E-Companion for the article: "Modeling Crew Itineraries and Delays in the National Air Transportation System"

Keji Wei, Vikrant Vaze

Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire 03755

{keji.wei.th@dartmouth.edu, vikrant.s.vaze@dartmouth.edu}

February 22, 2018

EC.1 Local Search Heuristic for the Calibration Problem

We describe the local search heuristic used to solve our calibration problem. It is as follows.

INITIALIZE:

Set all parameters to zero, i.e., $\alpha_i = \beta_i = 0, i \in \{1,2,3,4\}$ and $\gamma_i = 0, i \in \{5,6\}$.

LOOP:

FOR i=1:4

Perform local grid-search by varying (α_i, β_i) values to minimize $\sum_{i=1}^{6} \left| F^{\hat{x}}(i) - F^{x}(i) \right|$. Update (α_i, β_i) values.

END FOR

FOR i=5:6

Perform local line-search by varying γ_i values to minimize $\sum_{i=1}^{6} \left| F^{\hat{x}}(i) - F^{x}(i) \right|$. Update γ_i values.

END FOR

IF no parameter values got updated in the last iteration of the outer LOOP then EXIT.

EC.2 Data Preprocessing Steps

Preprocessing consisted of two major steps. The first step was to get aircraft tail numbers for all flights in the crew scheduling samples by matching each flight in the sample with exactly one flight in the AOTP database. This is performed by matching the departure airport, arrival airport, scheduled departure time, scheduled arrival time and the airline code. We are able to match around 95% of all flights in the airline crew scheduling samples.

The second preprocessing step was data filtering to account for the limitations of the AOTP database. Because tail number information is missing for some flights in the AOTP, we use, as input to our models, only those flights for which the tail number is present in the AOTP database. Since only domestic flights information is provided in the ATOP, we removed all international flights from our crew scheduling sample as well. Typically, cockpit crews are assigned to operate aircraft belonging to only one fleet family within a given pairing. Indeed, almost all the crew pairings in our confidential crew scheduling data contained flights operated by a single fleet family. We eliminated the few crew pairings (and their corresponding flights) which cut across multiple fleet families in our crew schedule data. As a result, the crew pairing problem can be considered separately for each fleet family. For the regional carrier, we used the first week of March 2014 as our calibration dataset and the first week of April 2014 as the validation dataset. For the network legacy carrier, we used the first week of January 2014 as our calibration dataset and the first week of one month in each quarter, namely, February 2014, April 2014, July 2014, and October 2013, as our validation datasets. We also eliminated all crew pairings (and all flights in those crew pairings) such that at least one flight in that crew pairing was already removed for any of the reasons mentioned above. Overall, this resulted in the removal of approximately 15-20% of all crew pairings and approximately 10-15% of all flights in our network legacy carrier crew schedule sample across different time periods in the sample. Also, it resulted in the removal of approximately 10-15% of all our crew pairings and

approximately 10-15% of all flights in our regional carrier crew schedule sample across different time periods in the sample.