The Market for Air Travel: What People Pay to Fly

Pacific Northwest AIAA
Technical Symposium
November 1, 2014

Roger A. Parker, PhD
Chief Technology Officer
AirMarkets Corporation

AirMarkets Corporation

Contents

- The Market Structure of the Global Airline Network
- Air Travel Passengers and How They Behave
- The AirMarkets Micro-Simulation
- The Air Fare Empirical Distribution Function

AirMarkets Corporation

The Market Structure of the Global Airline Network

A brief discussion of the structure and dynamics of the world's airline passenger network.

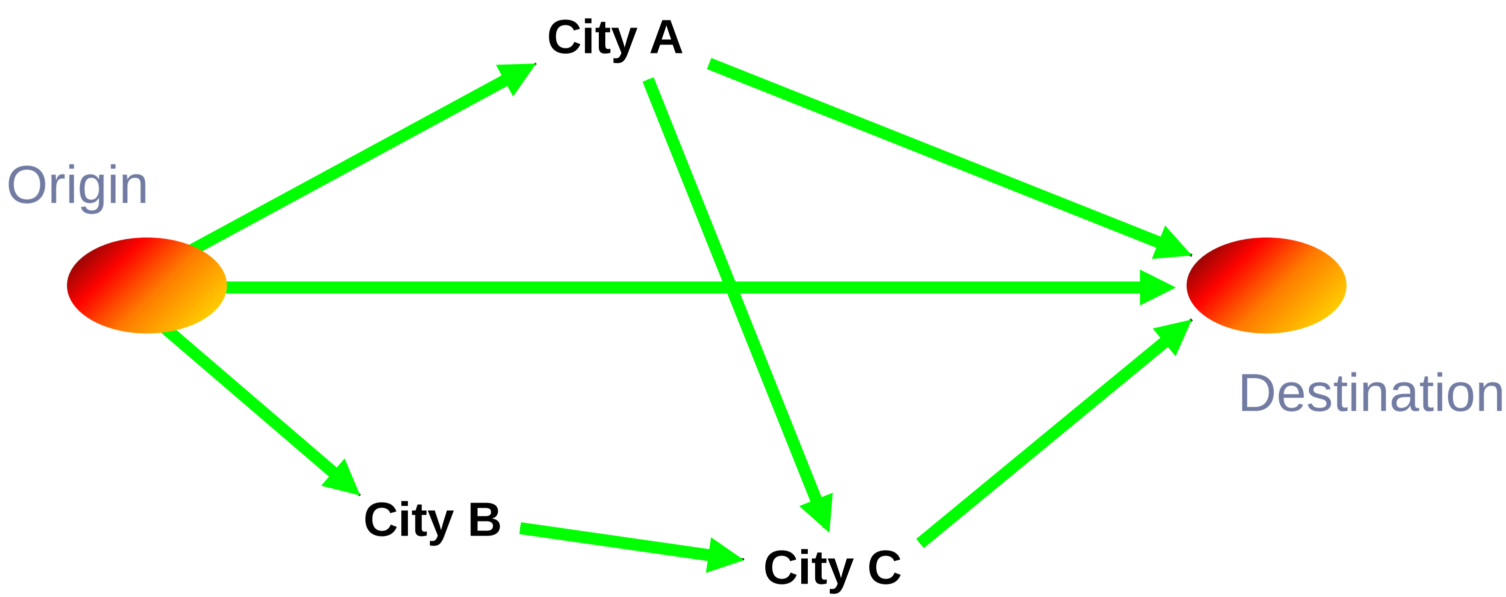
November 1, 2014

Airline Network Components

- Cities (the airports that serve them)
- The service between airports (flight legs)
- Itineraries are sequences of legs connecting cities
- Flight legs have capacity limits
- Key properties
  - Exists in three dimensions – space(2) and time
  - Legs are unique in time, but not in space
  - Network is a *directed graph,* and is *scale-free*
  - Circuit free – no leg starts and ends at the same city
  - Network is *everywhere connected* – any city can be reached from any other city (although not necessarily conveniently)

AirMarkets Corporation

A Simple Airline Network



Four distinct itineraries from origin to destination: O>D, O>A to D, O>B>C>D, and O>A>C>D.

Seven flight legs.

Nine markets: O>D, O>A, O>B, O>C, A>D, A>C, B>C, B>D, and C>D

AirMarkets Corporation

Dynamics of Capacity and Demand

- The legs in the network are capacity constrained
- The origin-destination demand which the network serves is a *random variable* for each market
- Capacity for a flight is consumed over the time leading up to the departure, and its use is a stochastic process (the 'ticketing curve')
- A ticket sold to a passenger in one OD may remove an itinerary choice from future booking passengers in that or other OD's served by the legs of the itinerary

How Big is the World Network?

- In August of 2008
  - 3319 cities have regularly scheduled service (at least once/week)
  - 694,784 unique flight legs
  - 11,012,422 markets
  - 39,520 nonstop markets
  - 733,132 one-stop markets
  - 3,948,632 two-stop markets
  - 5,042,051 three-stop markets
  - 1,249,087 four-stop markets
  - Approximately 42,000,000 passengers
- Recent changes have made this smaller

Air Travel Passengers and How They Behave

How passengers choose a travel itinerary, and how it is represented in AirMarkets

November 1, 2014

AirMarkets Corporation

What Drives Flight Choice?

Which airline should I take?

When do I want to arrive?

Are the kids coming?

How much will I pay?

Is my company paying?

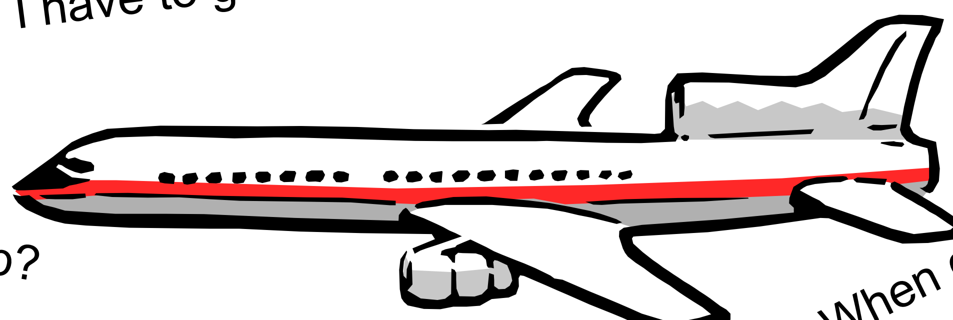
Do I have to make stops?

Can I take car instead?

Do I have to take a regional jet?

Will my knees be crushed into the seat ahead of me?

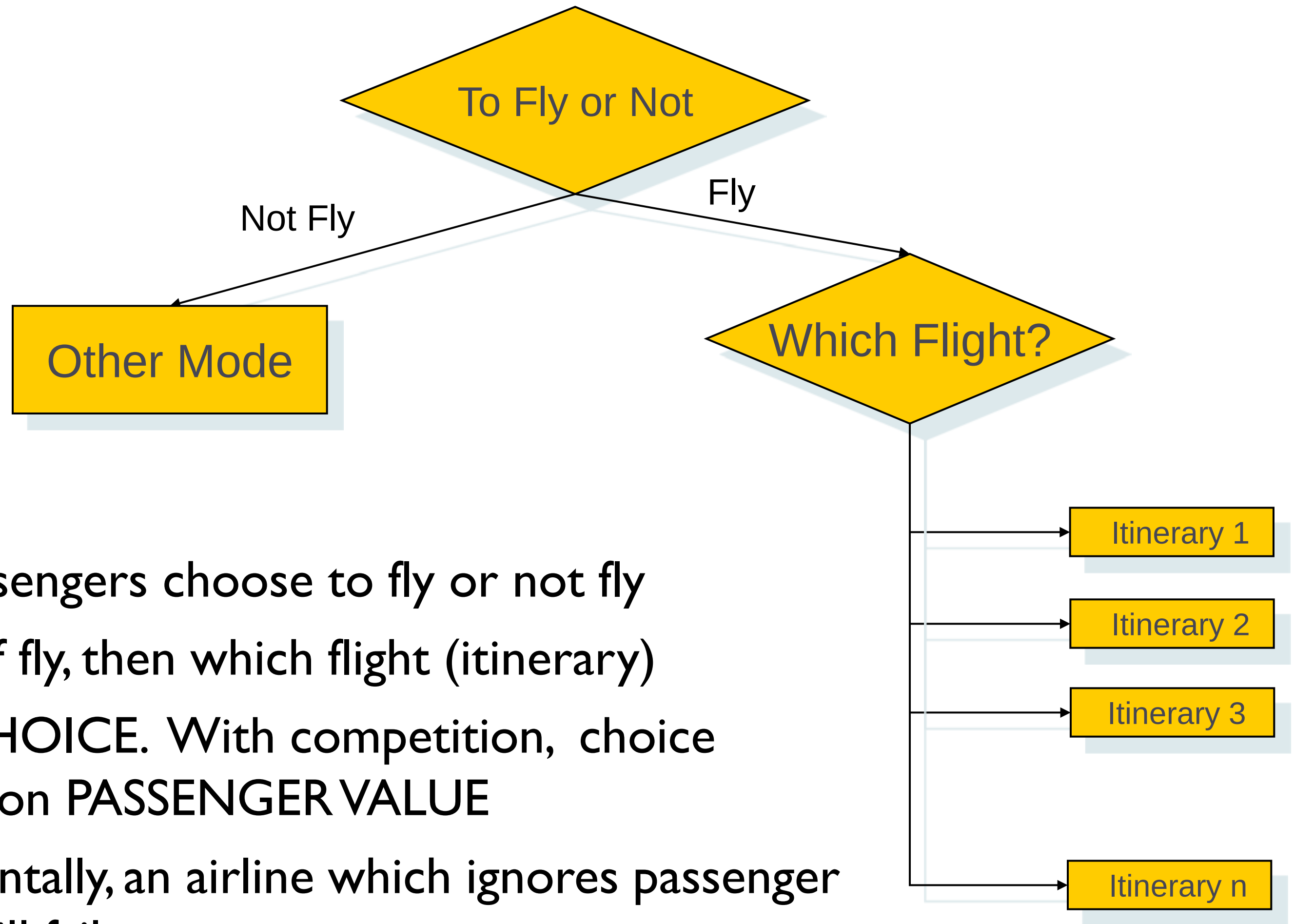
How far do I have to go?



How urgent is the trip?

When do I want to depart?

The Passengers Make *Two* Choices



- First, passengers choose to fly or not fly
- Second, if fly, then which flight (itinerary)
- Key is CHOICE. With competition, choice depends on PASSENGER VALUE
- Fundamentally, an airline which ignores passenger choice will fail

Example: Value of Travel Time

Leisure and self-pay business
Value of time ~ $40 / hour

Reimbursed business
Value of time ~ $150 / hour

Value of time for business travelers is consistent with industry practice of premium pricing non-stop flights (which save 1-2 hours) by $200 -$300.

AirMarkets Corporation

Passenger Choice Model

Passengers and their trips have *Characteristics*

- Ability to Pay
- Gender
- Age
- Income
- Occupation
- Country of Origin
- Previous travel
- FF membership
- Trip Purpose
- Desired dep/arr times
- Duration of stay

\+

Alternative itineraries have *Attributes*

- Fare
- Cabin
- Number & type of stops
- Depart/arrival times
- Total trip duration
- Airline
- Aircraft
- FF Awards

→

Choice Probability

The highest ranked is the most likely to be chosen

Passengers evaluate each alternative given its attributes and their characteristics

Agent Itinerary Choice Model

The utility function agent i uses to evaluate itinerary option j is

$$V(i,j) = \beta_f(i)\ln f(j) + [\beta_d(i) + \beta_{bd}(i)\ln d_{base}]d(j)$$
$$+[\beta_{16}(i)X_{16}(j) + \beta_{710}(i)X_{710}(j) + \beta_{1120}(i)X_{1120}(j)]d(j)$$
$$+\beta_{dc}(i)N_{dc}(j) + \beta_{ic}(i)N_{ic}(j) + \beta_{1st}(i)X_{1st}(j) + \beta_{ec}(i)X_{ec}(j)$$
$$+G(\tau(i) - t(j))$$

where the function G is the *schedule delay disutility*, defined as

$$G(\tau(i)-t(j)) = \begin{cases} \beta_E^G(i)\dfrac{(t(j)-\tau(i)-a+1)^{\lambda_E}-1}{\lambda_E} & \tau(i)-t(j) < -a \\ 0 & -a < \tau(i)-t(j) < b \\ \beta_L^G(i)\dfrac{(\tau(i)-t(j)-b+1)^{\lambda_L}-1}{\lambda_L} & \tau(i)-t(j) > b \end{cases}$$

Itinerary Choice Utility Variables

AirMarkets Corporation

- $f(i)$ = fare for itinerary i
- $d(i)$ = duration of itinerary i
- d_{base} = shortest duration in the market
- $X_{16}(i)$, $X_{710}(i)$ $X_{1120}(i)$ = dummy variables for trip journey structure
- $N_{dc}(i)$ = number of direct connects (same airline or alliance) in itinerary i
- $N_{ic}(i)$ = number of indirect connects (different airline or alliance) in itinerary i
- $X_{1st}(i)$, $X_{ec}(i)$ = dummy variables for the cabin used by itinerary i
- $G(\ldots)$ = schedule delay disutility (discussed later)
- Ψ = the set of airlines in the scenario
- $I(a) = 1$ if itinerary j uses airline a, 0 otherwise
- $F(i, a)$ = the frequent flyer mileage value of pag i on airline a.

The $\beta(i)$'s are random coefficients with normal distributions (truncated as appropriate) with empirical means and standard deviations.

The Logit Model

Then we have

$$U(j) = V(j) + \varepsilon$$

where

$U(j)$ = the utility of alternative j,
$V(j)$ = the observed utility of alternative j,
ε = the unobserved portion of the utility.

If the unobserved utility ε is distributed with an extreme value type I distribution, then

$$\Pr[\text{choice is } j] = P(j) = \frac{e^{V(j)}}{\sum_{i=1}^{J} e^{V(i)}}$$

where

$e = 2.71828182845\ldots$
$P(j)$ = the probability of making choice j.

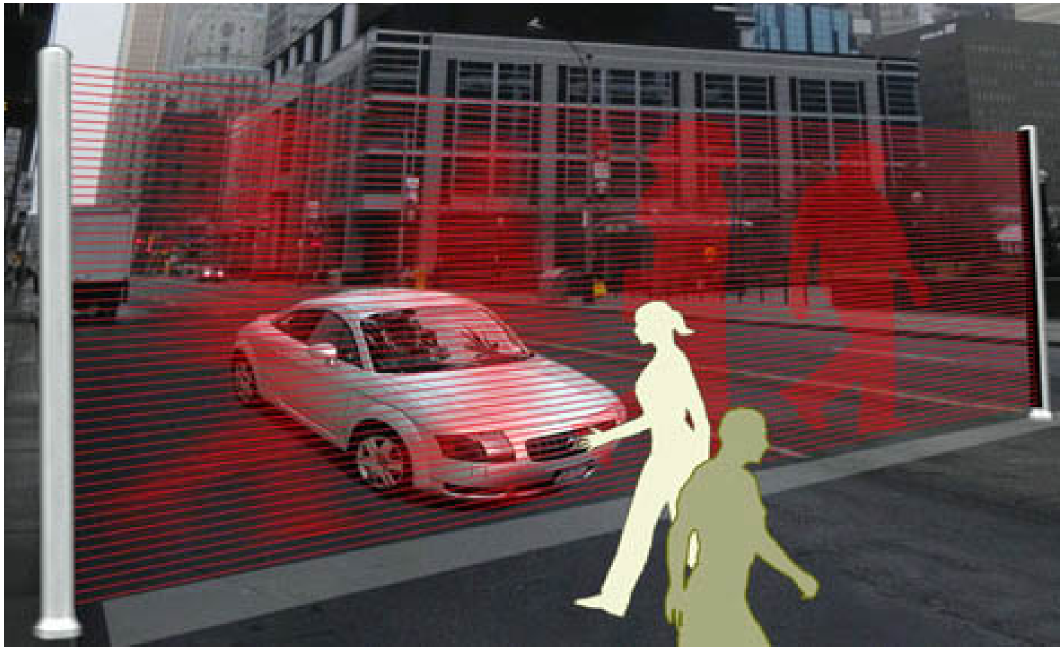
The AirMarkets Agent-Based Micro-Simulation

A computational science technique for studying complex systems which cannot otherwise be analyzed.

November 1, 2014

Agent-Based Modeling

Agents are computer objects which duplicate the important behavior of customers and suppliers observed in the real world. Agent-based models are virtual worlds.



Advantages of Agent-Based Modeling

- Easily represents dynamics, especially stochastic dynamics (based on probabilities).
- Does not need to assume "representative" averages, so heterogeneity is expressly represented.
- Directly handles complexity.
- Accounts for the interactions between individuals.
- Allow bounded rationality using incomplete information, even irrational behavior.
- Can run hundreds of replications (Monte Carlo), thus producing probability distributions.
- An example of computational science

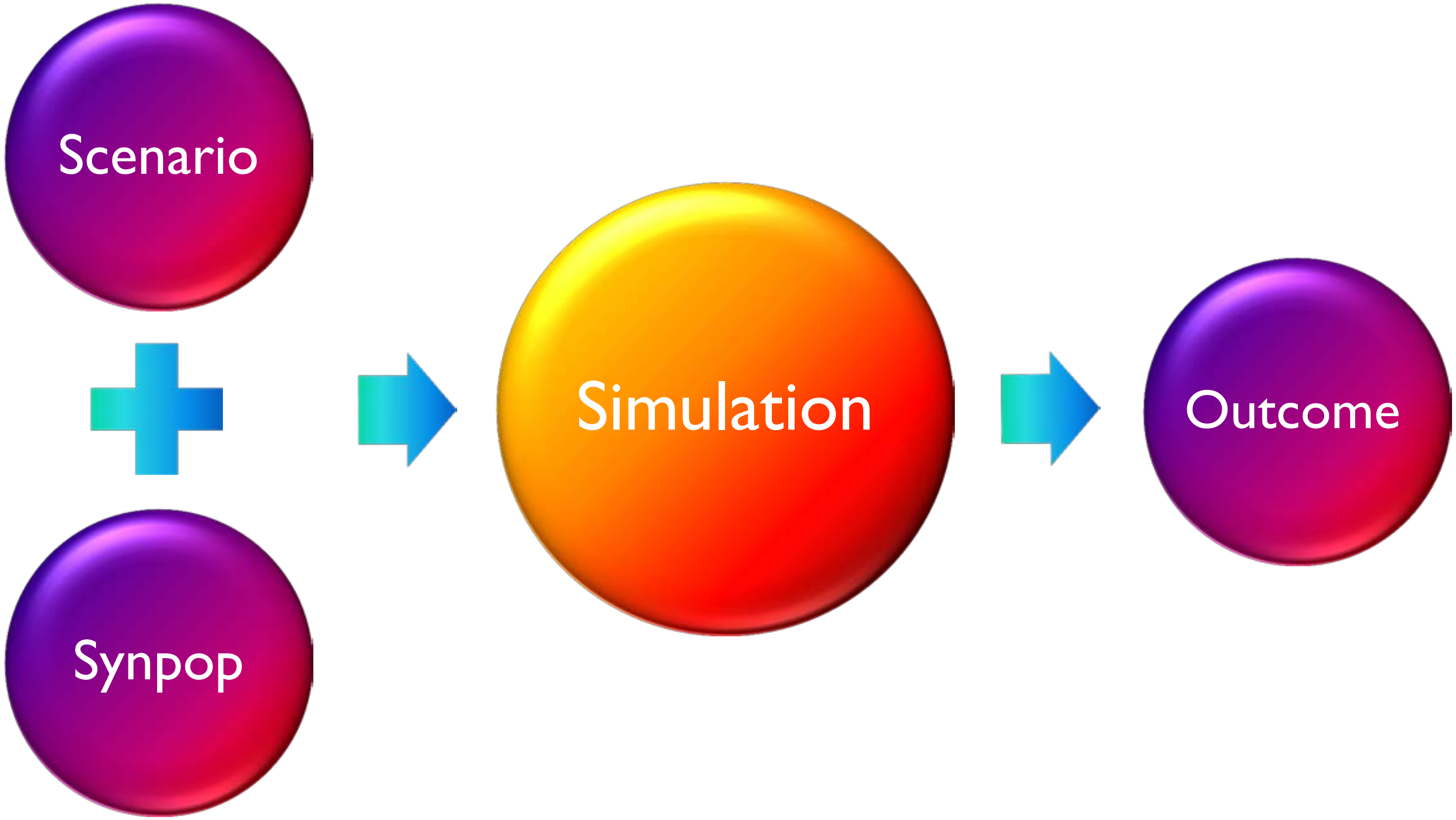
AirMarkets Features

- Basic perspective is the passenger
- Passengers purchase tickets for an itinerary
- Flight legs serve more than one itinerary
- Passengers on a given flight are in many markets
- Passengers buy tickets in advance
- Airlines can vary prices as demand grows
- Airlines compete on price and service
- The entire global airline network is represented
- Travel for one calendar week (the *standard* week)

Passenger Agent Properties

- Each agent has a "ticketing instant" which sets up the fundamental dynamics of the network.
- Each agent may buy one or more tickets
- Each agent has a unique
  - Trip purpose
  - Journey structure
  - Willingness-to-pay
  - Ticketing instant time
  - Ideal departure/arrival time
  - Itinerary choice utility parameters

Scenarios, Synpops, Simulations and Outcomes



AirMarkets Corporation

Data Sources

- Industry data
  - Schedule data from OAG, Innovata
  - Ticketing data from GDS's, IATA or ARC
  - Pricing data from GDS's, IATA, ARC
  - Observed origin-destination patterns from GDS's, IATA, ARC
- Passenger survey data
  - Trip purpose
  - Journey structure
  - Ideal departure/arrival times
  - Itinerary choice model parameters
  - Willingness-to-pay model parameters
  - Ticket cancellation parameters
  - Origin-destination patterns
- Government data
  - Origin-destination patterns
  - Economic activity (unemployment, fuel price, etc.)

The AirMarkets Distribution of Fares in a Market

An important validation of AirMarkets results is the duplication of the observed distribution of fare revenue

November 1, 2014

What's an Empirical Distribution Function (EDF)

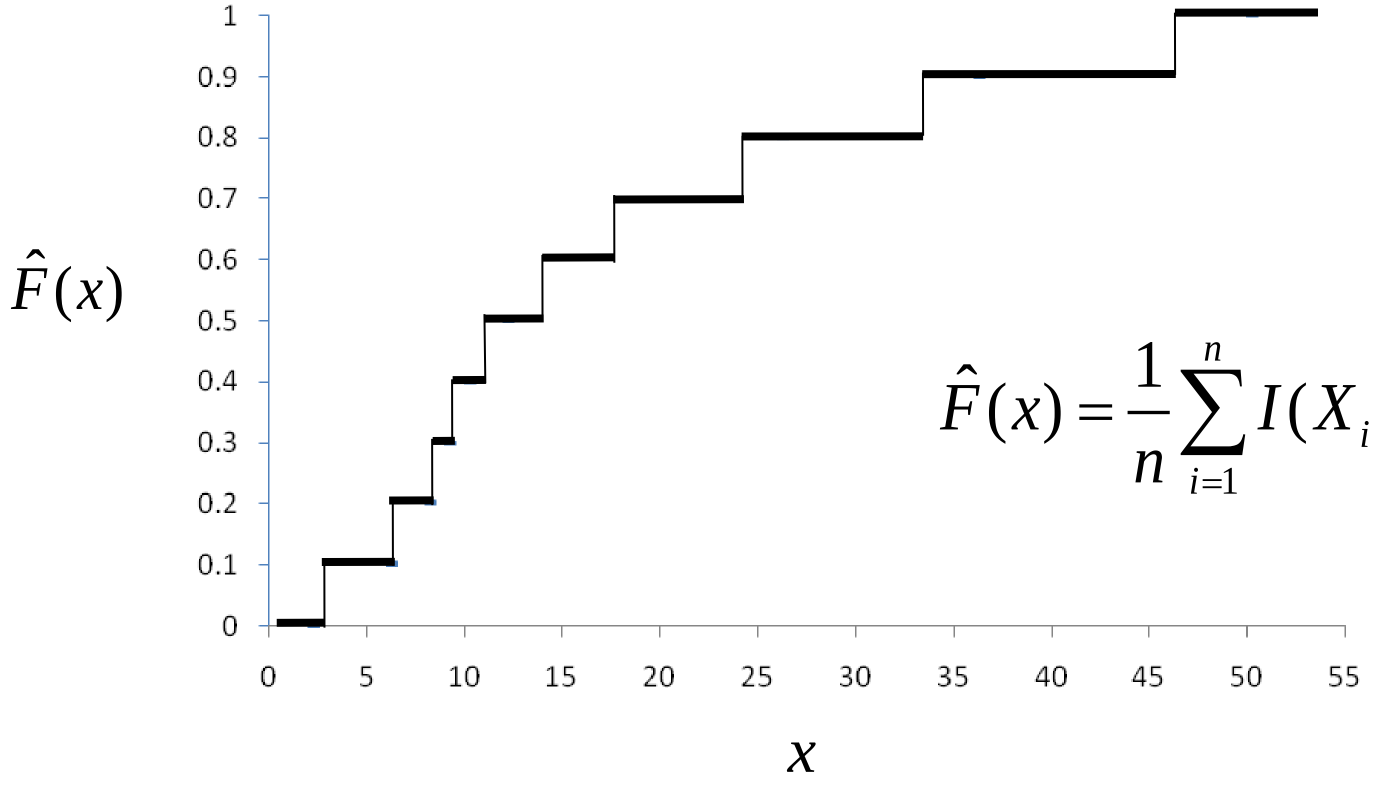
No general distribution form for the fares in a market is intuitively evident. So use the fare Empirical Distribution Function (EDF) to represent the fares. An EDF is defined as

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)$$

where *I(x)* is the indicator function

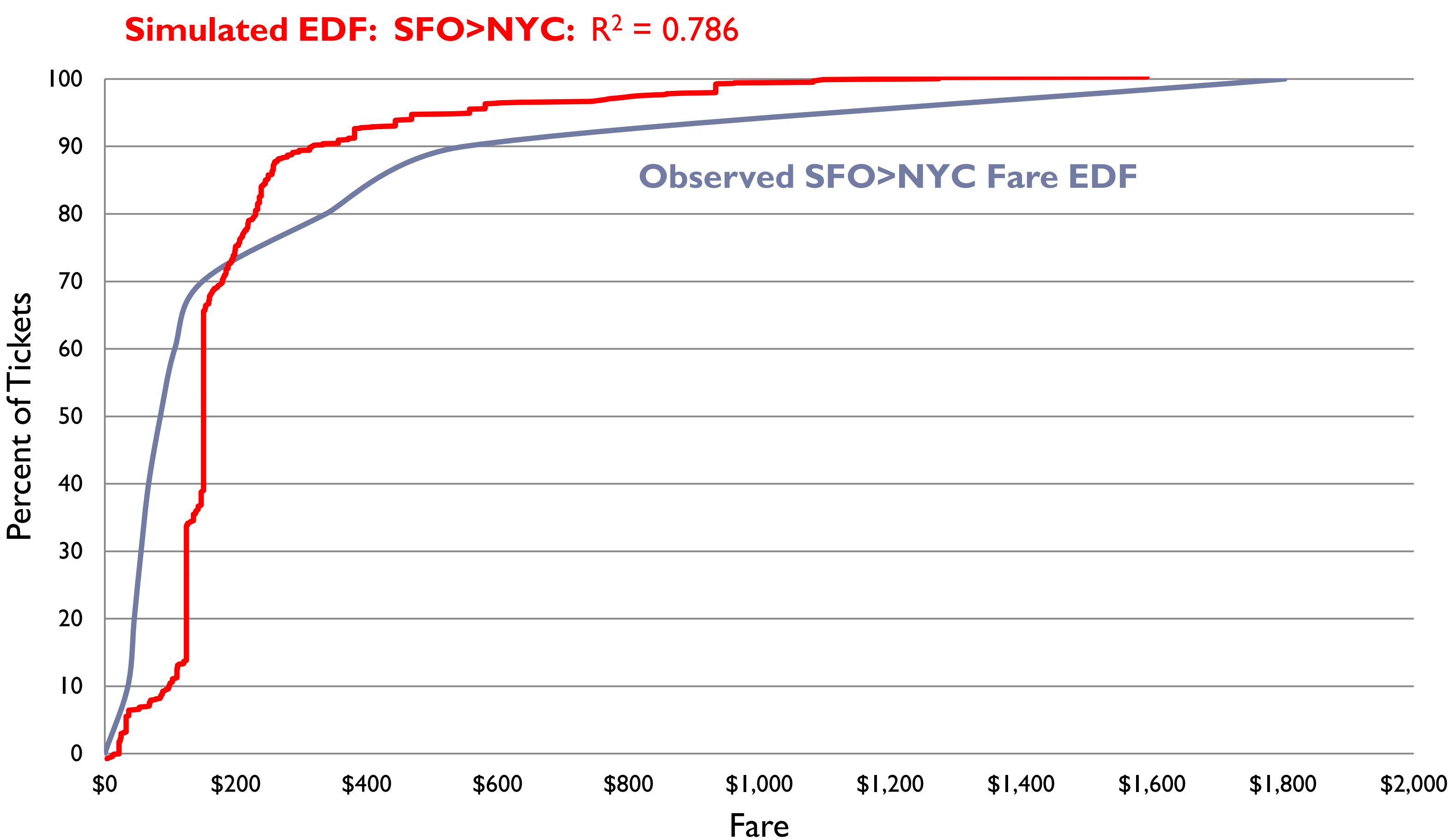
$$I(X_i \le x) = \begin{cases} 0 \text{ if } x < X_i \\ 1 \text{ if } \mathrm{X}_i \le x \end{cases}.$$

EDF Plot



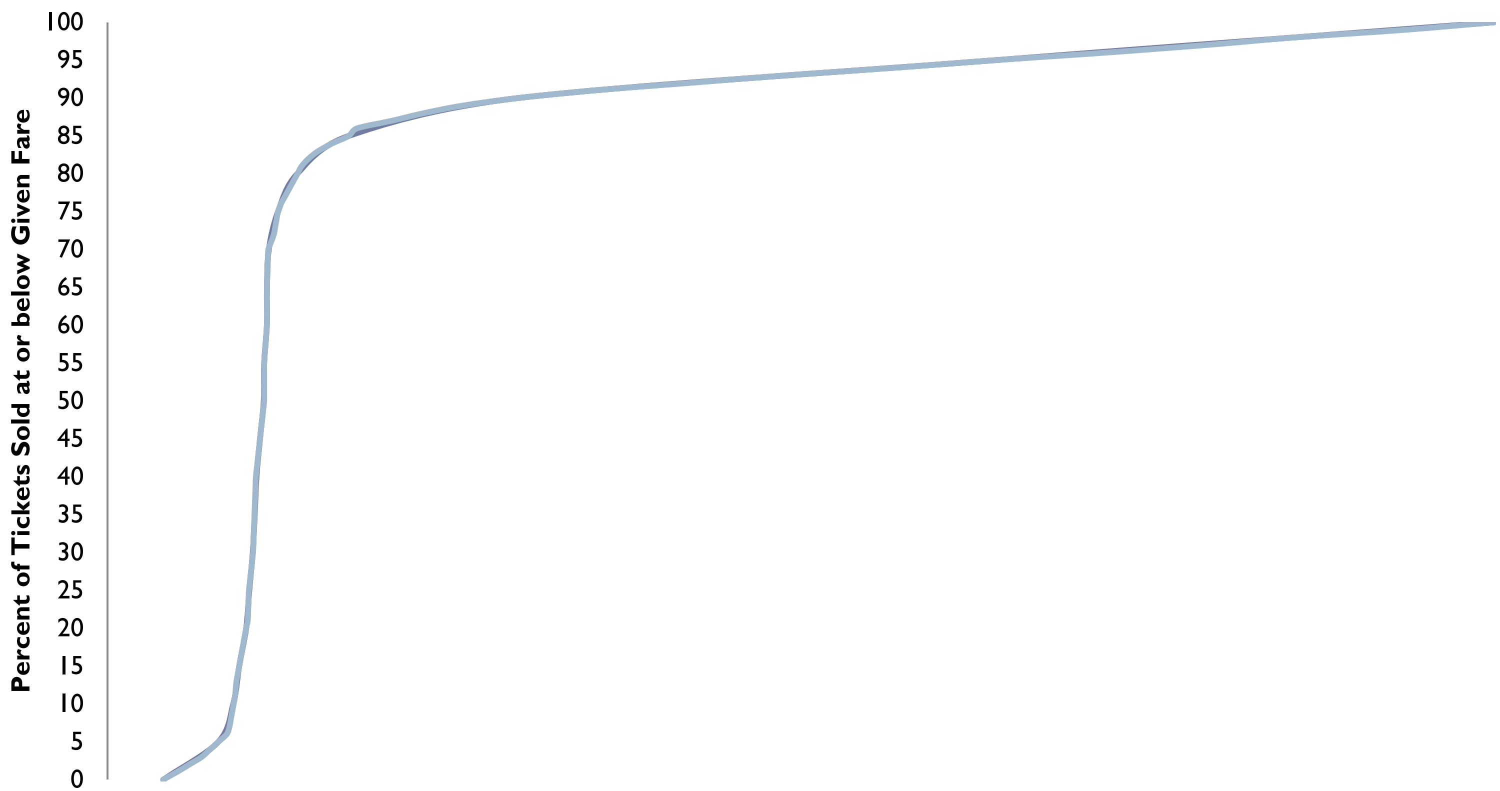
$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x)$$

Observed vs Simulated Fare EDF's



Data source: ARC; AirMarkets

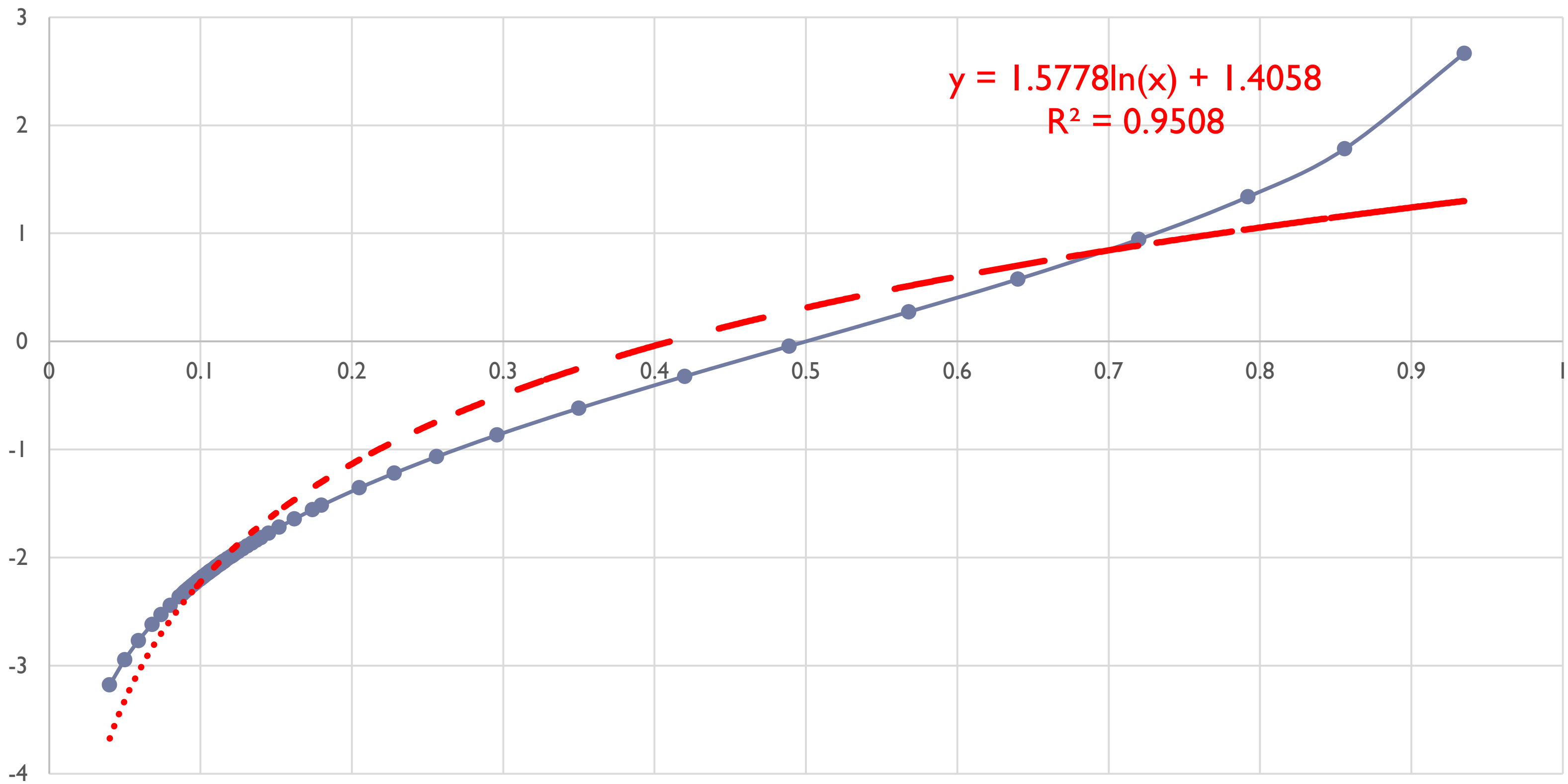
Generic Fare Profile *F(s)*



Based on data from >17,500 markets in US and Europe (ARC and IATA data).
Data fits a known equation form (more or less).

Generic EDF Transforms

Let $F(s)$ be the generic EDF curve above. Then apply the Fisher-Pry (log odds) transform $\ln[F(s)/1\text{-}F(s)]$



This curve is reasonably approximated by a log linear regression line.

Fare Profile Generic Equation

The fraction of tickets sold at fare s, $F(s)$, is well represented by the following:

$$F(s) = \frac{e^B s^A}{1 + e^B s^A}$$

where A and B are empirical constants found by computing the following regression from data:

$$\ln\left[\frac{\hat{F}(s)}{1 - \hat{F}(s)}\right] = A\ln(s) + B$$

The Flat Tail Question

- Data is based on observed ticket sales, which has an absolute maximum price
- Log-linear approximation to the log-odds transform doesn't fit upper right tail of the EDF
- If the EDF is adjusted to reach one in the limit (thus allowing prices to rise indefinitely), then the log-linear fit is much better
- Suggests a "luxury good" property for very high priced tickets; i. e. price does not matter

Example: LON>MIA Supersonic

- AirMarkets simulation executed with the availability of a private, supersonic aircraft.
  - Private aircraft has five times the comfort of a first class commercial carrier.
  - Speed (over water) is Mach 1.4
  - Fare was set at $55,000 (~$15.00/mile)
- Total tickets purchased in the London>Miami market is about 5,800 (for a week).
- The simulation showed 48 tickets sold for the supersonic option, or about 0.83%.
- Is this reasonable?

More Research Questions

- Why the differences between Actual and Simulated EDF?
- Why does the log odds formulation fit so well so often?
- What are the relationships between the values of A and B and network parameters such as
  - Distance/travel time between the origin/destination
  - airline pricing policies, including RM systems
  - aircraft performance capabilities

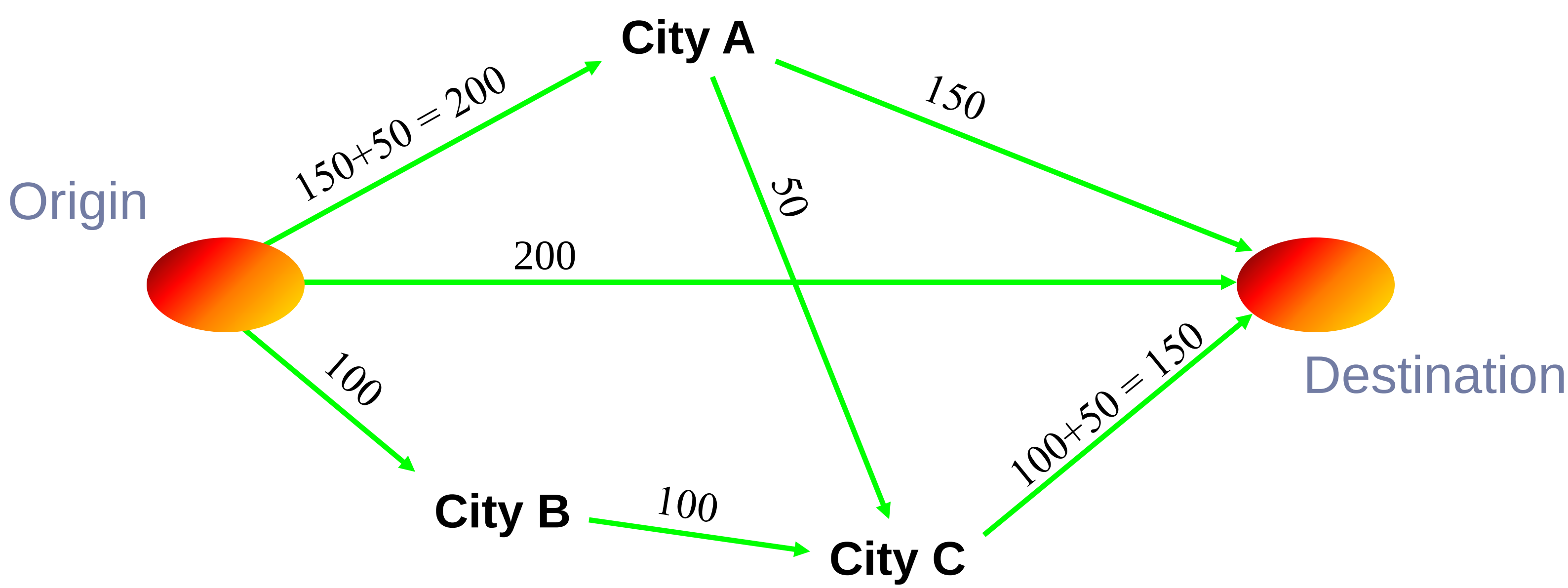
Thank You
Questions

AirMarkets Corporation

Back-Up Slides

Passenger Loads on Flights

If total demand is from O to D is 500, and the probabilities of each itinerary are as shown below, then the demand on each itinerary is just the probability of choosing that itinerary times 500.

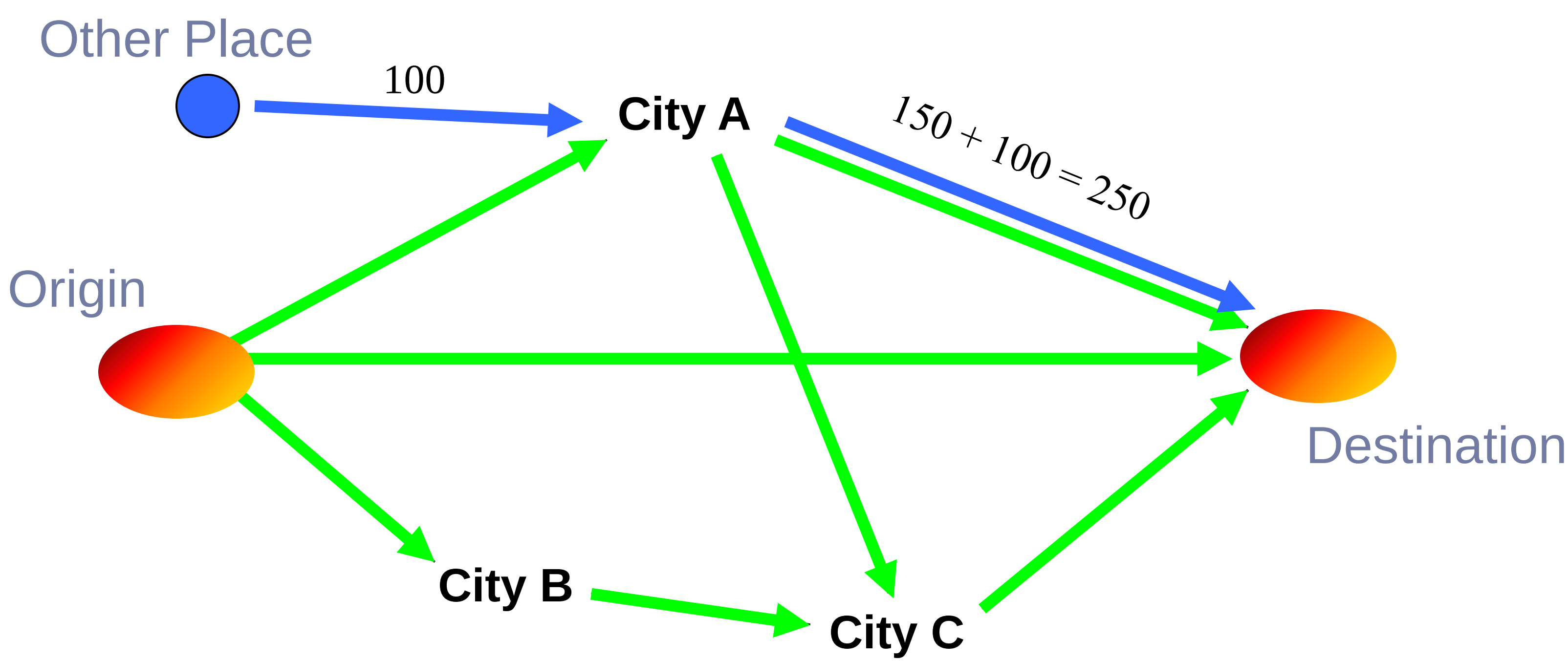


Prob(O>D) = 40%	Prob(O>A>D) = 30%	Prob(O>B>C>D) = 20%	Prob(O>A>C>D) = 10%
Dem(O>D) = 200	Dem(O>A>D) = 150	Dem(O>B>C>D) = 100	Dem(O>A>C>D) = 50

The *load* on each *leg* depends on the network geometry

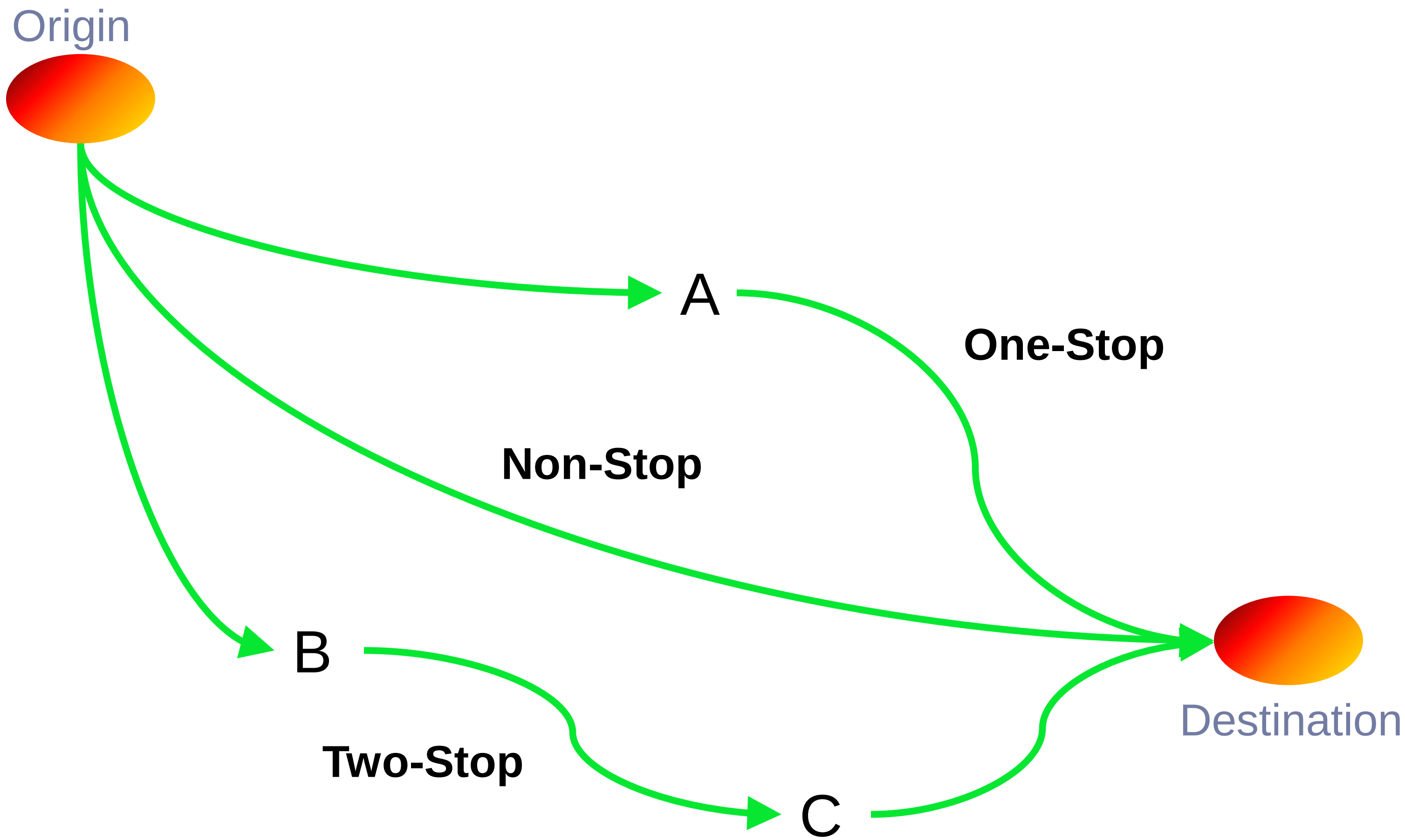
AIRMARKETS CORPORATION

Who is Traveling on a Flight



Not only does a leg carry traffic for O>D, it also carries local traffic, e. g. City A to City D. Plus traffic from Other Places coming through city A going to D.

Itineraries and Flights



AirMarkets Corporation

Passengers Choose Based on Trade-Offs

- Passengers don't decide solely on price – they make trade-offs.
  - Avoid a stop
  - Arrive at a better time
  - As cheap as possible
  - Favored airline
- Trade-offs depend on the flight circumstances
  - Holidays differ from business trips
  - With the family different than traveling alone
  - Who is paying
- Trade-offs depend on passenger characteristics
  - Young people are price sensitive; retired people travel mid-day;
  - Income sets willingness-to-pay
  - Gender, background, prior experience etc. all contribute

Random Utility Models

Utility = Observable Utility + Unobservable Utility

Total utility of the Alternative

Observable utility of the Alternative

Unobservable utility of the Alternative

A *Random* Variable

Choice can be Known Only Up to a Probability

- *We can know only how likely each choice will be, not what that choice will be*
- Probability will depend on the option attributes, and the chooser characteristics.
- Probability is defined in terms of attributes and characteristics, and probability changes as they change.
- Probability = market share, (demand x probability) so we can compute the effects on market share of attribute changes.

Synthetic Populations (Synpops)

- An AirMarkets synpop is the collection of passengers moving in all markets.
- The size of a synpop depends on the data in the OD matrix.
- The same scenario can be analyzed with different synpops.
- The same synpop can be applied to many scenarios.

The AirMarkets Synpop Models

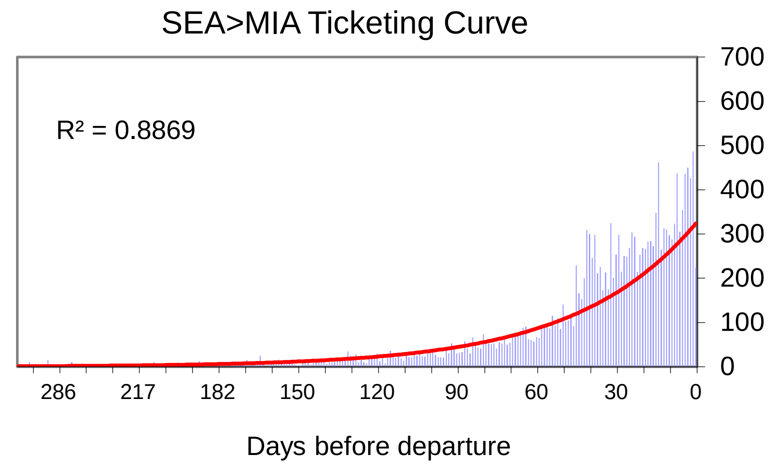
- Trip purpose
- Journey structure
- Travel group size
- Ticketing instant
- Itinerary choice
- Arrival/departure sensitivity
- Ideal departure/arrival time
- Willingness-to-pay
- Ticket cancellation

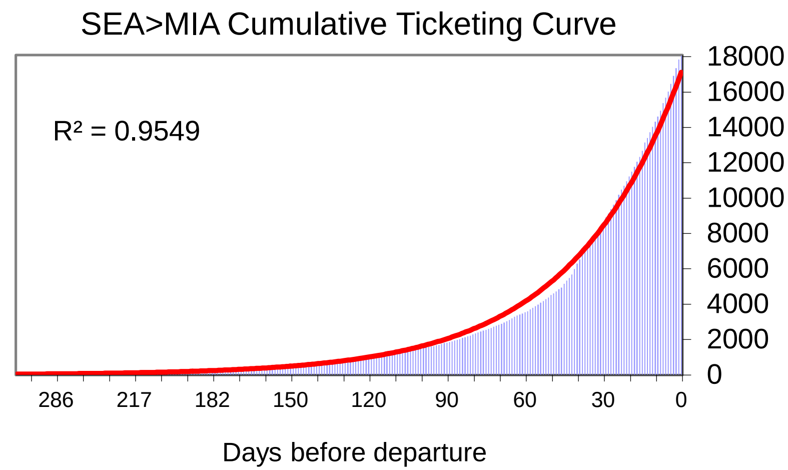
Note: All models have system-wide default parameters which can be overridden with market-specific values.

Example Ticketing Curve Data

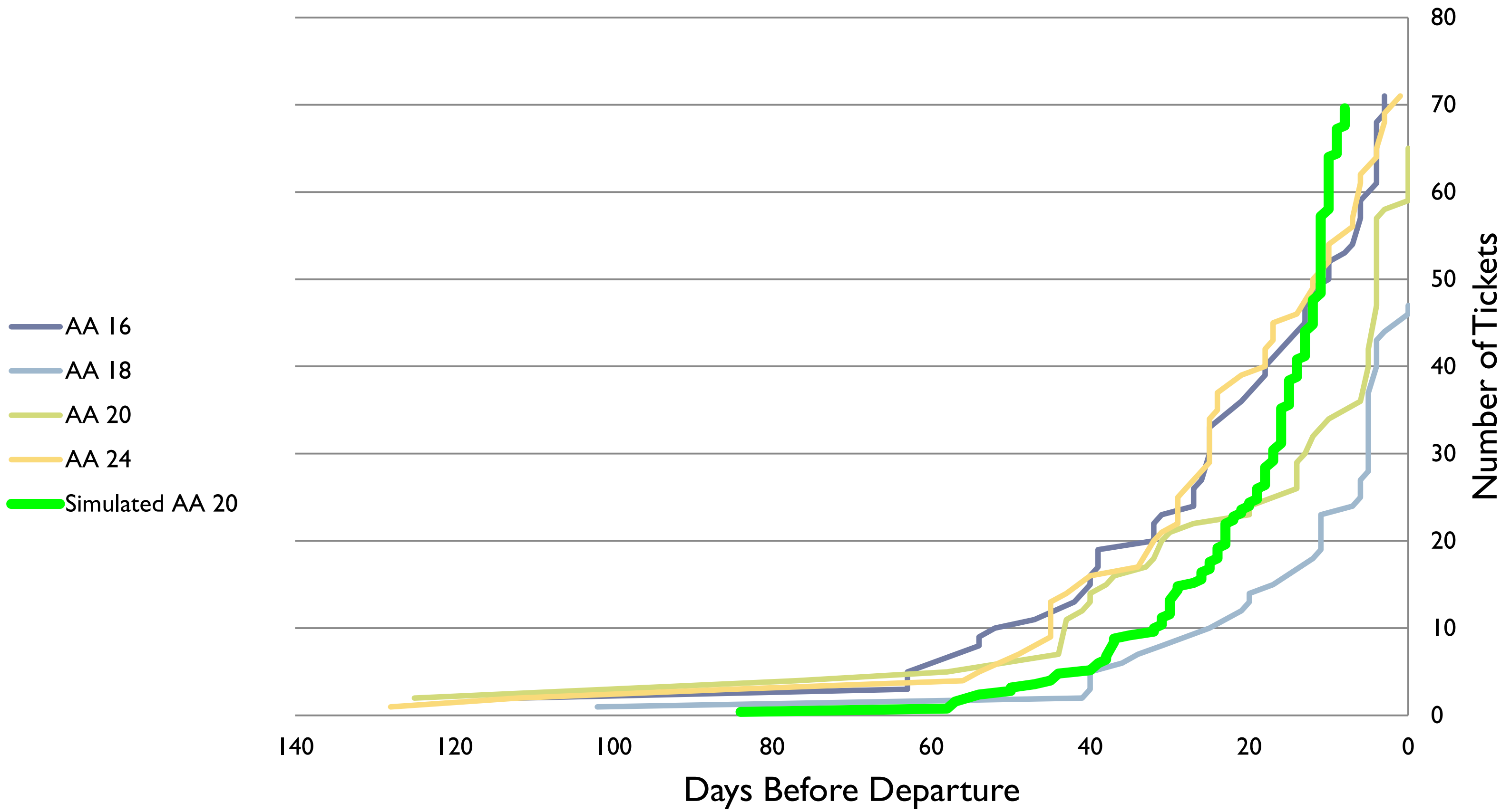
Comparison of total market tickets sold curve with theoretical values shows a reasonably good fit. R^2 values of 0.887 for the ticketing instance curve, and 0.955 for the total tickets sold ticketing curve. Similar results for many markets.

Data source: Industry data; passenger surveys



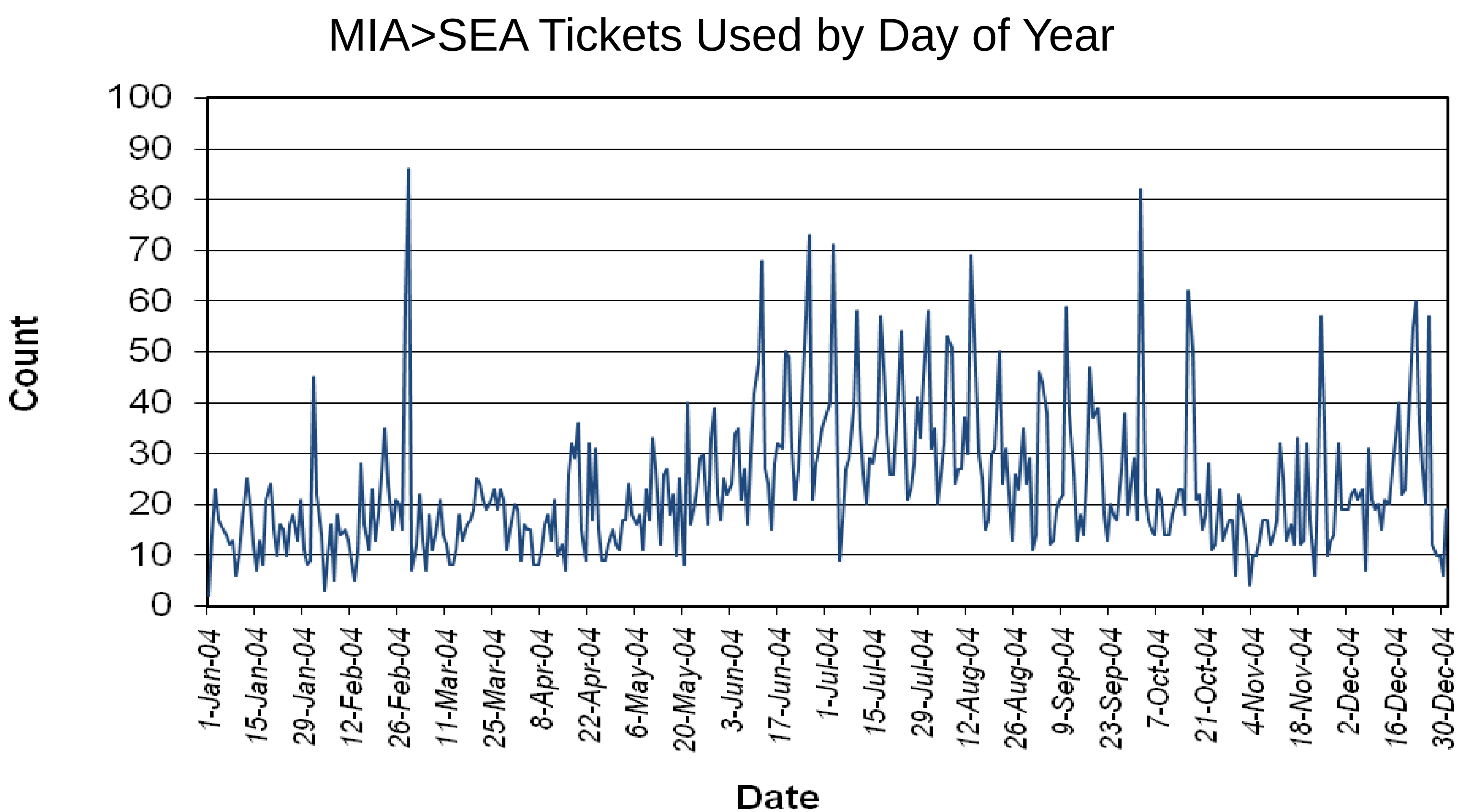


Simulated vs. Observed Ticketing



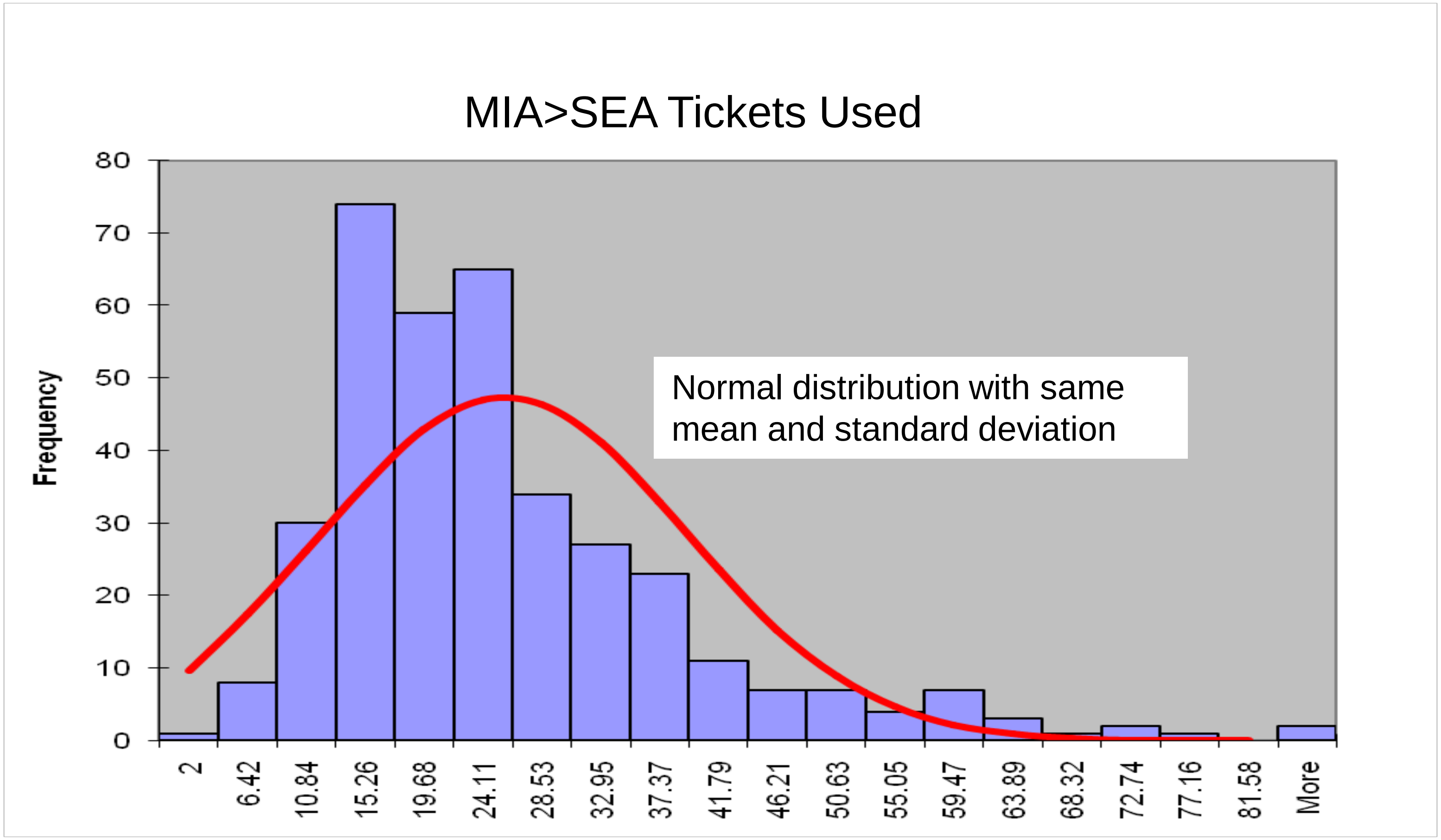
Source: Industry data; AirMarkets

Inherent Variation



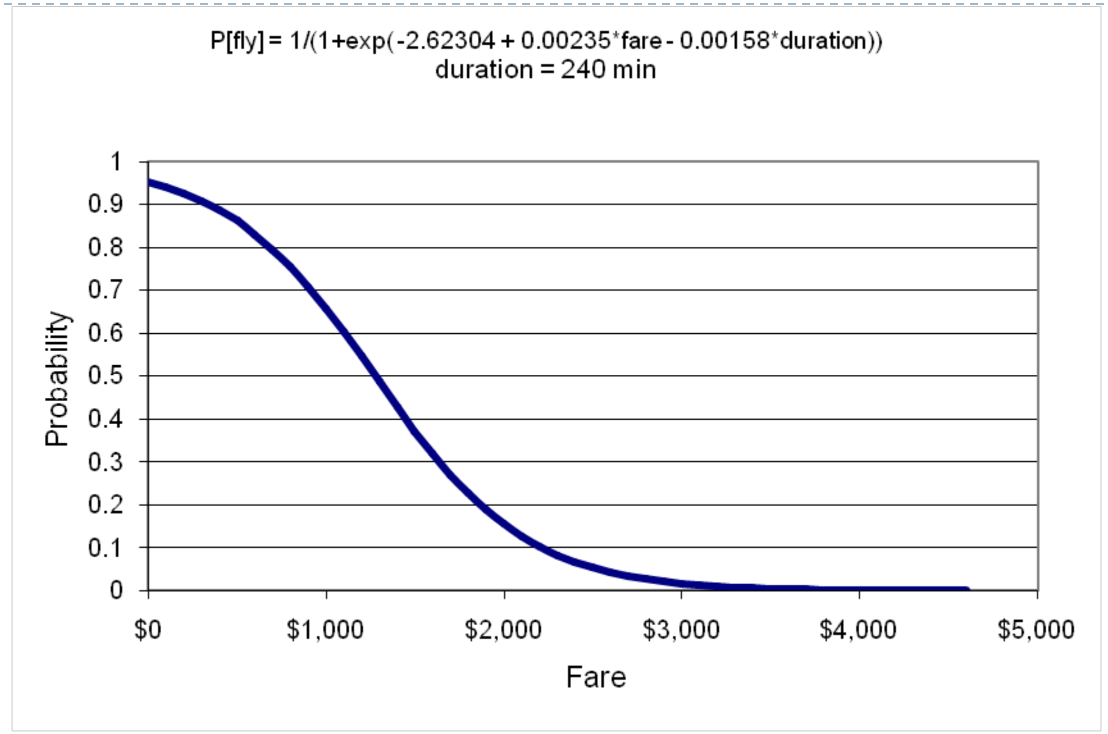
Source: ARC

Inherent Variation Distribution



Source: AirMarkets

Willingness-to-Pay Distribution



This formulation allows a WTP computation for any market, but much additional work needs to be done in this area.

Data source: Passenger surveys