Forecasting Airfare Prices: Evidence from an Online Travel Agency

George Dounias[1] Vasilakis Chrysovalantis[2]

[1]University of Aegean, Financial Engineering Department
[2]Bangor Business School University, IZA

July 9, 2025

Outline

Motivation

- Tourism powers the global economy, supporting 319 million jobs and 10.4% of GDP in 2018 (Scott, 2019).
- Airfare prices swing unpredictably, making it hard to plan affordable trips (Shin et al., 2022).
- Travelers want tools to find the best booking times and save money.
- Online Travel Agencies (OTAs) can stand out by offering accurate price forecasts.
- Our goal: Build a forecasting tool to help travelers book smarter and boost OTA success.

Literature Review

Pricing: Dynamic pricing (EMSR) (Belobaba, 1989); prices peak last-minute (Etzioni et al., 2003).

LCCs stabler than HCCs (Etzioni et al., 2003); price dispersion from competition (Varian, 1980).

Forecasting: LR, time-series (Goolsbee and Syverson, 2008); ANNs, SVMs for non-linearities (Chen et al., 2015).

Genetic algorithms in energy/finance (Gonzalez, 2015); hybrid models improve accuracy (Jin et al., 2020).

Gap: Limited high-frequency, customer-level airfare forecasting.

Paper Overview

Objective: Best method for airfare price forecasting (2011–2014 OTA data).

Data: Customer searches for 7 routes (e.g., MOW-KRR, ATH-AMS).

Methods: Linear Regression (LR), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Hybrid Genetic Algorithm (HGA).

Contributions: HGA outperforms; simulation shows €15.23/ticket savings.

Implications: OTA tool to enhance customer decisions, competitiveness.

Data Description: Source

Source: Anonymized European Online Travel Agency (OTA)[1].

Period: June 2011–December 2014.

Scope: 21 round-trip routes (7 analyzed: MOW-KRR, ATH-AMS, HER-SKG, LCA-SOF, RHO-ATH, LCA-SKG, LCA-ATH).

Flight Types: Direct and stop-over, all fare classes (economy, business, first).

Constraints: No multi-leg trips, max 8 passengers.

[1] OTA refers to web-based platforms where customers search and book flights, hotels, and other travel services.

Data Description: Variables

Search Parameters (10):

- Search date/time (day, hour: morning, afternoon, evening).
- Departure/return dates, travel month.
- Route, airline type (LCC/HCC), number of passengers (1–8).
- Seat type (economy, business, first).

Output Parameters (4):

- Proposed ticket price (in euros).
- Flight operator, direct/stop-over status, seat availability.

Data Description: Insights

Booking Timing: Half of bookings occur within 30 days of travel; 90% within 100 days, with more bookings closer to departure.

Passengers: 75% of searches are for one traveler, 18.65% for two, suggesting mostly solo or couple trips.

Flight Preferences:

- 97.95% of customers are open to direct or connecting flights.
- 58% don’t specify a class; 42% choose Economy, reflecting budget-conscious travelers.

Price Trends: Evening searches (e.g., after 6 PM) and shorter trips (e.g., 3–7 days) often yield lower prices.

Forecasting Methods: Overview

Linear Regression (LR): Simple method to predict prices using basic patterns like route or travel time.

Artificial Neural Networks (ANN): Learns complex price changes, like sudden drops, with fine-tuned settings.

Support Vector Machines (SVM): Flexible method to predict price trends accurately.

Hybrid Genetic Algorithm (HGA): Combines neural networks and smart optimization to find the best price predictions.

Evaluation: Use MAPE (average % error) and RMSE (average error size) to compare accuracy.

Forecasting Methods: LR, ANN, SVM

LR: Predicts prices using a simple formula based on route, time, and other factors (Bertrand et al., 2004).

ANN:

- Uses 4 layers to detect complex price patterns, like seasonal spikes.
- Fine-tuned with 20–75 units per layer for better accuracy.
- Trained on 78% of data, tested on 22% to ensure reliability.

SVM: Predicts prices with high accuracy; allows small errors for stability (Chen et al., 2015).

Forecasting Methods: HGA

Combines neural networks with a smart search to predict prices (Tsakonas et al., 2025).

Steps:

- Start with many possible prediction models.
- Test and improve models through repeated trials, keeping the best ones.
- Use five separate groups (islands) to explore different solutions.
- Share top models between groups to find the best overall prediction.

Achieves 85% accuracy, excels at handling complex price changes (Whitley et al., 1998).

Results: Model Comparison

Table: Average MAPE and RMSE Across 7 Routes

Method	MAPE (%)	RMSE (%)
HGA	**11.57**	**12.77**
SVM	13.99	13.65
ANN	14.77	13.27
LR	18.69	17.65

HGA leads with 11.57% error, 2–7% lower than others.

Excels for low-cost carriers and stable routes (e.g., HER-SKG).

Enables savings of €15–€22 per ticket via accurate forecasts.

Results: Temporal Patterns

Days: Tuesday (10.30%) and Thursday (10.92%) most predictable.

Months: January (9.89%), November (10.17%), August (10.53%) lowest errors; June (17.07%), December (19.30%) highest.

Airlines: LCCs (9.37%) more predictable than HCCs (11.57%).

Evening searches, shorter trips correlate with lower prices.

Customer Utility Analysis: Why Simulate?

Purpose: Show how HGA forecasts save customers money.

Airfare prices fluctuate, making timing critical for savings.

Simulation tests if HGA-guided booking beats naive approaches.

Benefits customers (lower costs) and OTAs (more bookings).

Supports forecasting as a practical tool (Jin et al., 2020).

Customer Utility Analysis: How It Works

Data: Anonymized European OTA, June 2011–December 2014.

Setup: Simulate 1,000 bookings per route (e.g., MOW-KRR, HER-SKG).

Choices: Random route, month, day (e.g., Tuesday), time (e.g., evening), days to departure (1–100).

Prices: Average per route (e.g., €150 for MOW-KRR) with 20% variation.

Strategies:

1. **HGA-Guided**: Book on lowest predicted price day in 7 days (e.g., 12.20% error for MOW-KRR).
2. **Last-Year**: Use last year's price (20% error).
3. **Random**: Book any day, no prediction.

Customer Utility Analysis: Savings

Table: Average Savings per Ticket with HGA

Route	vs. Last-Year (€)	vs. Random (€)
MOW-KRR	14.56	21.78
ATH-AMS	12.34	18.76
HER-SKG	18.56	27.32
LCA-SOF	13.89	20.45
RHO-ATH	15.67	23.12
LCA-SKG	17.89	26.45
LCA-ATH	13.76	20.34
Average	**15.23**	**22.47**

Average savings: €15.23 vs. last-year, €22.47 vs. random.
Highest for routes like HER-SKG (low error, 6.70%).

Customer Utility Analysis: Why It Matters

Customers save €15–€22 per ticket using HGA predictions.

Best savings: Tuesdays (MAPE 10.30%) and January (MAPE 9.89%) due to high predictability.

OTAs gain loyal customers by offering a forecasting tool.

Helps travelers plan smarter, book cheaper (Tsakonas et al., 2025).

Policymakers: Forecasting promotes fair, transparent pricing.

Conclusions

HGA most accurate (MAPE 11.57%, RMSE 12.77%).

Simulation: €15.23/ticket savings with HGA-guided booking.

Tuesdays, January best; June, December volatile.

LCCs more predictable than HCCs.

OTAs can offer forecasting service to attract customers.

Future: Explore XGBoost, hybrid models.

References

Belobaba, P. P. (1989). Application of a probabilistic decision model to airline seat inventory control. *Operations Research*, 37(2), 183–197.

Bertrand, M., et al. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1), 249–275.

Bitran, G., Caldentey, R. (2003). An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, 5, 203–229.

Gupta, M., et al. (2008). Matching information security vulnerabilities to organizational security profiles. *Decision Support Systems*, 41(3), 592–603.

Chen, P. H., et al. (2015). Support vector regression. *Encyclopedia of Machine Learning*, Springer.

Dana, J. D. (1999). Equilibrium price dispersion under demand uncertainty. *RAND Journal of Economics*, 30(4), 632–660.

Etzioni, O., et al. (2003). To buy or not to buy: Mining airfare data. *Proceedings of SIGKDD '03*.

Goolsbee, A., Syverson, C. (2008). How do incumbents respond to the threat of entry? *Quarterly Journal of Economics*, 123(4),