Analysis of Call Center Services

IEOR, UC Berkeley

What is a call center

- Point of contact between a firm and customers
- Large pool of customer service representatives (CSRs) who
  - *Incoming* – respond to inquiries, take orders, handle customer requests
  - *Outgoing* – call customers

Call center resources

- People
- Computers
- Telecommunication equipment
- Software
  - Routing calls
  - Supporting CSRs

Typical types of Call Centers

- Customer service
- Help desk
- Emergency response services
- Telemarketing

Generic vs. Skill-based routing

- Generic
  - All calls require same **low** level of service
  - All CSRs can handle all calls

- Skill based routing
  - Language based
  - Problem/program specific
    - Information, New sales, Returns, …
    - Word, Excel, Powerpoint, …

Single vs. multi-layered

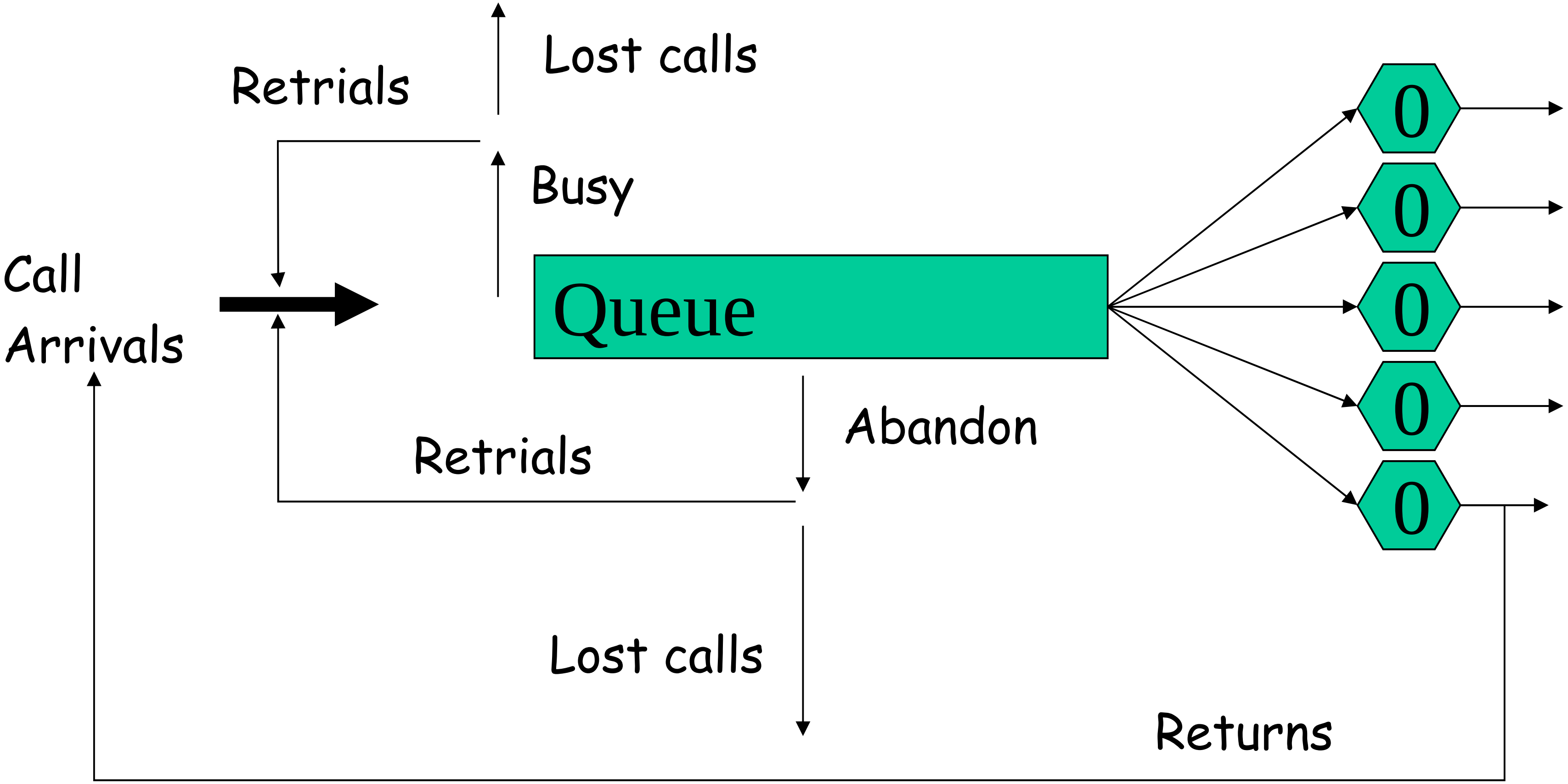
- Single
  - One CSR can handle a full range of issues
- Multi-layered
  - Customers moved between CSRs depending on needs

What are the key tradeoffs?
(in required training,
in customer experience)

IVR –
Interactive Voice Response

- CSRs are replaced (partially) by a voice or touchpad activated computer
  - Airline arrival/reservation info
- Reduces need for employees? Maybe?

Basic model structure



Key issue

- Call centers operate with MANY servers, high arrival rates, but with low probabilities of waiting and low wait times.

How do they do it?

Key decisions

- # of trunk lines
- # of CSRs in each period
  - – *Short term* – number to start each hour of the day
  - – *Long term* – number to train

Outline

- Square root law for Service Center staffing
- Long term planning – number of CSRs to train to ensure we have the right number τ months from now, where τ is the training time in months.

Derivations of square root rule

- BOTTOM LINE: Basic equation for number of CSRs to employ to attain a given level of service:

$$N = R + \beta\sqrt{R}$$

where

N is the number of CSRs

$R = \lambda/\mu$

β is a parameter related to the service level or P(wait)

Approximate derivation of square root law

If we had an infinite server system, the number of busy servers found by an arriving customer would be Poisson with parameter R=λ/μ. This is true for an M/G/∞ queue. To see this for an M/M/∞ queue, see the slide two after this.

Now if delays are not prevalent, the number of busy servers is still almost Poisson and we approximate the Poisson by a Normal with mean R and variance R.

$$\text{So } P(\text{wait}) = P(\#\text{ busy servers} \geq \text{N}) = 1 - \Phi\left(\frac{N-R}{\sqrt{R}}\right)$$

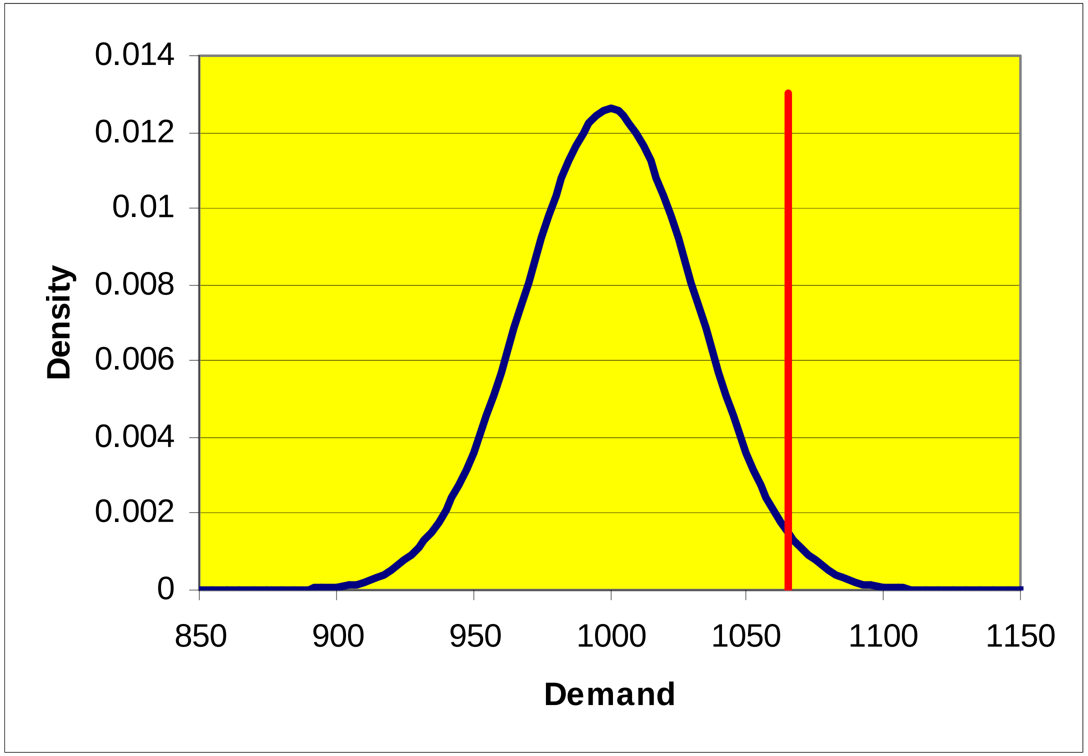
$$\text{So } \beta \approx z_{1\text{-}P(\text{wait})}$$

$$\text{So } \Phi\left(\frac{N-R}{\sqrt{R}}\right) = 1 - P(\text{wait})$$

$$\frac{N-R}{\sqrt{R}} = z_{1-P(\text{wait})}$$

$$N = R + z_{1-P(\text{wait})}\sqrt{R}$$

Example



$$\text{For } \lambda/\mu = 1000$$

$$1065 \text{ servers are needed for}$$

$$P(\text{wait} > 0) = 0.02$$

$$\text{Since}$$

$$\Phi\left(\frac{1065\text{-}1000}{\sqrt{1000}}\right) = 0.98$$

$$\text{So } \Delta = \beta\sqrt{R} = \beta\sqrt{\lambda/\mu} = 65$$

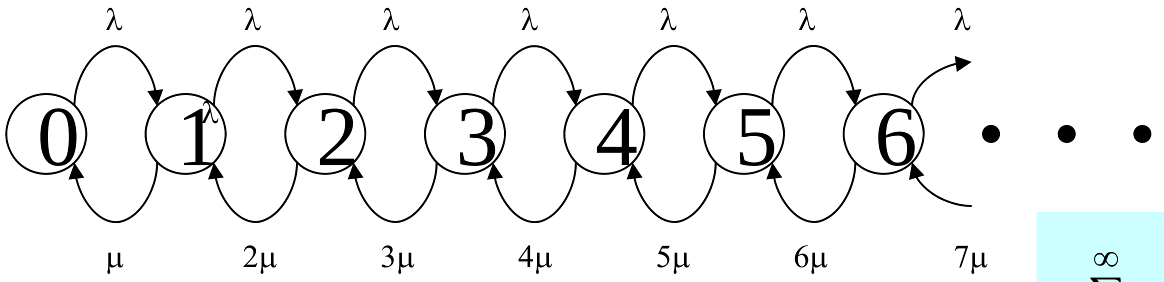
$$\text{or } \beta = 65\Big/\sqrt{\lambda/\mu} = 2.0555$$

For P(wait)=0.01, we get β=2.326

Square root law

Long term planning

Aside: M/M/∞ (Self-service) Queue



$$P_1 = \frac{\lambda}{\mu} P_0$$

$$P_2 = \frac{\lambda}{2\mu} P_1 = \frac{1}{2}\left(\frac{\lambda}{\mu}\right)^2 P_0$$

$$P_3 = \frac{\lambda}{3\mu} P_2 = \frac{1}{3!}\left(\frac{\lambda}{\mu}\right)^3 P_0$$

$$P_n = \frac{\lambda}{n\mu} P_{n-1} = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0$$

$$\sum_{n=0}^{\infty} P_n = \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 = 1$$

$$P_0 \sum_{n=0}^{\infty} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n = 1$$

$$P_0 e^{\lambda/\mu} = 1$$

$$P_0 = e^{-\lambda/\mu}$$

Poisson → $$P_n = \frac{(\lambda/\mu)^n e^{-\lambda/\mu}}{n!}$$

Questions

- So what is the mean number of busy servers? $\frac{\lambda}{\mu}$
- The variance of the number of busy servers? $\frac{\lambda}{\mu}$
- How can we approximate the distribution of the number of busy servers? $N\left(\frac{\lambda}{\mu},\frac{\lambda}{\mu}\right)$

Square root law

Long term planning

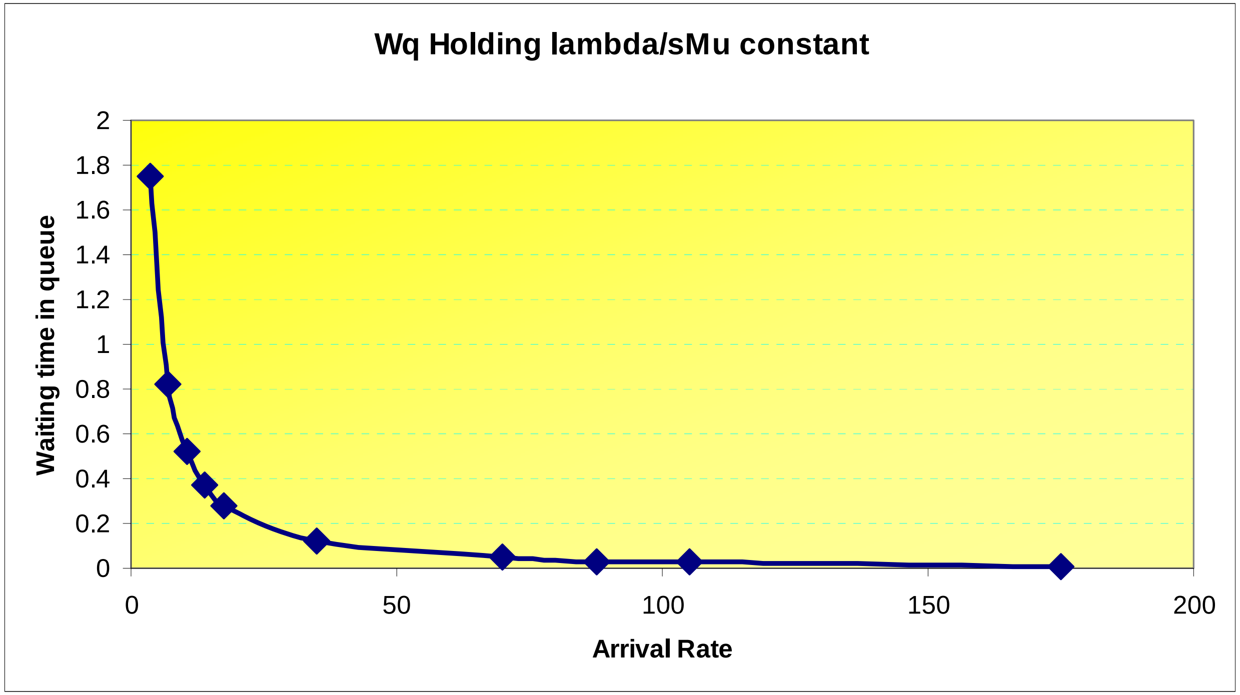
M/M/s queue holding $\lambda/(s\mu)$ constant

INPUTS					OUTPUTS			
Lambda	**Mu**	**Servers**	**$\lambda/(s\mu)$**	**R**	**L**	**W**	**Wq**	**Lq**
3.5	4	1	0.875	0.875	7.00	2.00	1.75	6.13
7.0	4	2	0.875	1.750	7.47	1.07	0.82	5.72
10.5	4	3	0.875	2.625	8.04	0.77	0.52	5.41
14.0	4	4	0.875	3.500	8.67	0.62	0.37	5.17
17.5	4	5	0.875	4.375	9.33	0.53	0.28	4.95
35.0	4	10	0.875	8.750	12.93	0.37	0.12	4.18
70.0	4	20	0.875	17.500	20.74	0.30	0.05	3.24
87.5	4	25	0.875	21.875	24.79	0.28	0.03	2.91
105.0	4	30	0.875	26.250	28.89	0.28	0.03	2.64
175.0	4	50	0.875	43.750	45.62	0.26	0.01	1.87

W_q goes down indicating economies of scale in queueing

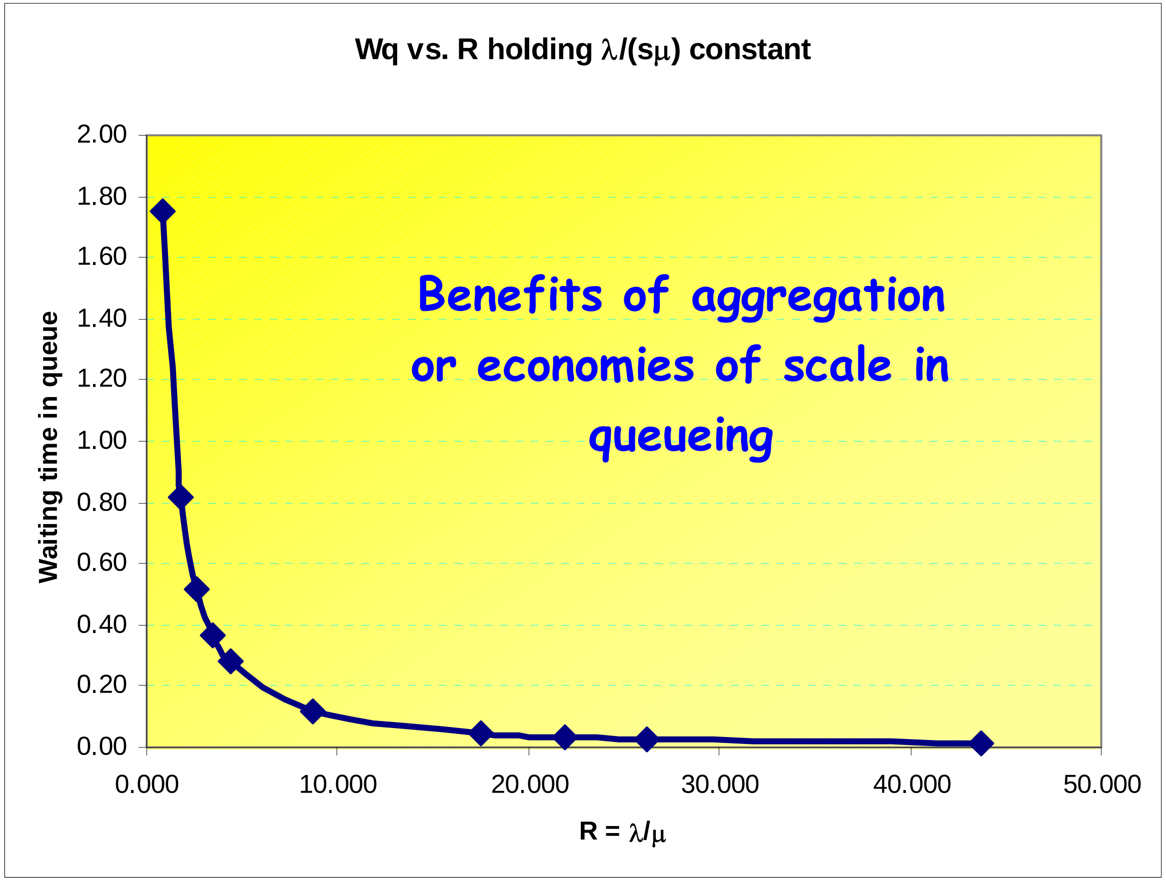
Long term planning

M/M/s queue holding λ/(sμ) constant



W_q goes down indicating economies of scale in queueing

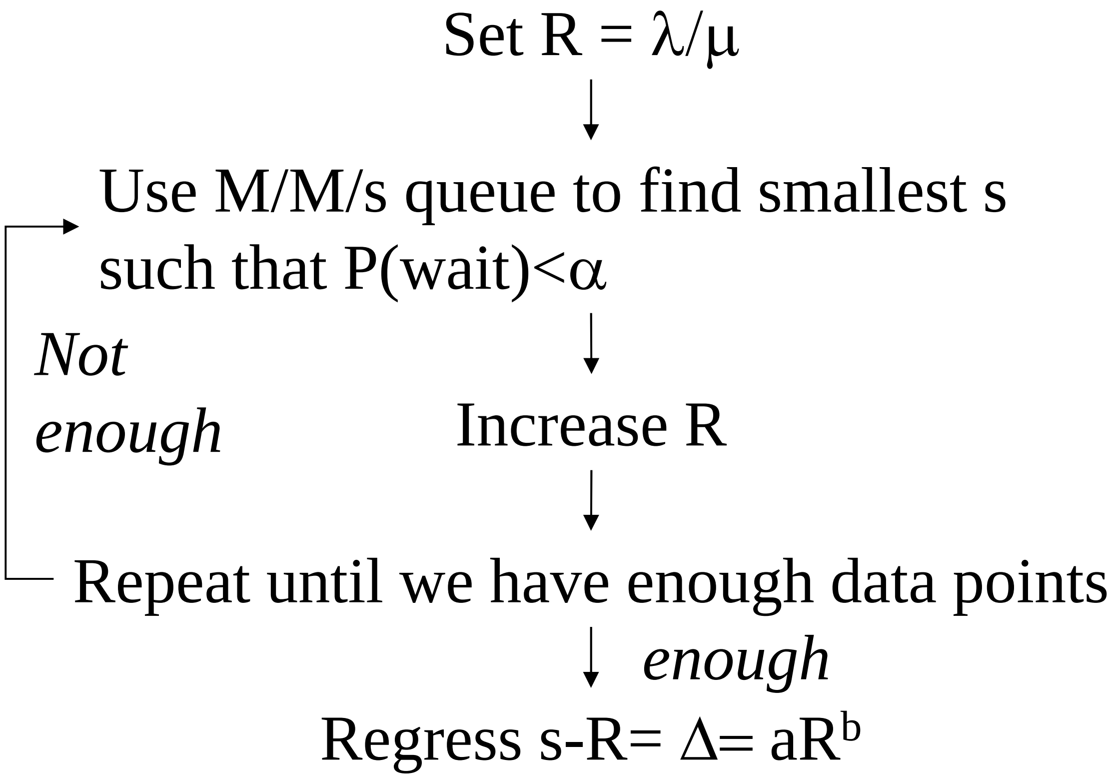
Economies of scale in queueing



Benefits of aggregation or economies of scale in queueing

Square root law Long term planning

Now set service standard [P(wait)] and find # needed



Set $R = \lambda/\mu$

Use M/M/s queue to find smallest s such that $P(\text{wait})<\alpha$

Not enough

Increase R

Repeat until we have enough data points

enough

Regress $s\text{-}R= \Delta= aR^b$

Results (???)

Multiplier	R	Number(0.01)	Number (0.05)	Number (0.2)	Number (0.5)	Delta(0.01)	Delta(0.05)	Delta(0.2)	Delta(.5)	Number(0.01)
1	0.99	5	4	3	2	4	3	2	1	5
2	1.98	7	6	4	3	5	4	2	1	7
3	2.97	9	7	6	4	6	4	3	1	9
4	3.96	10	9	7	6	6	5	3	2	10
5	4.95	12	10	8	7	7	5	3	2	12
10	9.9	19	17	14	12	9	7	4	2	19
15	14.85	26	23	20	17	11	8	5	2	26
20	19.8	32	29	25	23	12	9	5	3	32
25	24.75	38	35	31	28	13	10	6	3	38
30	29.7	44	40	36	33	14	10	6	3	44
35	34.65	50	46	42	38	15	11	7	3	50
40	39.6	56	52	47	43	16	12	7	3	56
45	44.55	62	57	53	49	17	12	8	4	62
50	49.5	68	63	58	54	18	13	8	4	68
60	59.4	79	74	68	64	19	14	8	4	79
70	69.3	91	85	79	74	21	15	9	4	91
80	79.2	102	96	90	84	22	16	10	4	102
90	89.1	113	107	100	95	23	17	10	5	113
100	99	124	118	110	105	25	19	11	6	124
110	108.9	135	128	121	115	26	19	12	6	135
120	118.8	146	139	131	125	27	20	12	6	146

What does all this mean?

Small part of the table

Multiplier	R	Number(0.01)	Number (0.2)	Delta(0.01)	Delta(0.2)
2	1.98	7	4	5	2
10	9.9	19	14	9	4
20	19.8	32	25	12	5
50	49.5	68	58	18	8
100	99	124	110	25	11
120	118.8	146	131	27	12

Extra Number of servers needed so that P(wait)<0.2

Number of servers needed so that P(wait)<0.2

Implications

Multiplier	R	Number(0.01)	Number (0.2)	Delta(0.01)	Delta(0.2)
2	1.98	7	4	5	2
10	9.9	19	14	9	4
20	19.8	32	25	12	5
50	49.5	68	58	18	8
100	99	124	110	25	11
120	118.8	146	131	27	12

Demand increases by a factor of 12, but required number of servers to maintain same service level goes up by a much smaller ratio

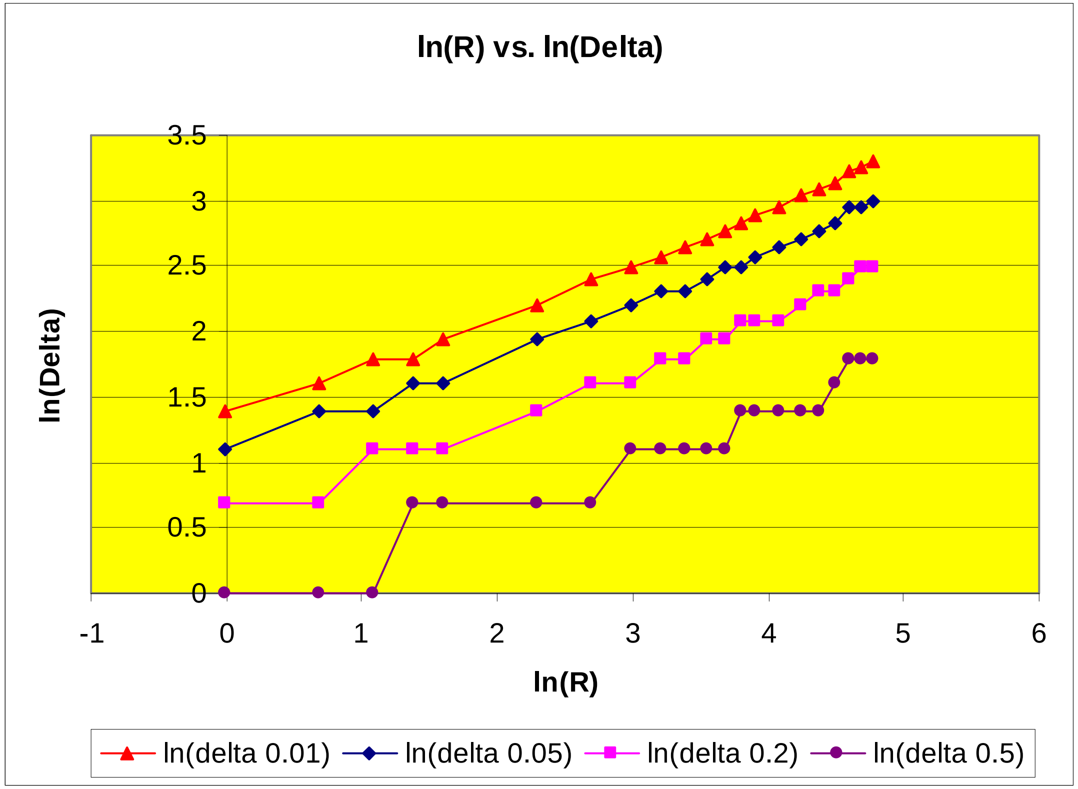
Implications

Multiplier	R	Number(0.01)	Number (0.2)	Delta(0.01)	Delta(0.2)
2	1.98	7	4	5	2
10	9.9	19	14	9	4
20	19.8	32	25	[illegible]	[illegible]
50	49.5	68	58	[illegible]	[illegible]
100	99	124	110	25	11
120	118.8	146	131	27	12

5 more servers needed to improve service with a small system.

With a system 12 times as big, only 15 more servers are needed.

Implications



Ln(Δ=# extra)

is linear in

ln(R=offered demand)

particularly for high service levels

Estimation of square root law

- Fix α=P(wait)
- Find # of servers needed to ensure P(wait)$\leq\alpha$ for a range of R using M/M/N queue
- Compute *ln(R)* and *ln(Δ)*
- Regress *ln(Δ)* =a+b *ln(R)*

α=0.01

R	Min #	Number needed for α=0.01	Delta	ln(R)	ln(delta)
0.99	1	5	4	-0.010	1.386
4.95	5	12	7	1.599	1.946
9.9	10	19	9	2.293	2.197
19.8	20	32	12	2.986	2.485
29.7	30	44	14	3.391	2.639
49.5	50	68	18	3.902	2.890

Regression results

Regression Statistics		
Multiple R	0.9976	
R Square	0.9951	
Adjusted R Square	0.9948	
Standard Error	0.0208	
Observations	15	
ANOVA		
	df	*SS*
Regression	1	1.1549
Residual	13	0.0057
Total	14	1.1606
	Coefficients	*Standard Error*
Intercept	1.1463	0.0342
X Variable 1	0.4462	0.0087

$$\begin{aligned}\Delta &= e^{1.146+0.446\ln(R)} \\ &= e^{1.146}R^{0.446} \\ &= 3.147R^{0.446} \\ &\approx 3.147\sqrt{R}\end{aligned}$$

$$\beta = 3.147$$

$$\boxed{\begin{aligned}N &= R + 3.147R^{0.446} \\ &\approx R + 3.147\sqrt{R}\end{aligned}}$$

Example with $\alpha=0.01$; $\mu=12/\text{hr}$

R	Number of servers	Delta	rho
150	180	30	0.833
175	207	32	0.845
200	234	34	0.855
225	261	36	0.862
250	287	37	0.871
275	314	39	0.876
300	341	41	0.880
600	655	55	0.916

Demand quadruples, but number of excess servers less than doubles and utilization of servers increases

Large service centers are more efficient

Long term planning and the number of CSRs to train

β	monthly turnover rate
τ	time to train an employee (months)
y_t	number of employees on hand at time t (now)
y_j	number of employees who finish training at beginning of month j
η_t	number of employees needed in month t
z_t	number to hire as trainees in month t

$$\boxed{z_t = \max\left(0, \eta_{t+\tau} - \sum_{j=t}^{t+\tau-1} y_j (1-\beta)^{\tau-(j-t)}\right)}$$

Example: Inputs for $\tau=3; \beta=0.1$

j	Y(j)	Needed(j)	Train	Available
1	280			280.0
2	30			
3	25			
4		300		
5		320		
6		310		
7		340		
8		350		
9		360		
10		320		
11		310		
12		295		
13		280		

Number available now

Number of trainees to start in next 2 months

Number needed in future periods

Example: Inputs for $\tau=3$; $\beta=0.1$

j	Y(j)	Needed(j)	Train	Available
1	280			280.0
2	30			282.0
3	25			
4		300		
5		320		
6		310		
7		340		
8		350		
9		360		
10		320		
11		310		
12		295		
13		280		

$$\begin{aligned} & 280*(1-0.1)+30 \\ = \ & 280*(0.9)+30 \\ = \ & 252+30 \\ = \ & 282 \end{aligned}$$

Example: Inputs for τ=3; β=0.1

j	Y(j)	Needed(j)	Train	Available
1	280		50	280.0
2	30			282.0
3	25			
4		300		
5		320		
6		310		
7		340		
8		350		
9		360		
10		320		
11		310		
12		295		
13		280		

Max{0,
300
$-280*(0,9^3)$
$-30*(0.9^2)$
-25*(0.9)}
=max{0,
300
-204.12
-24.3
-22.5}
=max{0,300-250.92}
=50 (round up)

Example: Inputs for $\tau=3$; $\beta=0.1$

j	Y(j)	Needed(j)	Train	Available
1	280		50	280.0
2	30			282.0
3	25			
4	50	300		
5		320		
6		310		
7		340		
8		350		
9		360		
10		320		
11		310		
12		295		
13		280		

Example; τ=3; β=0.1

j	Y(j)	Needed(j)	Train	Available
1	280		50	280.0
2	30		50	282.0
3	25		22	278.8
4	50	300	61	300.9
5	50	320	44	320.8
6	22	310	45	310.7
7	61	340	0	340.7
8	44	350	18	350.6
9	45	360	16	360.5
10	0	320	15	324.5
11	18	310		310.0
12	16	295		295.0
13	15	280		280.5

Significant drop in needed results in (1) no training and (2) available>needed

Number available now

Number of trainees to start in next 2 months

$280(1-0.1)+30$

Max{0,300-
$280(0.9)^3$-
$30(0.9)^2$-
$25(0.9)^{)1}$}

Would have been negative without max operator

Key Reference

Daskin, Mark, lecture notes, University of Michigan.

Gans, N., G. Koole, and A. Mandelbaum, 2003, "Telephone Call Centers: Tutorial, Review, and Research Prospectus," Manufacturing and Service Operations Management, **5**:2, 79-141.