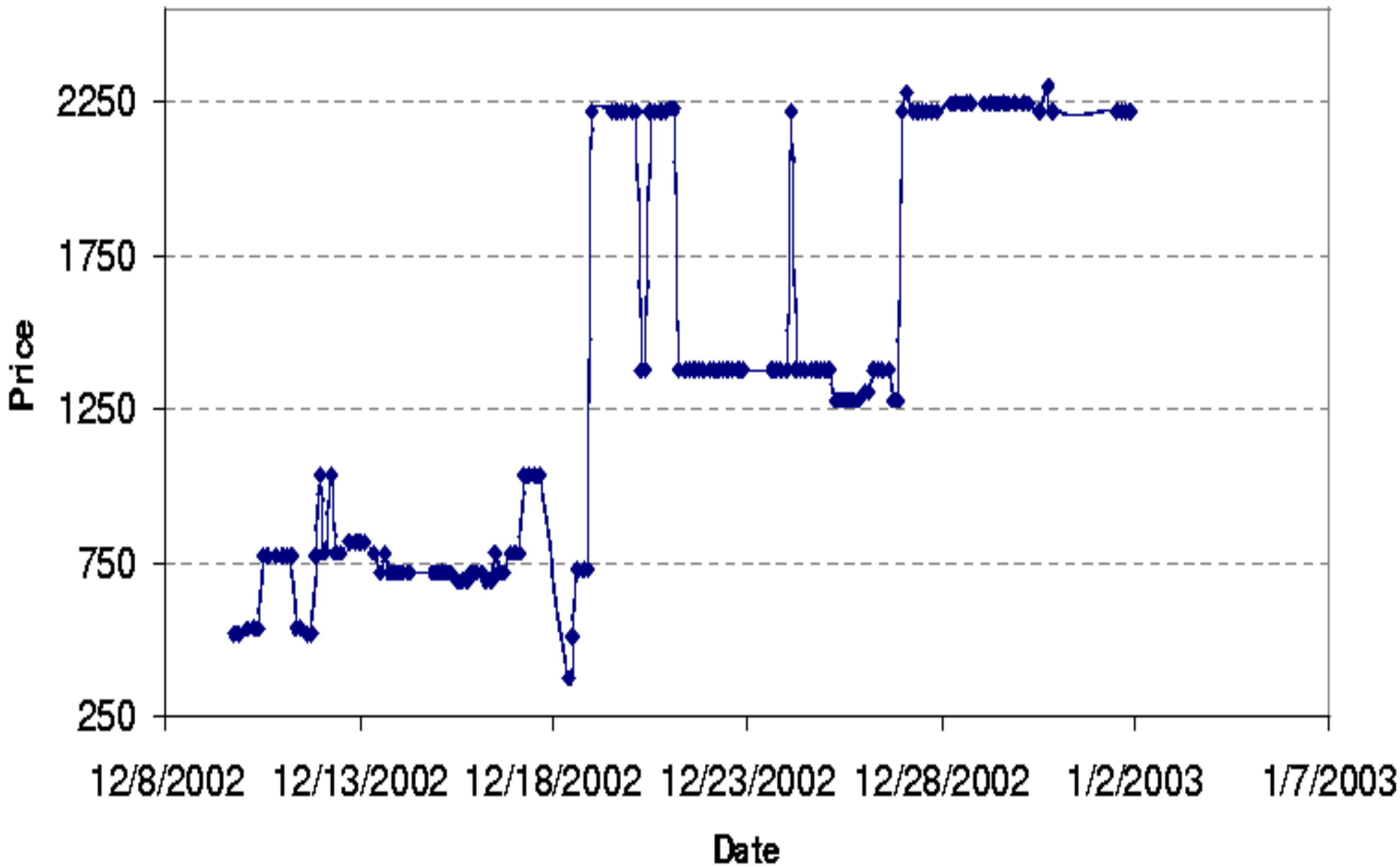
Mining Airfare Data to Minimize Ticket Purchase Price

Oren Etzioni (UW) Craig Knoblock (USC)
Alex Yates (UW) Rattapoom Tuchinda (USC)



Price change over time for American Airlines flight #192:223, LAX-BOS, departing on Jan. 2.

Consumers' Dilemma

To Buy or Not to Buy…that is the question..

Hamlet

Data mining → Price drops

Advisor Model

1. Consumer wants to buy a ticket.
2. Hamlet: 'buy' (this **is** a good price).
3. Or: 'wait' (a better price will emerge).
4. Notify consumer when price drops.

Arbitrage Model

1. “going price” is $900.
2. Hamlet anticipates a price of $400.
3. Hamlet offers a $600 fare.
4. Hamlet buys when the price drops to $400.
5. Consumer saves $300; Hamlet earns $200.

(of course, Hamlet could lose money!)

Will Flights sell out?

1. Watch the number of empty seats.
2. Upgrade to business class.
3. Place on another flight and give a free ticket.

In our experiment: upgrades were sufficient.

Is Airfare Prediction Possible???

- Complex "yield management" algorithms.
  - airlines have tons of historical data.
- Exogenous events create randomness.

How about the stock market?

- True markets *are* unpredictable.
- For Hamlet, prices are **set** by the airlines!

Surprising Experimental Result

Savings: buy immediately versus Hamlet.

Optimal: buy at the best possible time.

HAMLET's savings were 61.8% of optimal!

Though it be madness, yet there be method in it.

Data Set

- Used Fetch.com’s data collection infrastructure.
- Collected over 12,000 price observations:
  - Lowest available fare for a one-week roundtrip.
  - LAX-BOS and SEA-IAD.
  - 6 airlines including American, United, etc.
  - 21 days before each flight, every 3 hours.

Learning Task Formulation

Input: price observation data.

Algorithm: label observations (decision point);
run learner.

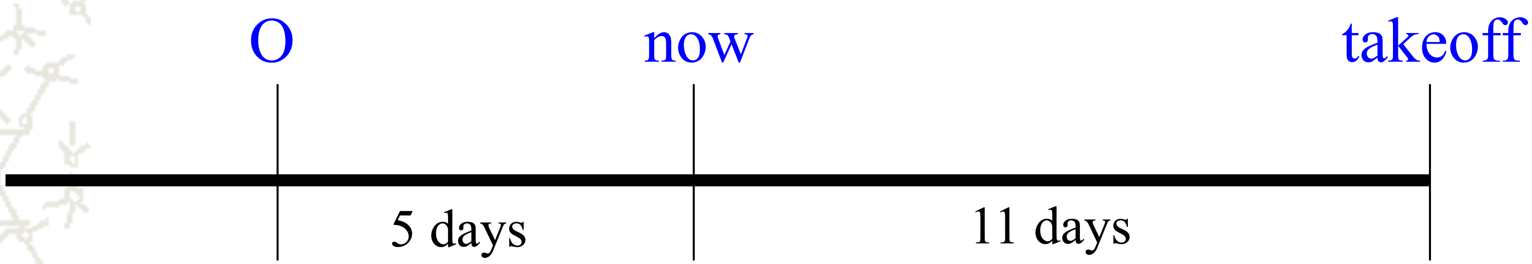
Output: Classify each decision point →
buy versus wait.

Formulation Fine Points

- Want to learn from the latest data.
- Run learner nightly to produce a new model.
  – Learner is trained on data gathered to date.
- Learned policy is a sequence of 21 models.

- Test set: 8 * 21 decision points for the last 1/3 of the flights.

Labeling Training Data



IF price drops between and now THEN label(O)=wait
ELSE label(O) → Pr(price will drop between now and takeoff)

We estimate Pr based on behavior of past flights.

Candidate Approaches

- Fixed: “asap”, 14 days prior, 7 days,…
- By hand: an expert looks at the data.
- Time series: $P_t = F(P_{t-1}, P_{t-2}, \ldots P_1)$.
  - Not effective at price jumps!
- Reinforcement learning: Q-learning.
  - Used in computational finance.
- Rule learning: Ripper, …

Ripper

- Features include price, airline, route, hours-before-takeoff, etc.
- Learned 20-30 rules…

IF hours-before-takeoff ≥ 252 AND price ≥ 2223 AND route = LAX-BOS THEN *wait*.

Simple Time Series

- Predict price using a fixed window of k price observations weighted by α.
- We used a linearly increasing function for α

$$p_{t+1} = \frac{\sum_{i=1}^{k} \alpha(i)\, p_{t-k+i}}{\sum_{i=1}^{k} \alpha(i)}$$

Q-learning

Natural fit to problem

$$Q(a,s) = R(a,s) + \gamma \cdot \max_{a'}\left(Q(a',s')\right)$$

$$Q(b,s) = -price(s)$$

$$Q(w,s) = \begin{cases} -300000 & \text{if flight sells out after } s. \\ \max\left(Q(b,s'), Q(w,s')\right) & \text{otherwise.} \end{cases}$$

Hamlet

- Stacking with three base learners:
  1. Ripper (e.g., R=wait)
  2. Time series
  3. Q-learning (e.g., Q=buy)
- Ripper used as the meta-level learner.
- Output: classifies each decision point as

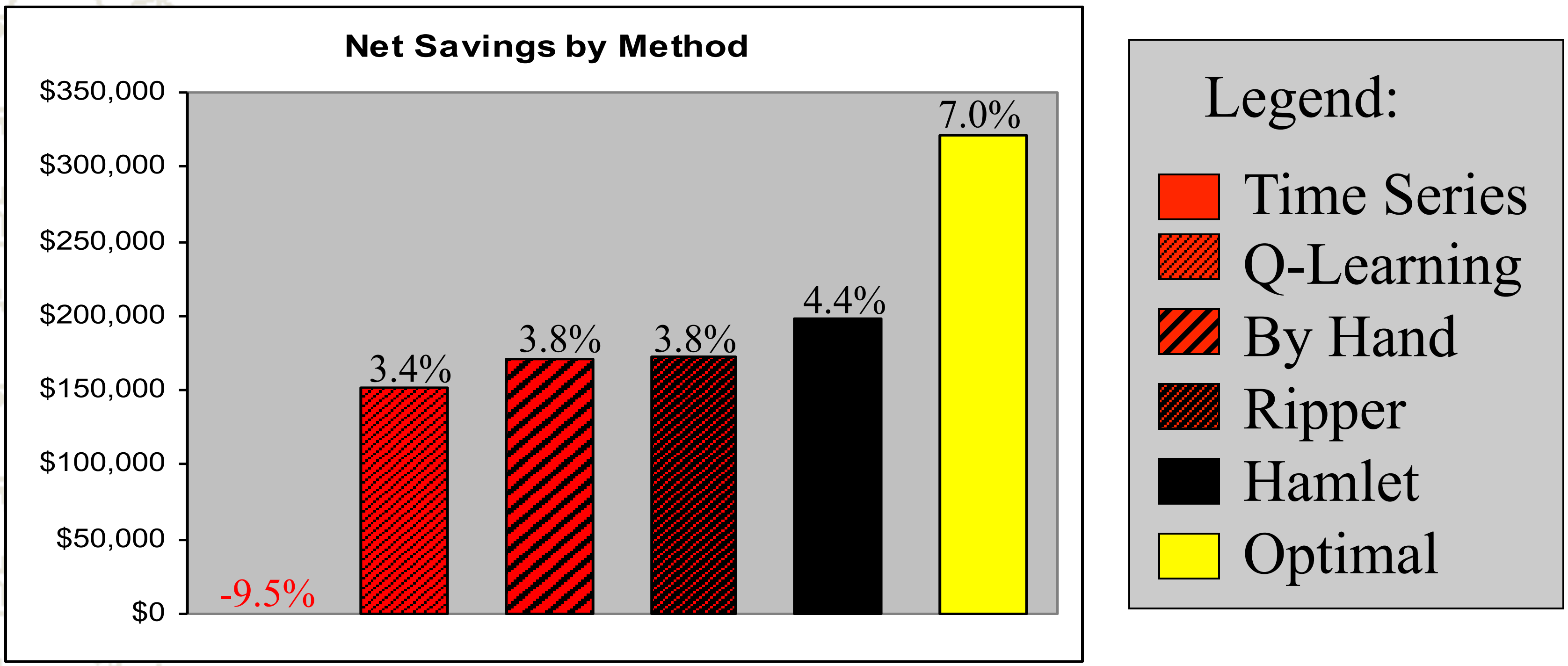
  'buy' or 'wait'.

Experimental Results

- Real price data; Simulated passengers.
  - Uniform distribution over decision points. (sensitivity) Requesting specific flights (also 3hr interval).
- Learner run once per day on "past data".
- Execution: label each purchase point until buy (or sell out).
- Compute savings (or loss).

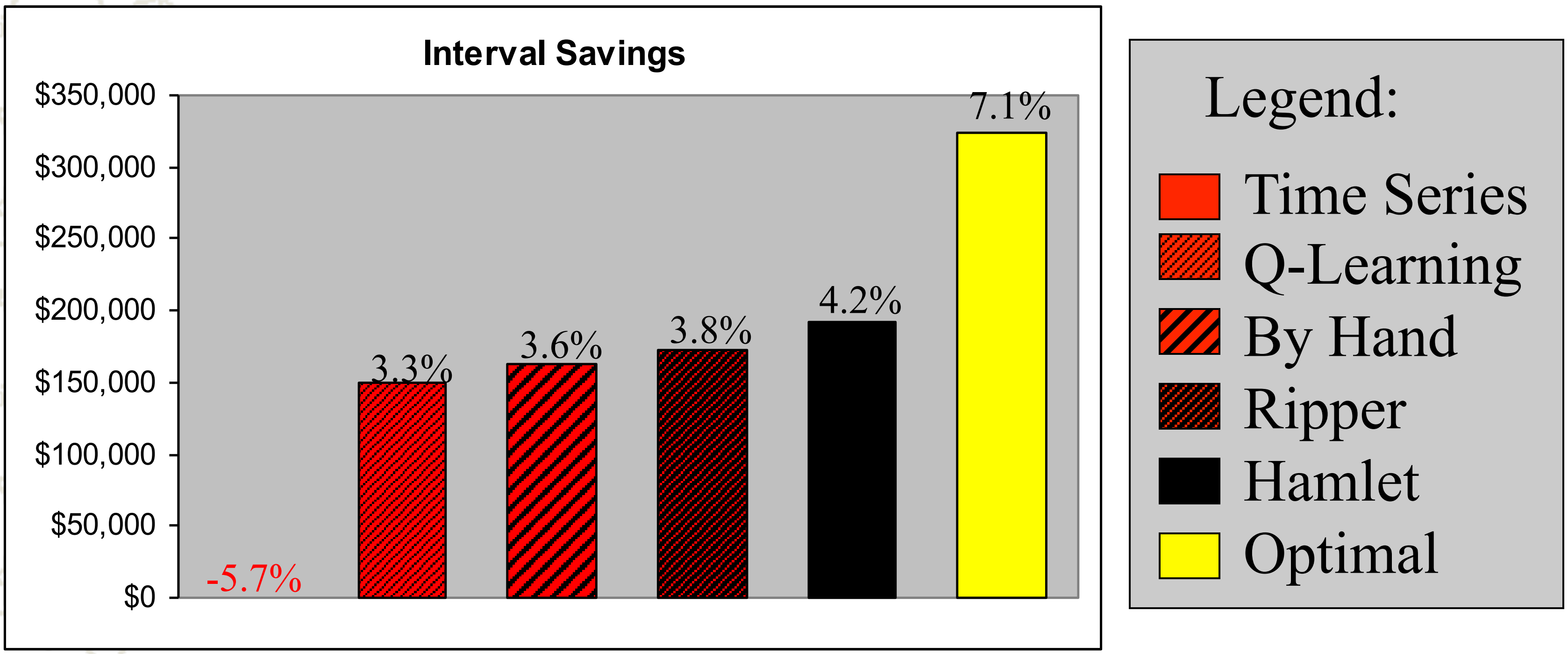
Savings by Method

- **Net** savings = cost now – cost at purchase point.
- Penalty for sell out = upgrade cost. 0.42% of the time.
- Total ticket cost is $4,579,600.



Sensitivity Analysis

- Passenger requests any nonstop flight in a 3 hour interval:



Upgrade Penalty

Method	Upgrade Cost	% Upgrades
Optimal	$0	0%
By hand	$22,472	0.36%
Ripper	$33,340	0.45%
Time Series	$693,105	33.00%
Q-learning	$29,444	0.49%
Hamlet	**$38,743**	**0.42%**

Discussion

- 76% of the time --- no savings possible.
- Uniform distribution over 21 days.
- 33% of the passengers arrived in the last week.
- No passengers arrived >21 days before.

Simulation understates possible savings!

Savings on “Feasible” Flights

Method	Net Savings
Optimal	30.6%
By hand	21.8%
Ripper	20.1%
Time Series	25.8%
Q-learning	21.8%
Hamlet	**23.8%**

Comparison of Net Savings (as a percent of total ticket price) on Feasible Flights

Related Work

- Trading agent competition.
  - Auction strategies
- Temporal data mining.
- Time Series.
- Computational finance.

Future Work

- More tests: international, multi-leg, hotels, etc.
- Cost sensitive learning (tried MetaCost).
- Additional base learners
- Bagging/boosting
- Refined predictions
- Commercialization: patent, license.

Conclusions

1. Dynamic pricing is prevalent.
2. Price mining a-la-Hamlet is feasible.
3. Price drops can be surprisingly predictable.
4. Need additional studies and algorithms.
5. Great potential to help consumers!

All's well that ends well.

Savings by Method

- Savings over "buy now".
- Penalty for sell out = upgrade cost.
- Total ticket cost is $4,579,600.

Method	Savings	Losses	Upgrade Cost	% Upgrades	Net Savings	% Savings	% of Optimal
Optimal	$320,572	$0	$0	0%	$320,572	7.0%	100.0%
By hand	$228,318	$35,329	$22,472	0.36%	$170,517	3.8%	53.2%
Ripper	$211,031	$4,689	$33,340	0.45%	$173,002	3.8%	54.0%
Time Series	$269,879	$6,138	$693,105	33.00%	-$429,364	-9.5%	-134.0%
Q-learning	$228,663	$46,873	$29,444	0.49%	$152,364	3.4%	47.5%
Hamlet	**$244,868**	**$8,051**	**$38,743**	**0.42%**	**$198,074**	**4.4%**	**61.8%**

Sensitivity Analysis

- Passenger requests any nonstop flight in a 3 hour interval:

Method	Net Savings	% of Optimal	% upgrades
Optimal	$323,802	100.0%	0.0%
By hand	$163,523	55.5%	0.0%
Ripper	$173,234	53.5%	0.0%
Time Series	-$262,749	-81.1%	6.3%
Q-Learning	$149,587	46.2%	0.2%
Hamlet	**$191,647**	**59.2%**	**0.1%**

Another Chart