sornette, “Robust dynamic classes revealed by measuring the response function of a social system”, *PNAS* 105(41), 2008, 15649–15653.

13 J. Katende and N. Mugalula, “Modelling crowd arrival at stadium turnstiles”, Loft Technologies, September 2026. `loft.ug` ↗

14 A. P. Dempster, N. M. Laird and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm”, *Journal of the Royal Statistical Society B* 39(1), 1977, 1–38.

15 A. P. Dawid, “Statistical theory: the prequential approach”, *Journal of the Royal Statistical Society A* 147(2), 1984, 278–292.

16 L. J. Tashman, “Out-of-sample tests of forecasting accuracy: an analysis and review”, *International Journal of Forecasting* 16(4), 2000, 437–450.

17 T. Gneiting and A. E. Raftery, “Strictly proper scoring rules, prediction, and estimation”, *Journal of the American Statistical Association* 102(477), 2007, 359–378.

18 T. Gneiting, F. Balabdaoui and A. E. Raftery, “Probabilistic forecasts, calibration and sharpness”, *Journal of the Royal Statistical Society B* 69(2), 2007, 243–268.

19 J. A. Hanley and B. J. McNeil, “The meaning and use of the area under a receiver operating characteristic (ROC) curve”, *Radiology* 143(1), 1982, 29–36.

20 A. H. Murphy, “A new vector partition of the probability score”, *Journal of Applied Meteorology* 12(4), 1973, 595–600.

21 B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman & Hall, 1993.