Predicting Operational Efficiency Using Cost Metrics in the Airline Industry

Group 3

18/11/2021

Introduction

The system-wide global commercial airlines constitute a high-cost industry, with expenses of approximately 795 and 484 billion dollars globally in 2019 and 2020 (Iata, 2021). This is why the revenue margins are highly dependent on its operational efficiency, the metric that measures the efficiency of profit earned as a function of operating costs. An industry is operationally efficient if conditions exist that allow participants to execute transactions and receive services at a price that equates fairly to the actual costs required to provide them. Airlines can become more efficient through improvement in a variety of areas, from which aircraft and crew utilization, payload, and fuel consumption are key. In this way, resources and the efficient use of such resources have a key impact in predicting profitability in airline firms.

The aim of this research is thus to predict operational efficiency in the airline industry by using key cost metrics. The research uses kernels as a method that allows for non-linearity, making the predictions more flexible although probably at the cost of bias. The predictions are relevant for both managers and stakeholders. On the one hand, they are key for managers to choose relevant strategies for the consolidation of the firms. On the other hand, the information allows for a better understanding of the predictors of financial development in the industry and provides more stable financial decisions.

The paper is structured as follows. In section 2, the airline data used for prediction. In section 3, the choice of kernel functions as a method is justified and outlined. In section 3, the results…. Finally, it is concluded that… DISCUSSION PERSON, PLEASE ADD TWO RELEVANT SENTENCES HERE (ONE FOR RESULTS AND ONE FOR DISCUSSION) TO SUMMARIZE THE SECTIONS.

Data

The dataset, obtained through Greene (2003), contains empirical evidence relating to the cost of the airline industry. It consists of a panel of 6 observations of Airlines in the United States over 15 years- from 1970 to 1984, yielding 90 observations in total. Data was collected for all six firms in the same fifteen years.

Each observation consists of six variables relating to cost: airline, year, pf (fuel price), lf (load factor, which is the average capacity utilization of the fleet), output (revenue, measured in revenue per passenger-mile flown), and cost (total cost per \$1,000). Fuel price, load factor, output, and cost are numerical variables. Airline is treated as an indicator variable that classifies each airline in a distinct category. We control for the year of observation; however, this is done by using a continuous variable instead of individual dummies, as this allows the prediction of future years. All summary statistics are shown in Table **??**.

All observations are used for the study. All variables except the airline indicator variable are scaled to create z-scores $(X - \mu)/\sigma$ where μ and σ represent the mean and the standard deviation, respectively. Scaling ensures that the variables are comparable, as our approach relies on the distance between observations.

Table 1: Summary Statistics

Variable	Count	Mean	Std	Min	25%	50%	75%	Max
Revenue per Passenger-Mile Flown (in USD)	90	0.54	0.53	0.04	0.14	0.31	0.95	1.94
Total Cost (in million USD)	90	1123	1192	69	292	637	1346	4748
Fuel Price	90	472	330	104	130	357	850	1016
Load Factor	90	56.0%	5.3%	43.2%	52.9%	56.6%	59.5%	67.6%

Methodology

In the aim to increase the flexibility of the model- to decrease variance at the cost of bias-, we move beyond the standard linear regression model and use basis expansion models, popular in machine learning for their ability to extract more predictive information from the data set. In particular, kernels are used as a method that allows the introduction of non-linear components to linear regression, thus maintaining the convenience and ability of linear methods to not overfit. Kernels do so by embedding the points of any space into a new space where linearity can be applied via a non-linear feature map. The embedding computation is not made directly but through the kernel functions, built to be equivalent to the scalar products of the transformation (Hastie, Tibshirani, Friedman, 2008). In the analysis, polynomial kernels, radial basis function kernels and polyharmonic spline kernels are used.

Specifically, let $X \to R$ be the space that contains the data points $x_1, ..., x_n \in X$, and $\Phi : X \to R^d$ a feature map where Φ denotes the $n \times d$ matrix containing data points $\Phi(x_i)$ as rows. Then, a symmetric function $k : X \times X \to R^d$ is a kernel function if for all $n \geq 1$ in X and $c_1, ..., c_n \in R^d$, it holds that:

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0. \tag{1}$$

The kernel function k must be a symmetric and positive semi-definite function. The kernel function has a corresponding kernel matrix k with entries $k_{ij} = k(x_i, x_j)$. By definition, a valid kernel matrix coincides with matrix $\Phi \cdot \Phi^T \in R^d$, having entries:

$$k_{ij} =< \Phi(x_i), \Phi(x_j) >= \Phi(x_i)\Phi(x_j)^T \text{ for all } x, y \text{ in } X. \tag{2}$$

Given X and k, $\mathcal{H}$ denotes a Hilbert space with scalar product $< \cdot, \cdot >_{\mathcal{H}}$. The Reproducing Kernel Hilbert Space (RKHS) is the vector space $\mathcal{H}$ in which equality (2) holds.

The aim of the kernel method is to solve the general regression problem as a problem of finding the function f that minimizes:

$$\min_{f \in \mathcal{H}} H[f] = \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i)) + \lambda ||f||_K^2, \tag{3}$$

where $||f||_K^2$ is a norm in the RKFH $\mathcal{H}$ defined by the positive definite function K, n is the number of data points and λ is the regularization parameter (Evgeniou, Pontil, & Poggio, 1999, p. 2). The term $\lambda||f||_K^2$ is built as a quadratic penalty factor based on the Ridge regression, which imposes the penalty on the size of the regression coefficients. This reduces the high dimensionality and high correlation of the regressors, that are problematic for conventional techniques such as the Ordinary Least Square optimization.

The introduction of kernels to the optimization problem (3) allows us to solve it purely in terms of kernel functions:

$$\underset{\alpha \in R^n}{maximize} \sum_{i=i}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j), \tag{4}$$

subject to:

$$0 \leq \alpha_i \leq \frac{c}{n} \forall_i = 1, ..., n,$$

$$\sum_{i=i}^{n} \alpha_i Y_i = 0.$$

The solution to the optimization problem can be found with the function k, $(X \times X)$ with $k(x_i, y_i) =< \Phi(x_i), \Phi(x_j) >$.

Kernels not only allow us to find the minimization functional, but are able to do so in a finite-dimensional subspace even though its respective RKHS is an infinite dimensional vector space (Vapnik, 1995). This is because $\mathcal{H}$ can be subdivided in $\mathcal{H}_{data}$, the *span* $\{k_{xi}, ..., k_{xn}\}$, and $\mathcal{H}_{comp}$, its orthogonal complement. Each vector $w \in \mathcal{H}$ is $w = w_{data} + w_{comp}$. The prediction of all functions with the same w_{data} agree on all training points, and as such is not affected by w_{comp}. As such, the norm w is smallest if $w_{comp} = 0$, thus always being an optimal solution. In this way, kernels allows us to combat the computational difficulties related to very large number of parameters.

In this analysis, three kernels are computed: the radial basis function (RBF) kernel, the inhomogeneous polynomial kernel (IPK), and the polyhomogeneous spline kernel (PSK). Different kernels are used to compare their ability to predict the data. Their functions are:

$$\text{RBF: } k_{ii'} = e^{(-\gamma||x_i - x_i'||^2)}, \text{ with hyperparameter} \gamma > 0. \tag{5}$$

$$\text{IPK: } k_{ii'} = (1 + x_i^T x_i')^d, \text{ with fixed degree} d > 0. \tag{6}$$

$$\text{PSK:} k_{ii'} = r^d, \text{ if } d = 1, 3, 5, ..., \tag{7}$$
$$k_{ii'} = r^d ln(r), \text{ if } d = 2, 4, 6, ...$$

Intuitively, RDB captures the similarity at point y to a fixed point x as a function of their distance. IPK considers combinations of the features of input samples (interaction features) to determine similarity. PSK ... ADD ONE SENTENCE TO DESCRIBE INTUITIVELY WHAT IT DOES.

Diagnostics

In both the cross-validation and the test set our predictions will be tested using the mean-squared prediction error ($MSPE$). $MSPE = \frac{1}{n}\sum_{i=n}^{n+q}(y_i - \hat{y}_i)^2$, where y_i is an output variable, and $\hat{y}_i$ is the predicted value over that period. In this example, n is the number of observations in the training set, q the number of observations in the test set. In each case a lower MSPE will be considered better predictive performance.

Results

Discussion

References

Evgeniou, T., Pontil, M., & Poggio, T. (1999). Regularization Netowrks and Support Vector Machines. Advances In Computational Mathematics, 13, 1-50. Retrieved 22 November 2021, from.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer.

IATA Economics. (2021). Industry Statistics. IATA.

Vapnik, V. (1995). The Nature of Statistical Learning Theory (2nd ed.). Springer.

Appendix I: Code