Modeling the Customer Role in Services: Examples from Call Centers

Zeynep Akşin, Koç University
B. Ata, C. Su, University of Chicago
S. Emadi, University of North Carolina
O. Jouini, Y. Dallery, Ecole Centrale Paris
E. Güneş, L. Örmeci, Koç University
H. Özden, Amazon

Services ..

- … are settings where the process is the product
- …are provided to customers and cannot be produced independently of them
- …are produced, distributed and consumed simultaneously
- …mismatch between demand and capacity is inherent

Traditional queuing models

Passive and mostly homogeneous

Traditional queuing models

versus queues where customers react

Passive and mostly homogeneous

Traditional queuing models versus queues where customers react

Passive and mostly homogeneous

Strategic
Makes choices
Reacts to system features
Can be heterogenous

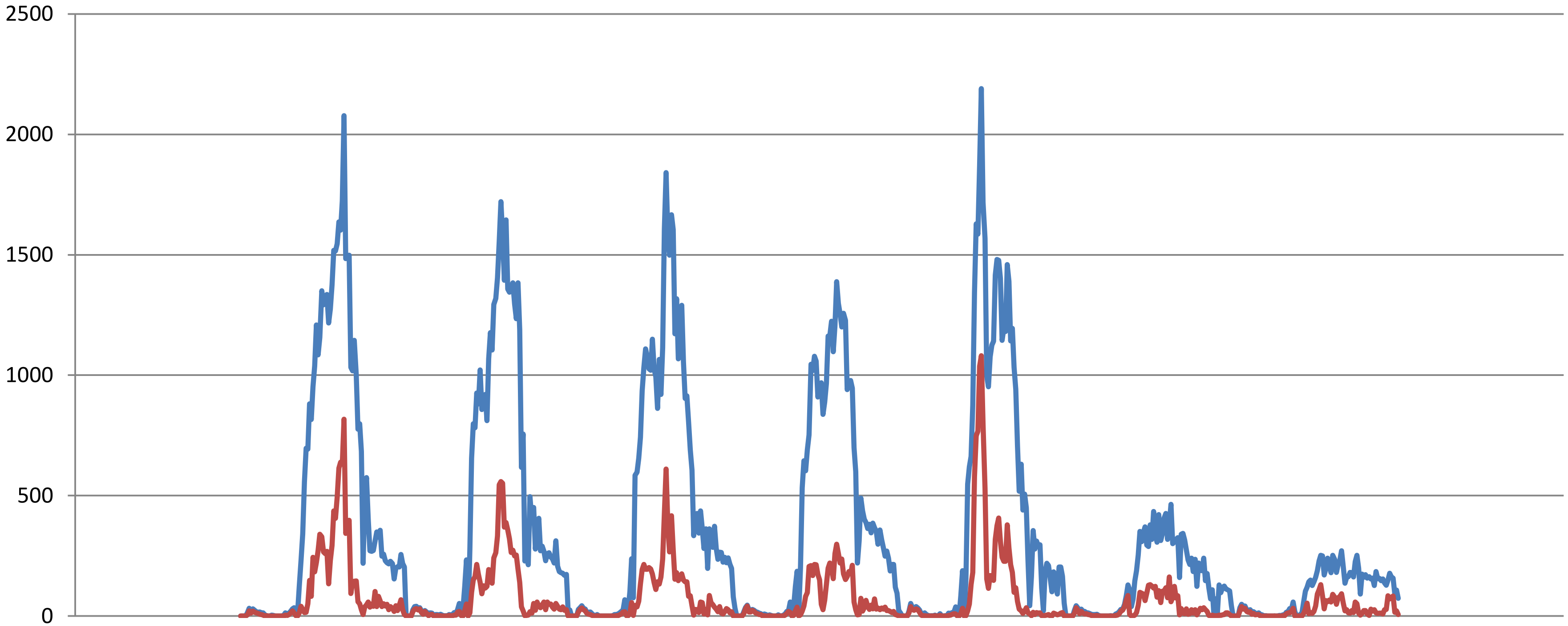
Overview

- Customer reactions to waiting
  - Patience in queues
  - Delay announcement practices aimed to control patience
- Customer reactions to sales attemps
  - Customer purchase
  - Cross-selling practices aimed to control customer purchase

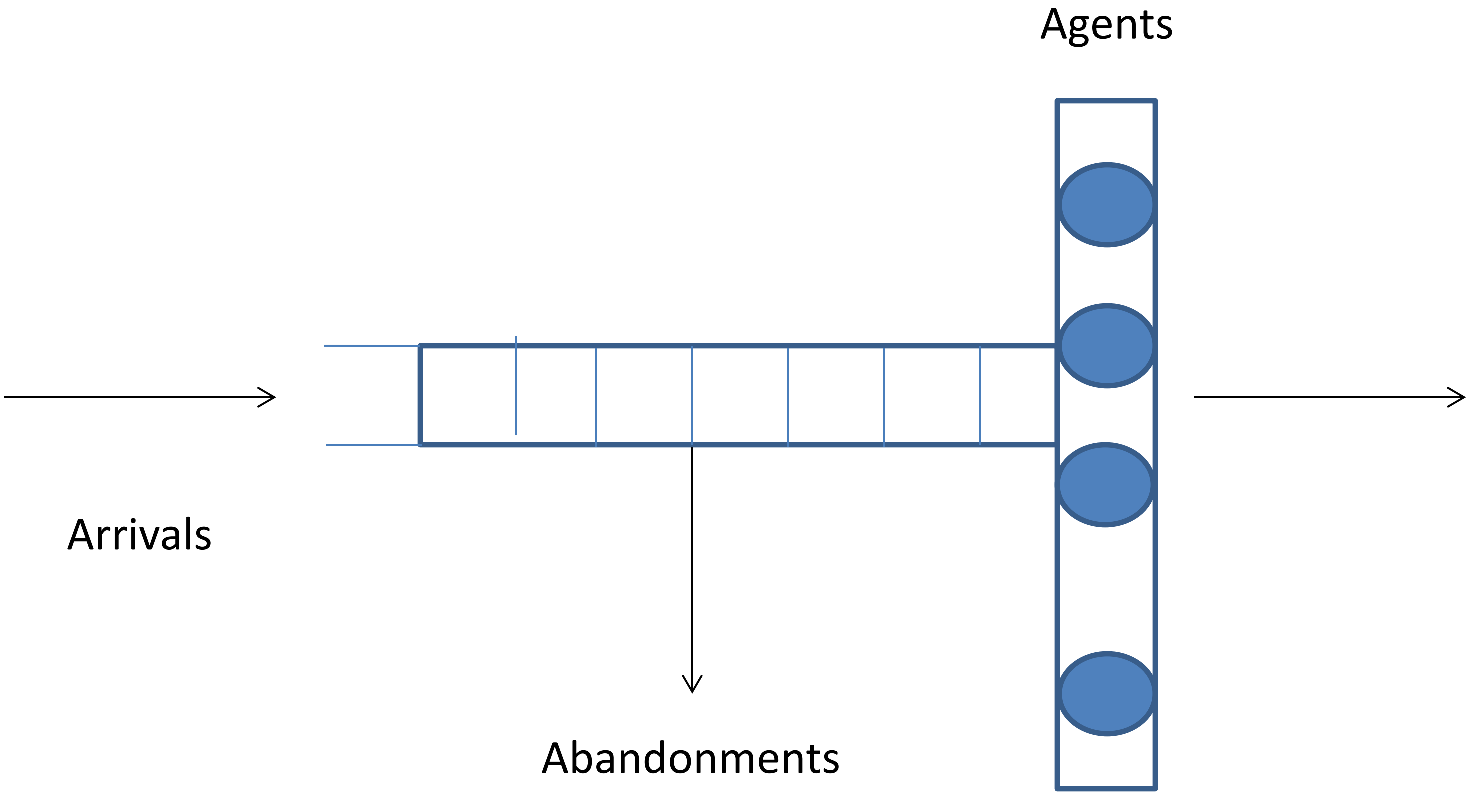
Understanding and Controlling Customer Patience

Setting: Call Centers

A bank call center: daily number of calls and abandonments



M/M/N+M: Erlang-A Model



Phone queues with abandonments / reneging

- Baccelli and Hebuterne (1981)
- Brandt and Brandt (1997, 1999)
- Akşin and Harker (1999)
- Garnett, Mandelbaum, Reiman (2002)

Understanding abandonments correctly makes a difference!

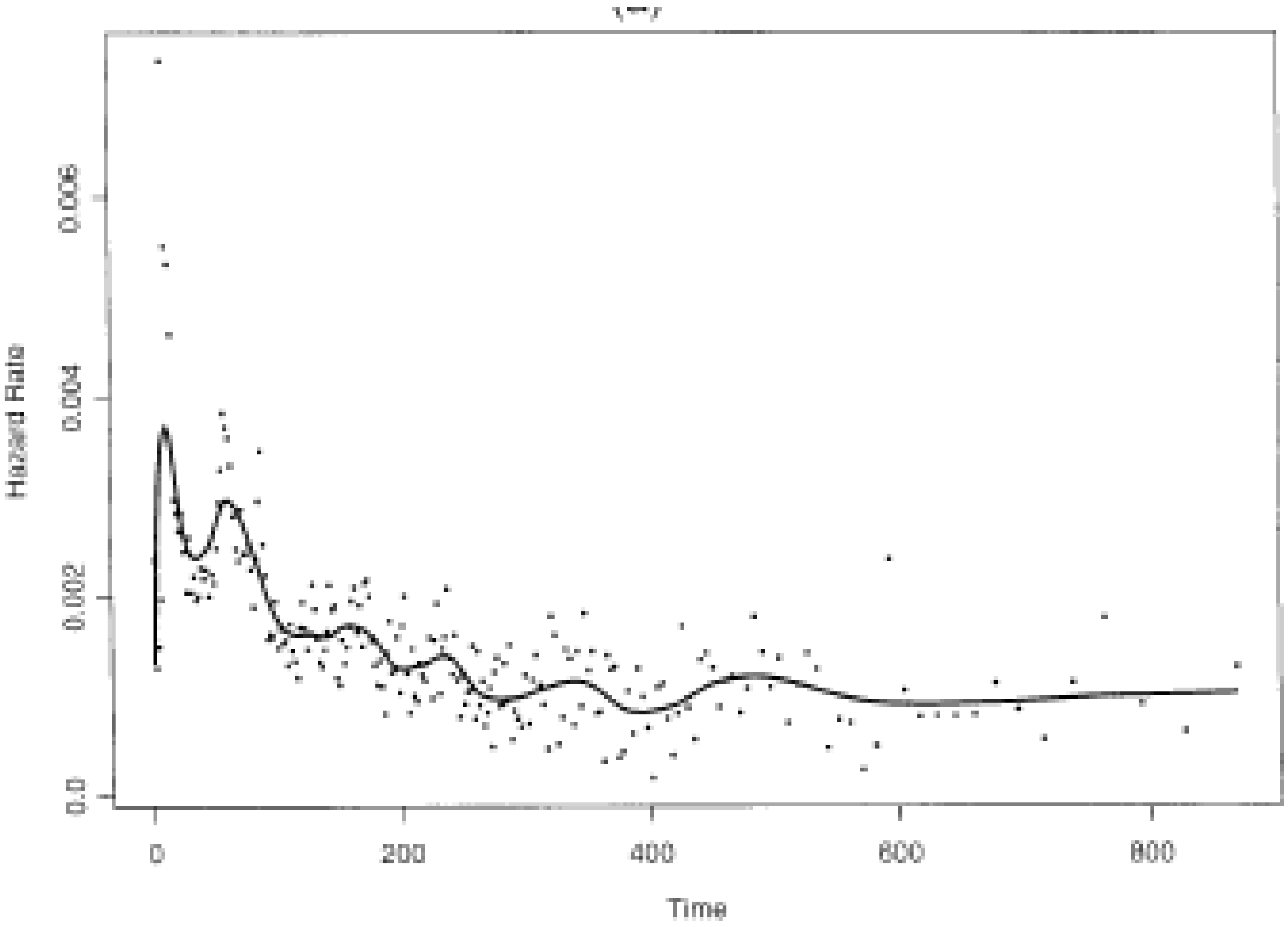
From Table 2 Garnett, Mandelbaum and Reiman, 2002

	M/M/N	*M/M/N +M*
Fraction abandoning	—	3.1%
Average speed of answer	20.8 sec.	3.6 sec.
Waiting time’s 90th percentile	58.1 sec.	12.5 sec.
Average queue percentile	17	3
Agents’ utilization	96%	93%

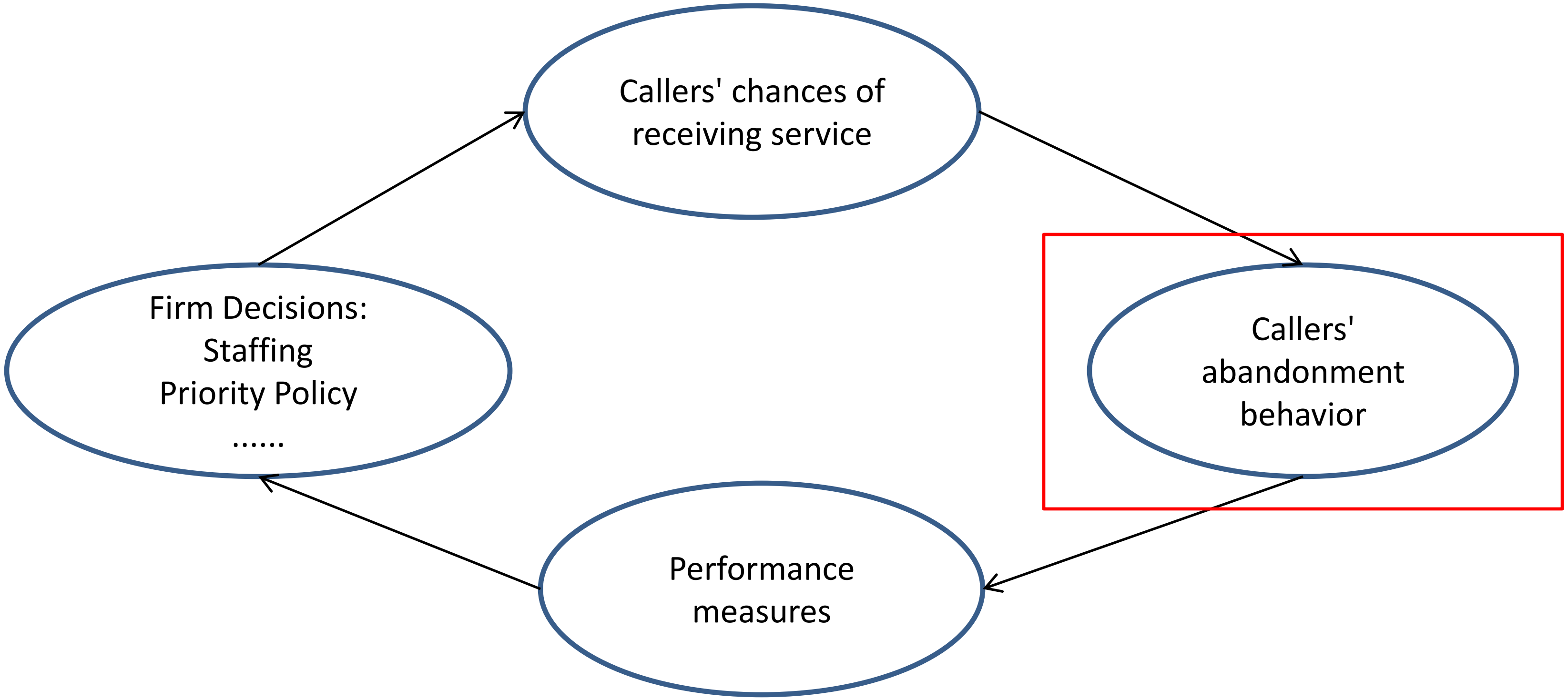
Patience modeling

- Assume a patience time distribution
- Estimate distribution parameter(s) using data on observed abandonments
- Issues:
  - Data is heavily censored
  - Call center data is aggregated in 15 min intervals
  - Observed patience for a **given** system

First documented use of individual call data Brown et al. (2005)- hazard rate



Abandonment is not independent from queue design



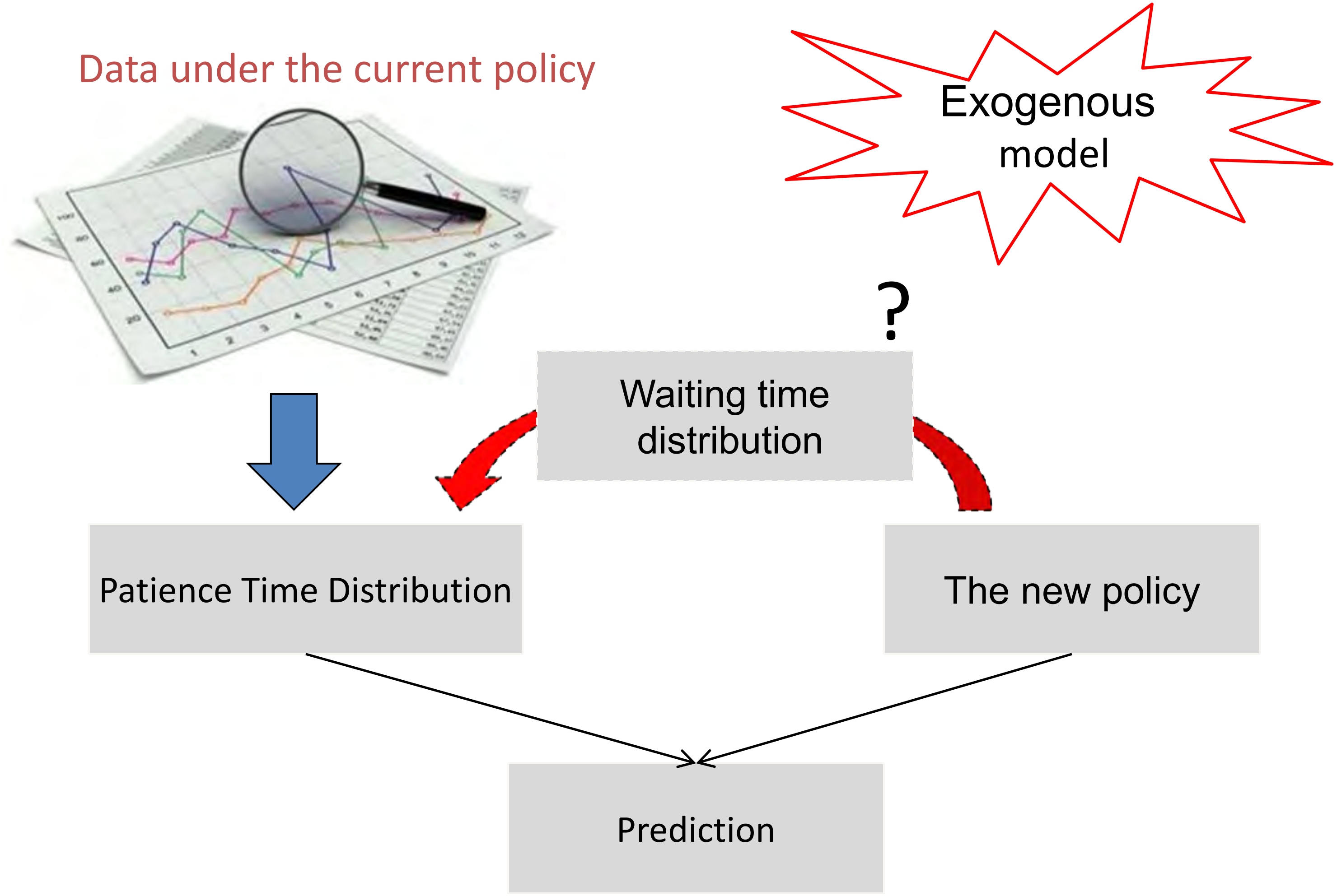
Callers' chances of receiving service

Callers' abandonment behavior

Performance measures

Firm Decisions:
Staffing
Priority Policy
……

The traditional approach to treating abandonments



Data under the current policy

Exogenous model

?

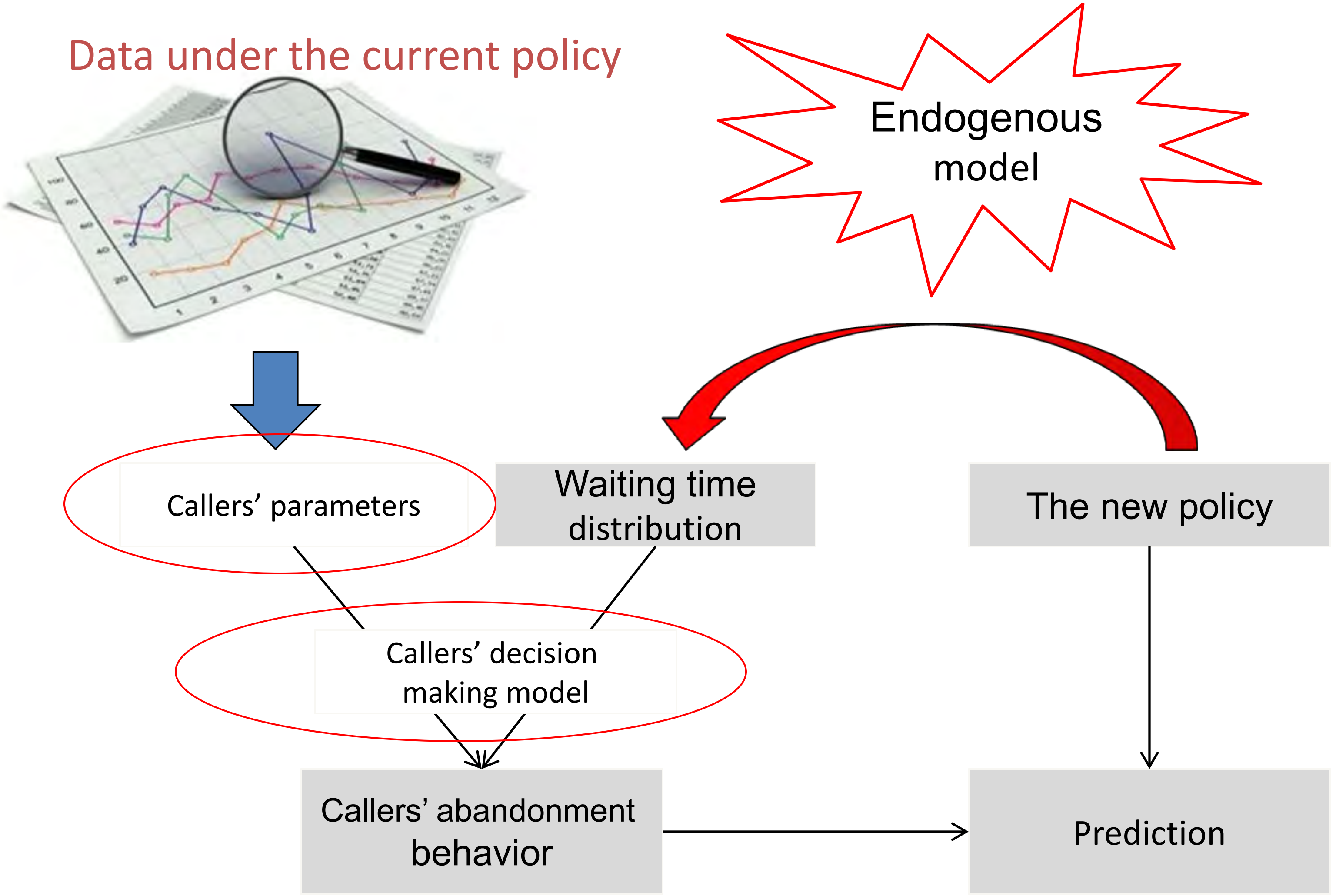
Waiting time distribution

Patience Time Distribution

The new policy

Prediction

The new approach which enables one to find the impact of policy changes



Data under the current policy

Endogenous model

Callers’ parameters

Waiting time distribution

The new policy

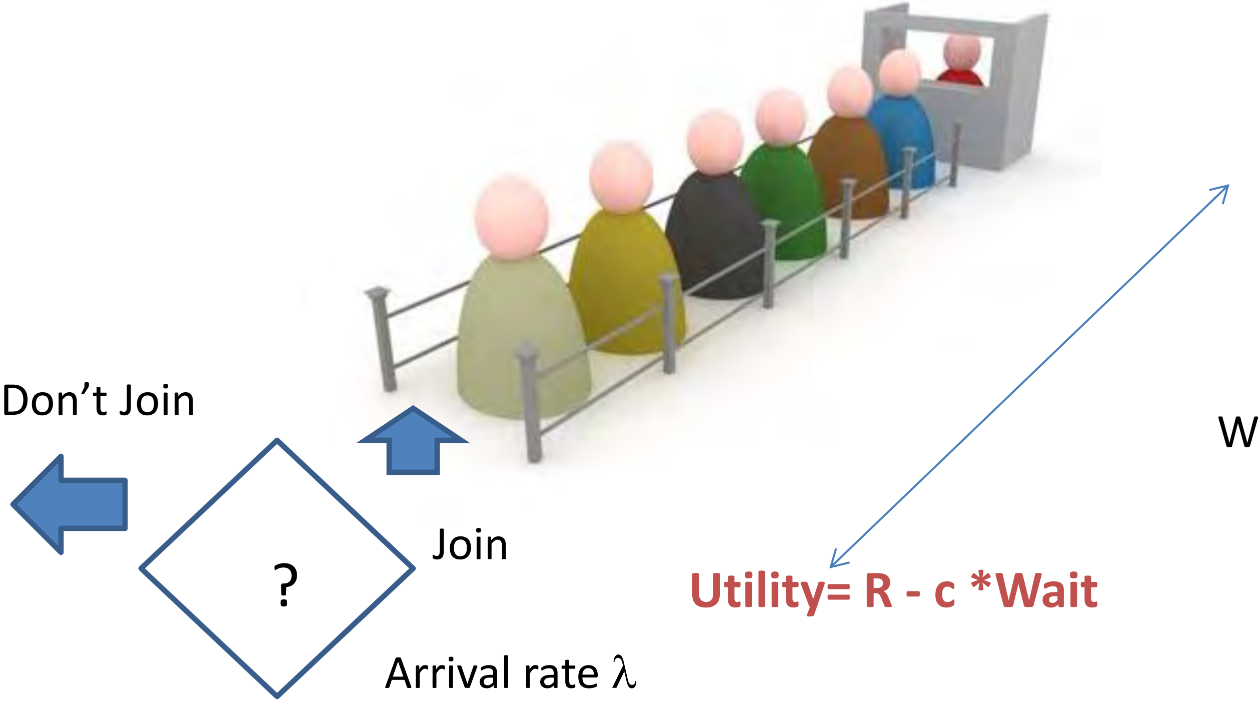
Callers’ decision making model

Callers’ abandonment behavior

Prediction

Aksin, Ata, Emadi and Su (2013)

Decision Making Rational Customers in Observable Queues (Naor, 1969)



Service has a value R

Service rate μ

Don't Join

Waiting has a cost c

Join

Utility= R - c *Wait

Arrival rate λ

Rational Customers in Queues

- Huge literature-Hassin and Haviv (2003) To queue or not to queue
    - Seminal paper by Naor (1969)
    - Observable or unobservable
    - Homogenous vs heterogeneous customers
    - Linear versus non-linear waiting costs
    - Etc.
- Rational abandonments
    - Hassin (1995)- reward drops to zero after a threshold wait
    - Mandelbaum and Shimkin (2000)- a queue with a fault state
    - Shimkin and Mandelbaum (2004)-non-linear waiting costs
    - Etc.

The typical OR/OM framework

ASSUME

R, c

Cu, Co

FORMULATE

Maximize Utility(R, c)

Maximize E(Profit(Cu, Co))

ANALYZE

Rational Join, Balk, Wait, Abandon decisions

Optimal order quantity decisions

Structural Estimation

OBSERVE FROM DATA

Join, Balk, Wait, Abandon decisions

Order quantity decisions

FORMULATE A DECISION MAKING MODEL

Maximize Utility(R, c, others)

Maximize E(Profit(Cu, Co, others))

INFER **R, c**

Cu, Co

- Berry (1994), Nevo (2000), Rust (1987), Nair (2007).

Dynamic model for Callers' decision making process

Callers are forward looking and make decision dynamically.

Callers' Parameters

- r_i : Reward from receiving service
- c_i : Cost incurred by waiting for a unit of time

- Heterogeneity of the callers:

$$\begin{aligned} r_i &= \exp(m_r + \sigma_r y_{1i}) \\ c_i &= \exp(m_c + \sigma_c y_{2i}) \end{aligned}$$

y_{1i} and y_{2i} have independent normal distributions.

- The structural parameters to be estimated:

$$\Theta = (m_r, m_c, \sigma_r, \sigma_c)$$

Callers’ Actions and Utilities

- $u(t, r_i, c_i, \varepsilon_{it}, d_{it}) = v(t, r_i, c_i, d_{it}) + \varepsilon_{it}(d_{it})$

$$d_{it} = \begin{cases} 1 & \text{Abandoning} \\ 0 & \text{Waiting} \end{cases}$$

$\varepsilon_{it}(d_{it})$ random shock to the utility with type-I extreme value distribution.

Callers’ Nominal Utilities

- Nominal utility from abandoning:

$$v(t, r_i, c_i, 1) = 0$$

- Nominal utility from waiting:

$$v(t, r_i, c_i, 0) = -c_i + \pi(t) r_i + (1 - \pi(t))\mathbb{E}\left[\underbrace{\max_{d\in\{0,1\}} u(t+1, r_i, c_i, \varepsilon_{i(t+1)}(d), d)}_{V(t, r_i, c_i)} \right]$$

Probability of receiving service

the integrated value function

The model

- The optimal action of a caller:

$$d_{it} = \arg \max_{d \in \{0,1\}} u(t, r_i, c_i, \varepsilon_{it}, d)$$

- Probability of abandoning:

$$P_{it}(1; r_i, c_i) = \frac{1}{1 + \exp(-c_i + \pi(t) r_i + (1 - \pi(t)) V(t, r_i, c_i))}$$

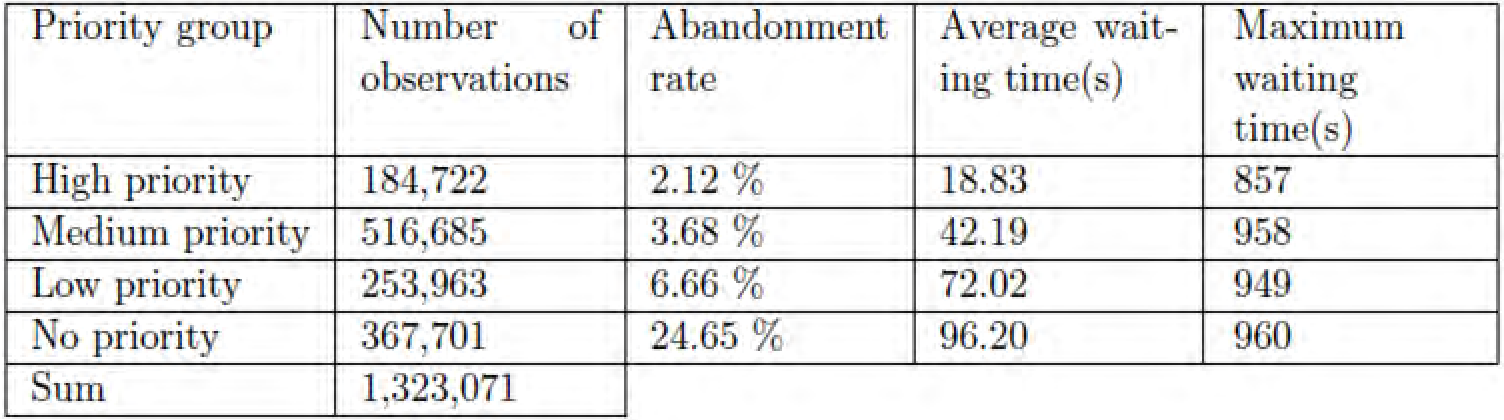
- The integrated value function:

$$V(t, r_i, c_i) = \log \Big(1 + \exp(-c_i + \pi(t+1) r_i + (1 - \pi(t+1)) V(t+1, r_i, c_i)) \Big)$$

$$V(T, r_i, c_i) = 0$$

Data

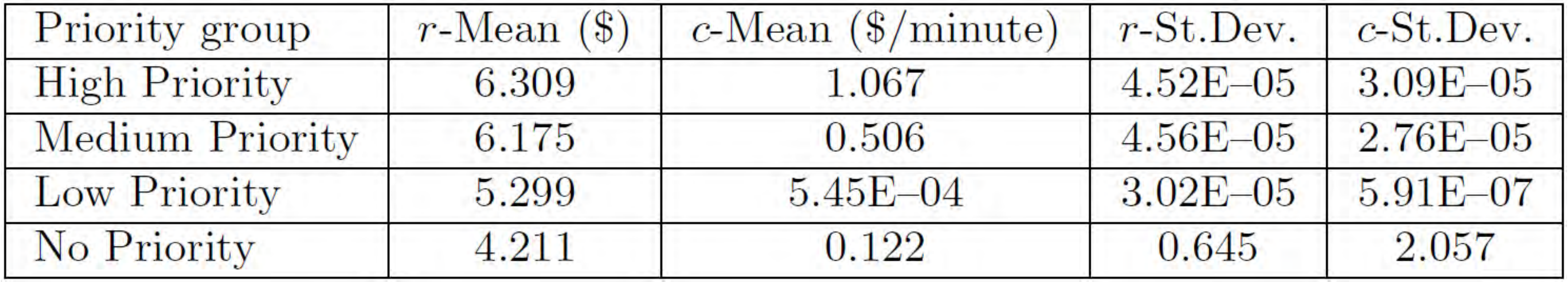
- The focus of our analysis is on calls
  - In the Retail service group.
  - Received during the working days during weeks without holidays between 9 a.m. and 2 p.m.
  - Entered the system through VRU and proceeded to wait in the queue
  - Having normal termination, transfer or abandonment as an outcome.
  - Their maximum waiting time is less than 960 seconds.



Priority group	Number of observations	Abandonment rate	Average waiting time(s)	Maximum waiting time(s)
High priority	184,722	2.12 %	18.83	857
Medium priority	516,685	3.68 %	42.19	958
Low priority	253,963	6.66 %	72.02	949
No priority	367,701	24.65 %	96.20	960
Sum	1,323,071			

Estimation

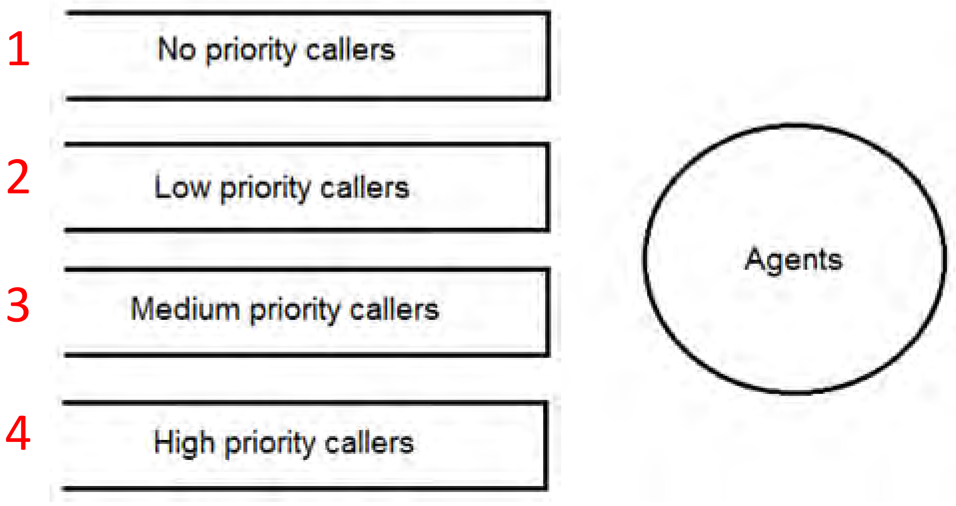
- Kaplan Meier to estimate the probability of service from the data
- MLE to estimate model primitives



Priority group	r-Mean (\$)	c-Mean (\$/minute)	r-St.Dev.	c-St.Dev.
High Priority	6.309	1.067	4.52E–05	3.09E–05
Medium Priority	6.175	0.506	4.56E–05	2.76E–05
Low Priority	5.299	5.45E–04	3.02E–05	5.91E–07
No Priority	4.211	0.122	0.645	2.057

Counterfactual illustrating the importance of modeling patience endogenously

Reversed strict priority policy


1 No priority callers

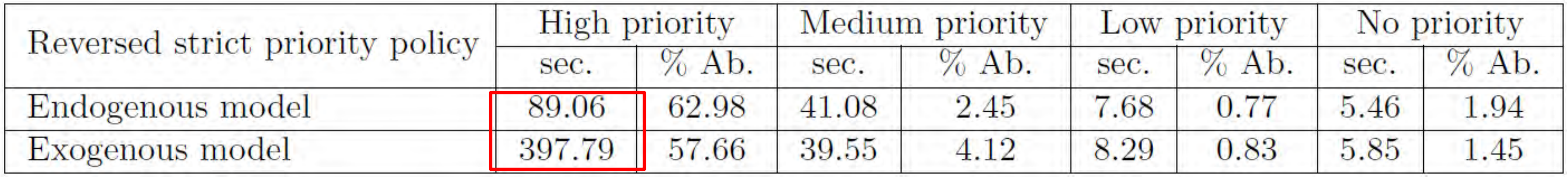
2 Low priority callers

3 Medium priority callers

4 High priority callers

Agents

Reversed strict priority policy



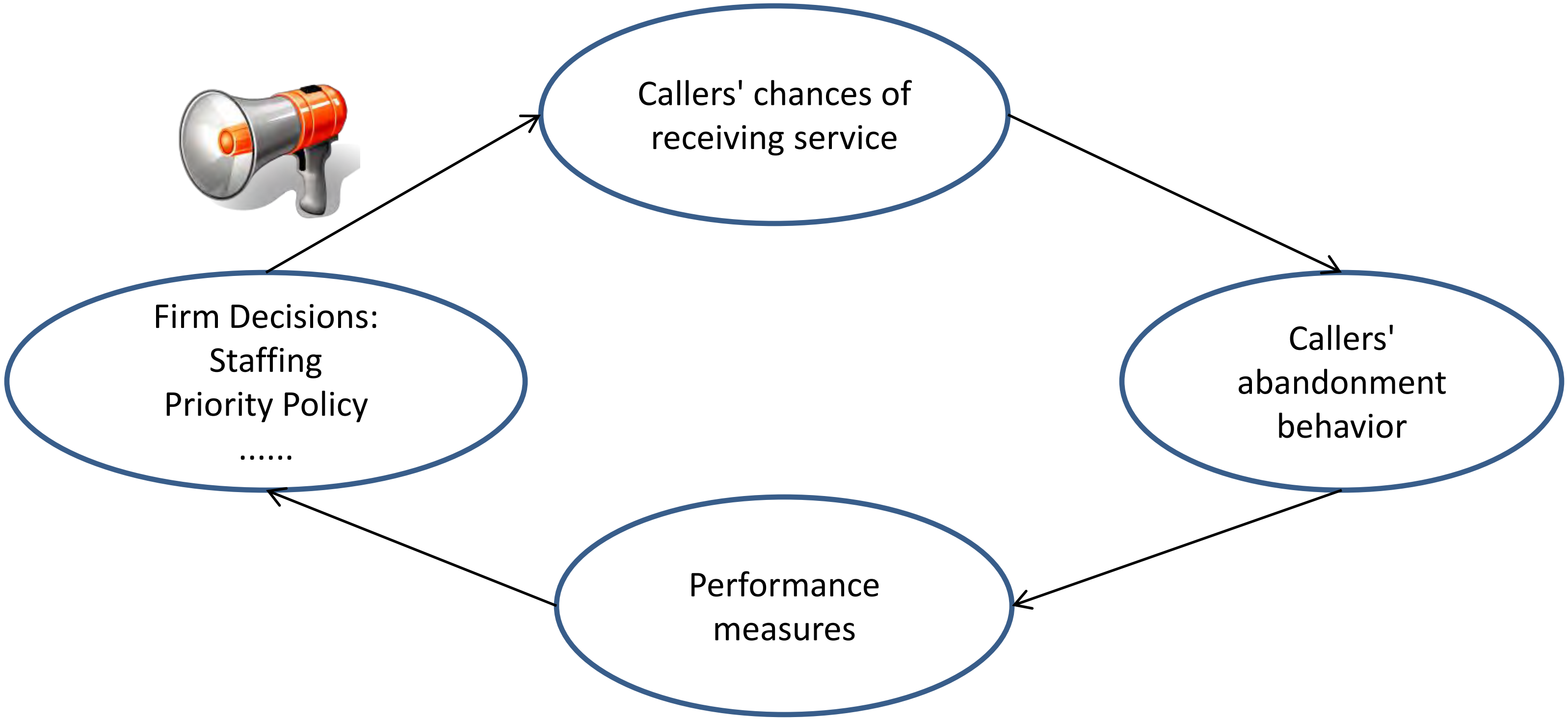
Reversed strict priority policy	High priority		Medium priority		Low priority		No priority	
	sec.	% Ab.	sec.	% Ab.	sec.	% Ab.	sec.	% Ab.
Endogenous model	89.06	62.98	41.08	2.45	7.68	0.77	5.46	1.94
Exogenous model	397.79	57.66	39.55	4.12	8.29	0.83	5.85	1.45

- Exogenous model overestimates the waiting time for the high priority group.
- In the endogenous model: for callers with large waiting cost,

  service quality ↓ → abandonment probability ↑
- In the exogenous model: for callers with large waiting cost,

  service quality ↓ → abandonment probability (no change)

Patience Control: Delay announcement with Customer Reaction



Callers' chances of receiving service

Callers' abandonment behavior

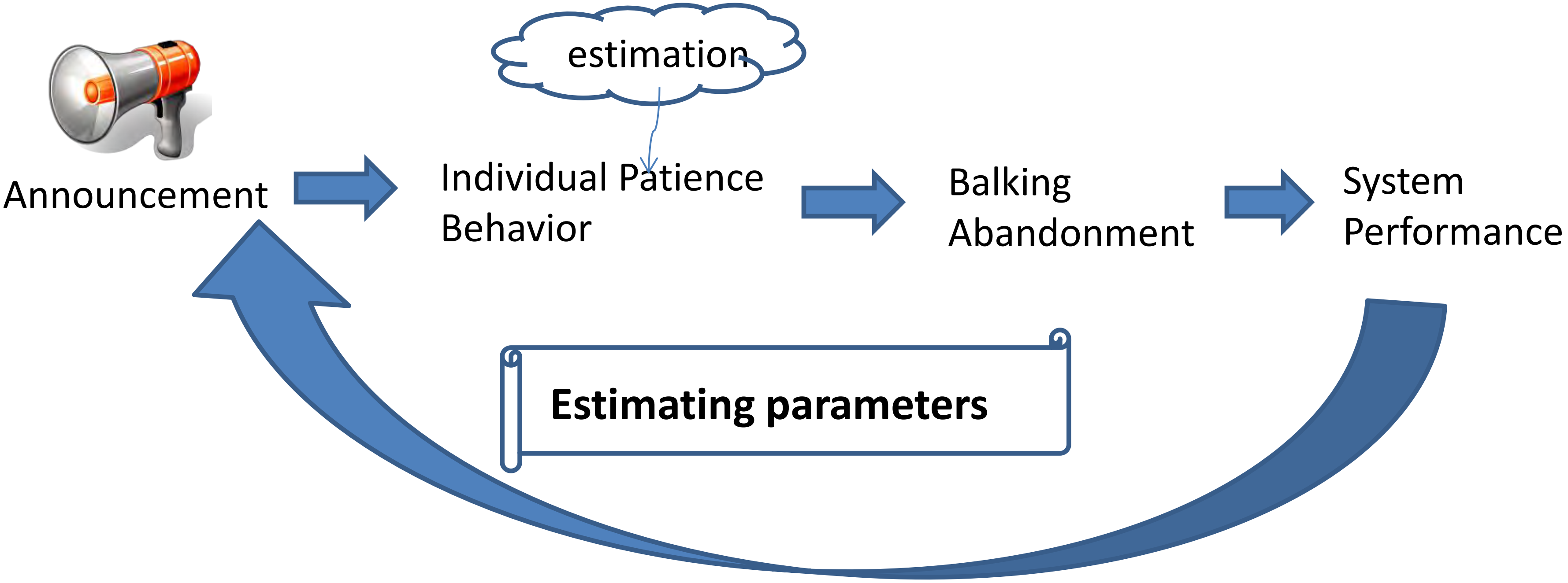
Performance measures

Firm Decisions:
Staffing
Priority Policy
......

Delay announcement problem challenges

How to model individual behavior?

Performing transient queueing analysis when individual level model is embedded in a queue w/ abandonment



estimation

Announcement

Individual Patience Behavior

Balking Abandonment

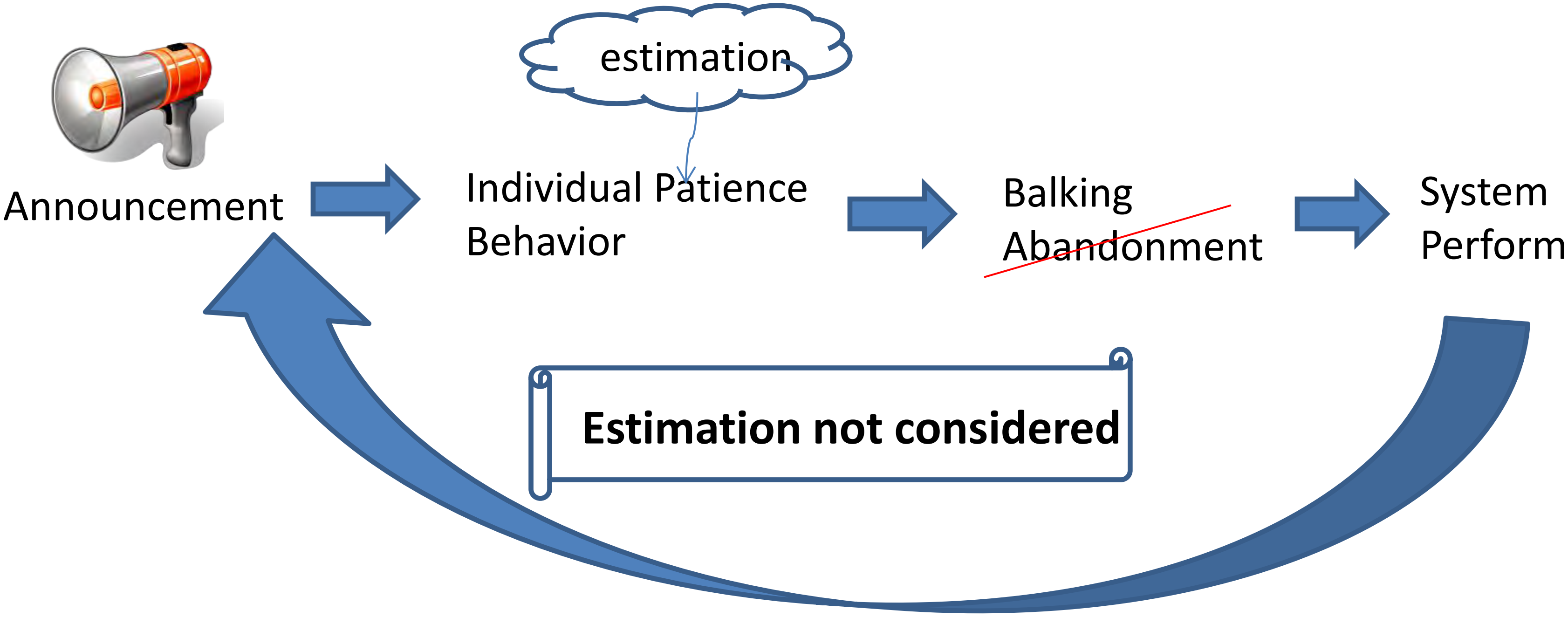
System Performance

Estimating parameters

Early solution by Whitt (1999)

Assume random patience threshold

Simplify by ignoring abandonment reactions; threshold only affects balking decisions



estimation

Announcement

Individual Patience Behavior

Balking ~~Abandonment~~

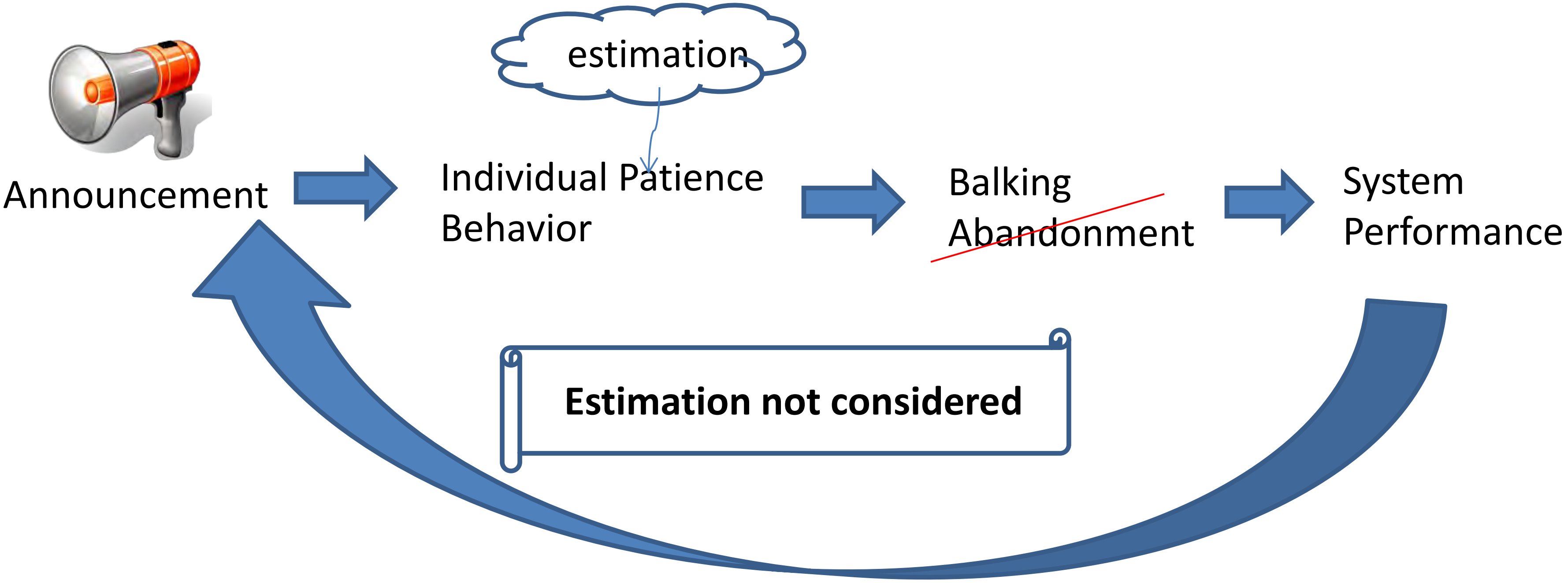
System Performance

Estimation not considered

Guo and Zipkin (2007)

Assume utility maximizing customers

Simplify by ignoring abandonment reactions



estimation

Announcement

Individual Patience Behavior

Balking ~~Abandonment~~

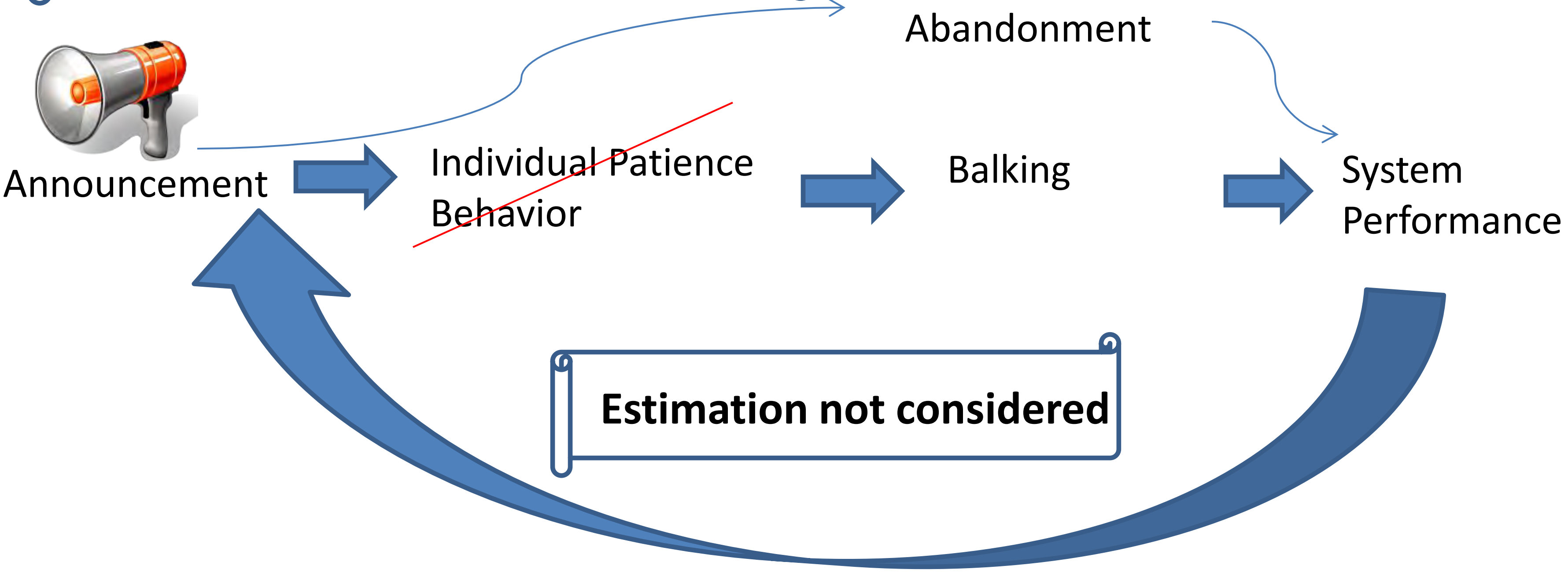
System Performance

Estimation not considered

Armony, Shimkin and Whitt (2009):
Include customer reactions in the form of abandonments

Assume an exogenously specified different abandonment rate

Simplify through heavy traffic analysis



Abandonment

Announcement

~~Individual Patience Behavior~~

Balking

System Performance

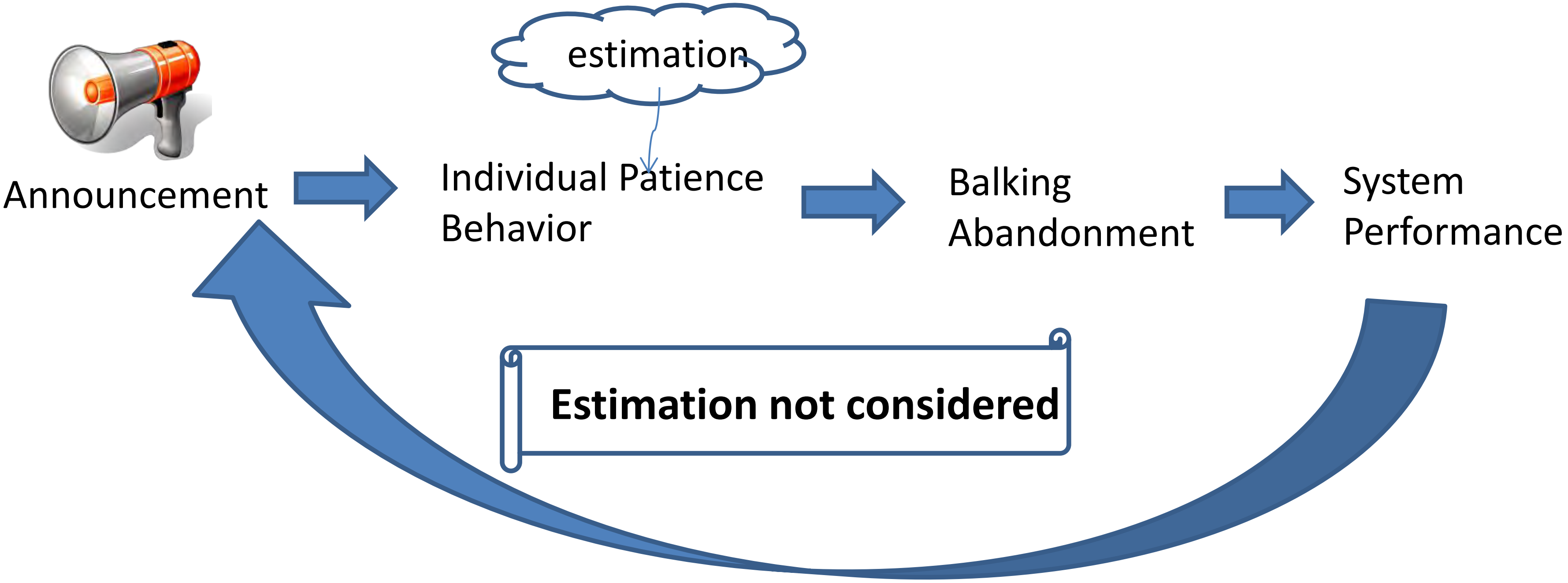
Estimation not considered

No modeling of individual behavior

Jouini, Aksin, Dallery (2011): individual customer patience reaction is modeled

Assume random patience threshold adjusts after announcement

Simplify by assuming aggregate abandonment remains exponential



estimation

Announcement

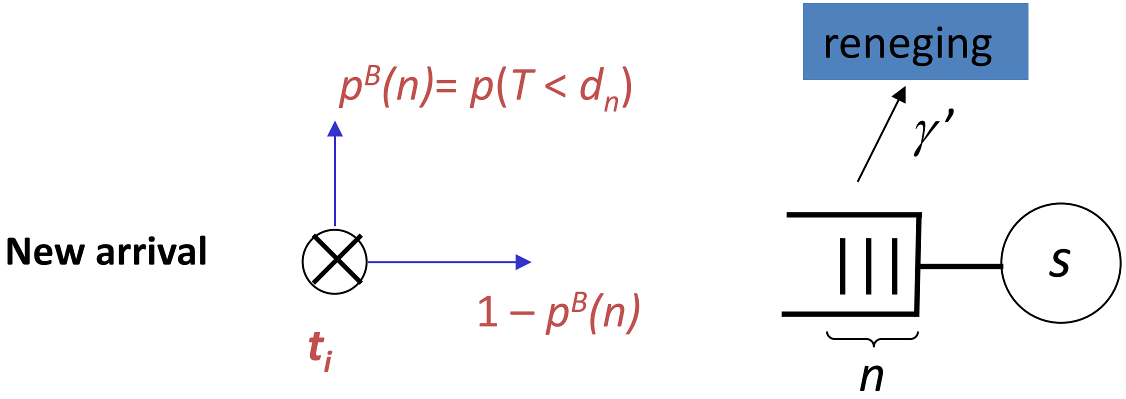
Individual Patience Behavior

Balking Abandonment

System Performance

Estimation not considered

Modeling reneging behavior



New arrival

$p^B(n) = p(T < d_n)$

$1 - p^B(n)$

t_i

reneging

γ'

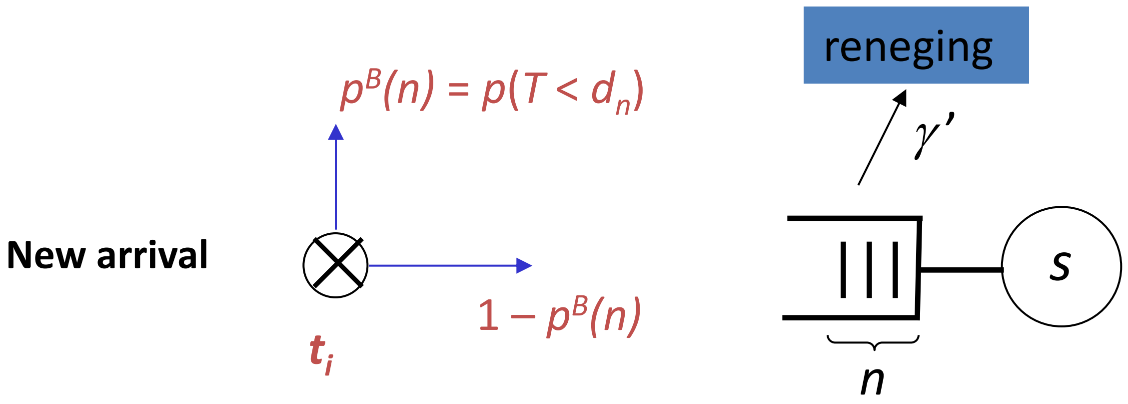
s

n

$$d_n = F_{Xn}^{-1}(\beta)$$

patience threshold= $\theta t_i + (1 - \theta)\, d_n$

New reneging behavior: analysis



$$d_n = F_{Xn}^{-1}(\beta)$$

- $r_n = r_n(\beta)$, probability that a customer who elects to wait initially will renege

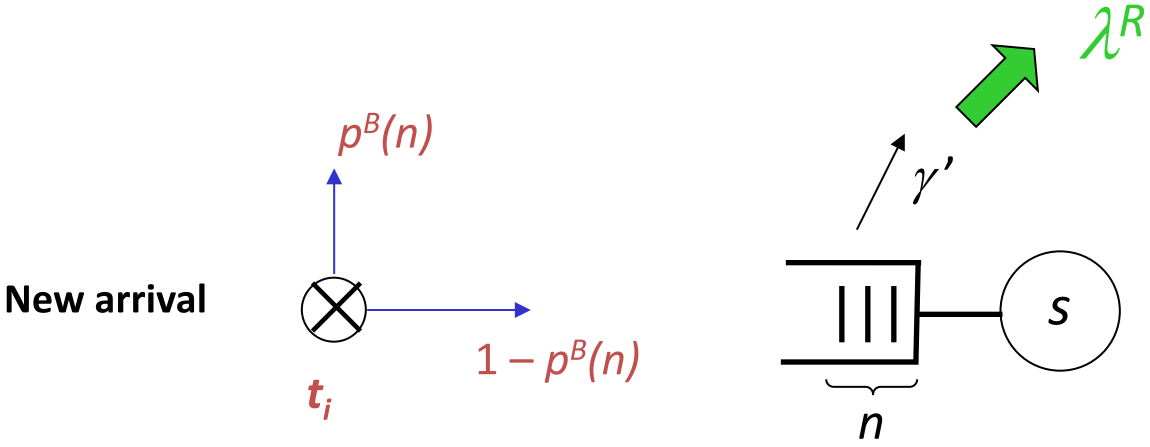
- r_n is the conditional probability that the realization of his waiting time D_n exceeds his patience threshold, given that he elects to join the queue

$$r_n = P(D_n > \text{patience threshold} \mid T > d_n)$$

$$\text{patience threshold} = \theta t_i + (1 - \theta)\, d_n$$

New reneging behavior: analysis

- We denote by λ^R the stationary flow of abandoning customers

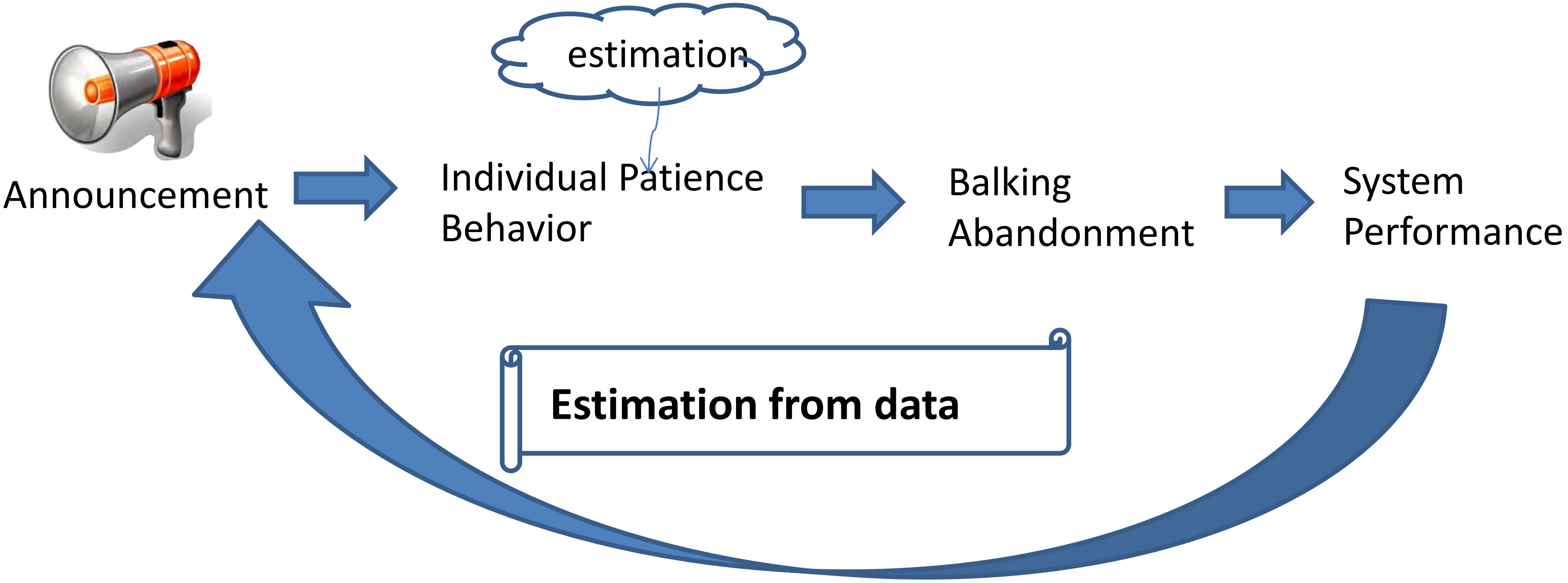


- On the one hand, we have due to exponential reneging: $\lambda^R = \gamma' L_q$
- On the other hand, we have using PASTA: $\lambda^R = \sum_{n=0}^{\infty} \lambda(1-\alpha_0)(1-p^B(n))p(s+n)r_n$
- All quantities are functions of γ'. Then, γ' is a point mapped to itself by a given continuous function
- Using a fixed-point algoritm we can compute γ'

Aksin, Ata, Emadi, Su (2014):estimation, reaction, queueing combined

Utility maximization: optimal stopping problem under delay announcement

Simplify by making Markovian approximations



estimation

Announcement

Individual Patience Behavior

Balking Abandonment

System Performance

Estimation from data

Understanding and Controlling Customer Relationships

Setting: Cross-selling in bank call centers

The motivating idea

- Customers have relationships with service firms
- Relationship dynamics can be modeled as Markov chains
- Firm actions may influence customer behavior
- Changes in customer behavior may in turn affect firm decisions
- Idea: modeling firm decisions in the presence of customer reactions

Related Literature

Descriptive models of customer relationship

- Schmittlein, Morrison and Colombo, 1987
- Schmittlein and Peterson, 1994
- Netzer, 2004

Optimizing customer equity

- Ho, Park and Zhou, 2005
- Rust, Lemon and Zeithaml, 2004
- Venkatesan and Kumar, 2004
- Ching, Wong, Altman, 2004
- Sun and Li, 2005
- Sun, Li and Zhou, 2006

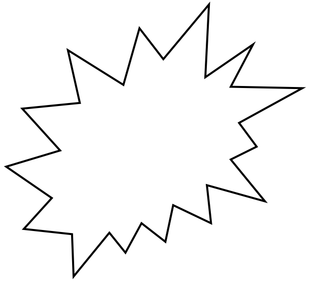
Markovian Life Time and Arrival Process

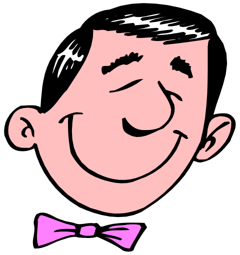
Endogenity problem (Rust and Chung, 2006)

Customer Reactions to Cross-selling

Customer reaction

Try to Cross-sell?

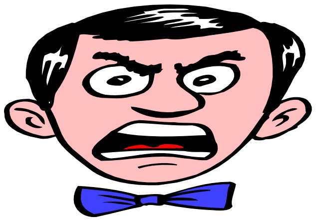




“This new loan option is exactly what I need!”

+ $$€€

- (+) Retention: Marple and Zimmerman, 1999
- (+) Reduce churn: Kamakura et al. 2003



“I don’t want another sales pitch, just transfer the money! ”

Lost time, annoyance

- (-) Switch: Kamakura et al. 2003

Cross selling in the literature

When to cross-sell,
Which product?-relationship

- Kamakura et al. (1991, 2003)
- Paas and Kuijlien (2001)
- Li, Sun and Wilcox (2005)

Sequential offers

When to cross-sell,
Which product?-system congestion

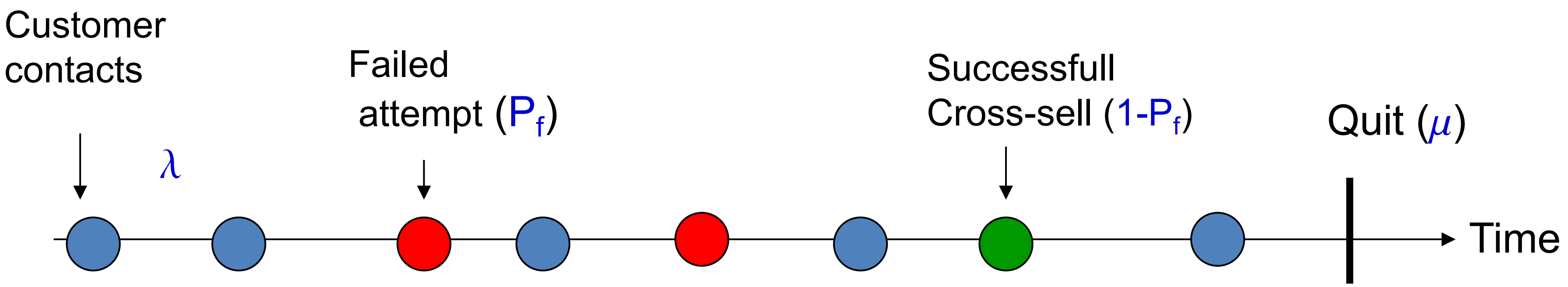
- Aksin and Harker (1999)
- Ormeci and Aksin (2006)
- Byers and So (2007)
- Armony and Gurvich (2006)
- Gurvich et al. (2006)

congestion vs. revenue

Gunes, Aksin, Ormeci, Ozden (2010)
Two Features of a Customer Relationship

j : # of contacts

Customer Evolution:
Maturity increases the chances of buying the
next product (Li et al. 2005, Kamakura et al. 1991)



Customer contacts

λ

Failed attempt (P_f)

Successfull Cross-sell ($1-P_f$)

Quit (μ)

Time

i : # of failed atempts

Customer Reaction:
Excessive x-selling may irritate customers
(Kamakura et al. 2003, Eichfeld 2006)

Customer reacts:
- lowers utilization of service? (λ)
- quits relationship earlier?(μ)
- less inclined to accept future attempts? (P_f)

Overview of results from MDP analysis

- Optimal cross-sell policy is of threshold type with no customer evolution or reaction
- Optimal cross-sell policies are of state-dependent threshold type when only evolution or reaction is present
- When both evolution and reaction is present, policies can take complex dynamic forms

Modeling $P_f(i,j)$

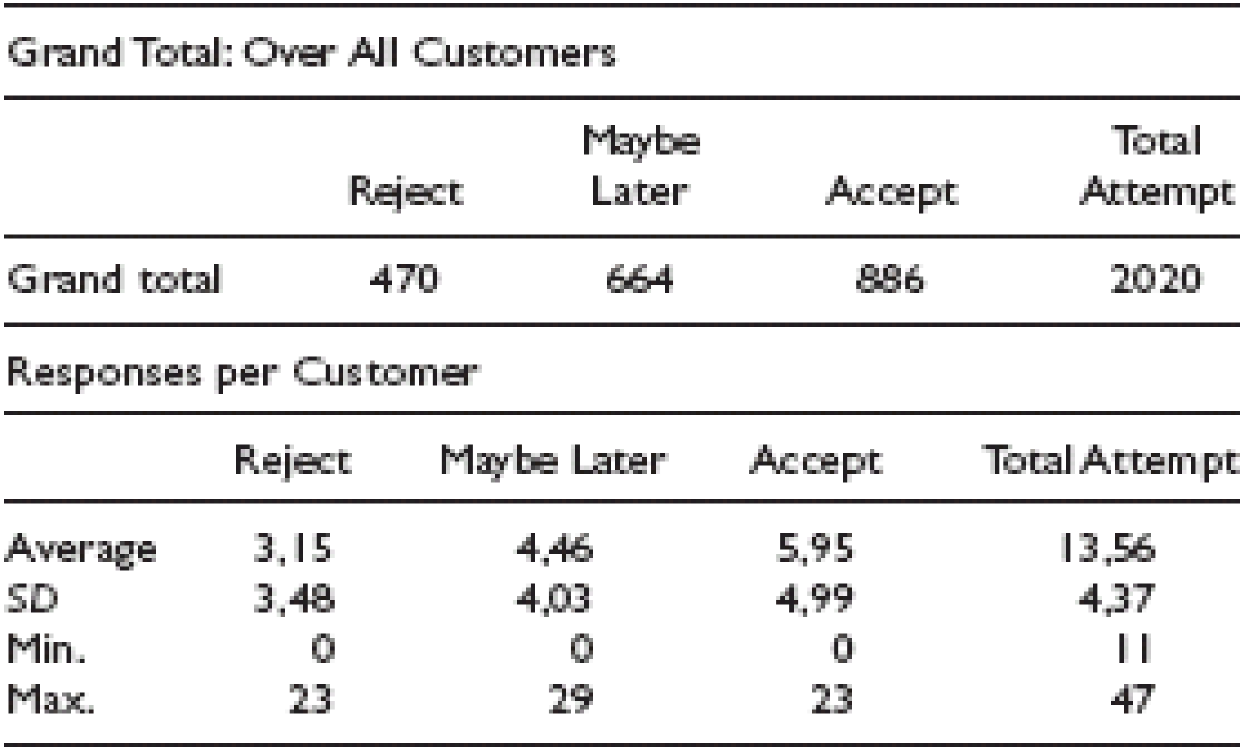
Customers' utility of not buying the proposed product relative to buying is $U(i,j)$

$U(i,j)= \beta_0+\beta_1 i+\beta_2 j+\varepsilon$

$$P_f(i,j) = \frac{1}{1+e^{-U(i,j)}}$$

How can $P_f(i,j)$ be estimated in practice?

- Using data from a bank
- 149 customers over a 2 year period



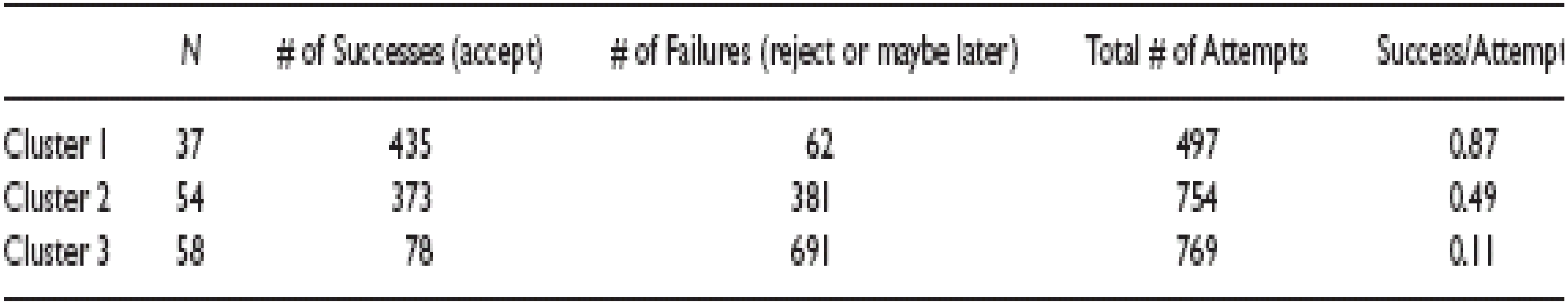
Grand Total: Over All Customers

	Reject	Maybe Later	Accept	Total Attempt
Grand total	470	664	886	2020

Responses per Customer

	Reject	Maybe Later	Accept	Total Attempt
Average	3,15	4,46	5,95	13,56
SD	3,48	4,03	4,99	4,37
Min.	0	0	0	11
Max.	23	29	23	47

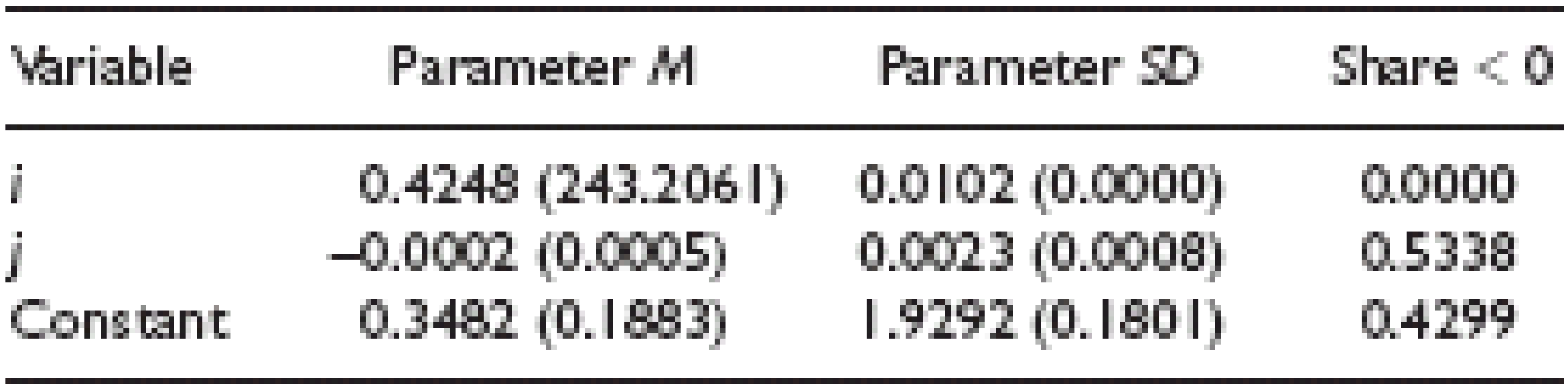
A clustering analysis based on success/attempt ratio



	N	# of Successes (accept)	# of Failures (reject or maybe later)	Total # of Attempts	Success/Attempt
Cluster 1	37	435	62	497	0.87
Cluster 2	54	373	381	754	0.49
Cluster 3	58	78	691	769	0.11

Estimation using random coefficient logit

- Dataset contains cross-sell attempt dates and outcomes
- We use time as a proxy for *j*
- The outcomes are used to determine *i*- we count both reject and maybe later as failure



Variable	Parameter *M*	Parameter *SD*	Share < 0
i	0.4248 (243.2061)	0.0102 (0.0000)	0.0000
j	–0.0002 (0.0005)	0.0023 (0.0008)	0.5338
Constant	0.3482 (0.1883)	1.9292 (0.1801)	0.4299

What makes all of this exciting?

- Modeling of customer role in services is an inter-disciplinary area
  - Economics framework
  - Behavioral framework: New models motivated by behavioral findings
- Choice models and structural estimation
  - Allows capturing strategic customers
  - Combines theoretical modeling with practice through estimation
- Data availability
  - Better field data
  - Lab experiments

Thank You