IMPACT OF OPERATIONAL PERFORMANCE ON AIR CARRIER COST STRUCTURE: EVIDENCE FROM US AIRLINES

Bo Zou, Mark Hansen

National Center of Excellence for Aviation Operations Research,
Institute of Transportation Studies, University of California at Berkeley, USA

ABSTRACT

The impact of operational performance on airline cost structure is empirically investigated using an aggregate, statistic cost estimation approach. We extend the conventional delay characterization by developing two distinct sets of operational performance metrics: one that considers delay against schedule and schedule buffer; the second based on the relationship between scheduled and actual time periods when flights are active. The two metrics sets provide different perspectives on the flight delay phenomena, and are incorporated into the airline cost models as arguments. The results from estimating a variety of airline cost models reveal that, both delay and schedule buffer are important cost drivers; flight activity outside schedule windows increases cost, whereas flight inactivity within schedule windows does not. Using the estimated cost models, we predict the cost savings to airlines of "perfect" operational performance, obtaining an estimate of about $10 billion for 2007.

INTRODUCTION

Flight delay is a serious problem that has garnered increasing attention in the United States. In 2007, nearly one in four US airline flights arrived at its destination over 15 minutes late (BTS, 2009). About a third of these late arrivals were a direct result of the inability of the aviation system to handle the traffic demands that were placed upon it, while another third resulted from airline internal problems. Most of the remainder was caused by an aircraft arriving late and thus having to depart late on its next flight (BTS, 2009). Between 2002 and 2007, as the air transport system recovered from the 9/11 attacks, scheduled airline flights increased about 22 percent, but the number of late-arriving flights more than doubled. Since 2007, traffic and delays have declined somewhat because of the recession, but the Federal Aviation Administration (FAA) expects growth to resume, with air carrier flight traffic reaching 2007 levels by 2012, and growing an additional 30 percent by 2025.

Substantial investments are required in order to modernize and expand the aviation infrastructure so that it can accommodate anticipated growth without large increases in delay. In the US, the Next Generation Air Transportation System (NextGen) will deploy improved systems for communications, surveillance, navigation, and air traffic management, and require flight operators to invest in new on-board equipment. The NextGen development and capital costs are estimated to be $30-40 billion, about half for the infrastructure and half for aircraft equipage through 2025. Substantial improvements in air transportation capacity also require airport infrastructure enhancement. Over the next five years alone, airports report a need for $19 billion in airport capacity expansion investments. If airport capacity investment continued at that rate, then combined NEXTGEN and airport capacity investment could reach $100 billion over the next 15 years (Hansen, 2010).

To justify the huge investment, an understanding of the return to the investment is of critical importance. It is well recognized that, much of the benefits from NEXTGEN and airport capacity investments will take the form of increased capacity and reduced delay. The business case for these large expenditures rests largely on the value of reducing delay and its associated costs. The cost of delay includes many elements. Passengers are inconvenienced when their flight arrives late, diminishing their willingness to pay for the air travel product and discouraging some from flying altogether. Adaptations to avoid or mitigate delay—such as leaving early for a business meeting to make sure it is not missed, or scheduling flights at less ideal times to avoid congestion—also entail costs. However, the element that receives the most attention is that incurred directly by airlines through increased operating expenses. Many in the community perceive that, since airlines must pay these costs with “real money”, they merit stronger consideration than the cost of passenger time loss.

Despite a large volume of literature on estimating delay cost to airlines, current knowledge and thinking about the impact of delay on airline cost structure remain limited. The prevailing estimation methods often involve assumptions that are rarely attested or justified. Moreover, the majority of existing studies ignore the fact that delay is measured against flight schedule, which has been manipulated by airlines in order to improve their official on-time performance statistics. Nowadays airlines routinely add extra time to the schedule, providing more leeway for flights to arrive on-time despite encountering congestion or other delays. The costs of this practice represent an integral part of the total delay cost; however, it is largely missed in existing research.

The goal of this paper is to address these concerns. We contribute to the airline delay cost literature, first, by developing alternative metrics for what we term "operational performance," a concept that is broader than but includes the traditional concept of delay against schedule. Multiple metrics sets are developed, providing different angles to look at the operational performance of airlines. We also introduce a different cost estimation methodology to quantify the impact of operational performance on airline cost structure. Using published, quarterly, airline-level data, we estimate the relationship between airline cost, output, factor prices, and other variables. Included among the latter are airline operational performance variables. Such models are capable of establishing the empirical basis for translating operational performance into monetary terms, and involve a minimum of assumptions about the mechanisms through which operational performance affects cost.

The rest of the paper is organized as follows. A critical review of current airline delay cost estimation methodology is performed in section 2. Section 3 describes different sets of metrics for characterizing airline operational performance. Section 4 presents the specification of airline cost models and the databases used. Estimation results are discussed in the ensuing section. In section 6, we apply the estimated models to assess the potential cost impact of imperfect operational performance. Section 7 offers further discussion and conclusions.

CURRENT PRACTICE OF ESTIMATING DELAY COST TO AIRLINES

Current practice of estimating the cost impact of imperfect operational performance on airlines can be classified into two approaches. The first, often called the cost factor approach, is based upon assigning unit costs to different categories of delay based on estimates of the resources consumed when a given category of delay occurs. The total cost of delay, *C*, is equal to the sum of delay cost in each category:

$$C = \sum_i P_i \cdot X_i \tag{1}$$

where P_i denotes the unit cost per minute for delay in the *i*th category, and X_i represents the corresponding total delay minutes. Equation (1) represents the general formula of the cost factor approach, which admits many possibilities for classifying delay. One is based upon the phase of flight in which delay is taken (JEC, 2008). Gate, taxi, and airborne delays are the primary categories. Alternatively, a flight delay can be distinguished according to whether it is propagated from delay on an earlier flight, or is caused by some occurrence on the delayed flight itself (ITA, 2000; Cook et al., 2004). Depending upon whether a propagated delay is caused by delay on previous legs flown by the same airframe, or delay on legs flown by a different aircraft, propagated delay can be further distinguished as rotational and nonrotational. Since the cost per delay minute probably varies by aircraft type, such differentiation is also suggested in the use of equation (1) (Cook et al., 2004).

While conceptually appealing, these possibilities are often subject to measurement uncertainties and constrained by the availability of data. If delay is decomposed into gate, taxi, and airborne components, identifying delay for each phase requires comparing the actual time spent and the nominal or unimpeded ones. The unimpeded taxi time is affected by factors such as airfield geometry and gate location, the information of which is difficult to obtain. Even the proper definition of nominal taxi time is unclear: FAA, for example, calculates nominal taxi times that allow sufficient time for a plane to wait for one aircraft ahead in the take-off queue (JEC, 2008), while others assume no interference from other aircraft in the nominal scenario. The unimpeded airborne time depends critically upon the optimal flight trajectory, winds, aircraft type, and the relative importance airlines attach to fuel and time, but such detailed information is rarely available. Distinguishing primary and propagated delay is also not intuitive unless explicit data records are available.[1] As pointed out by Cook (2009), if an aircraft arrives 30 minutes late inbound at the gate, and then leaves 45 minutes late on the next outbound leg, the portion in the 45 minutes that should be counted as delay propagation from the last leg is generally not known.

Determining cost factors rests on the assumption that delay causes additional consumption of largely the same inputs as the airlines' normal line production process. Judgement must be made about what cost components (e.g. fuel, labor, capital, airport charges) need to be included for a specific type of delay, and what are the unit cost per delay minute for each cost component. To determine these two methods are most commonly adopted, both involving some uncertainties due to the lack of detailed, e.g. flight-level, cost data. One method is to use more aggregated cost factors, e.g. cost for one-minute ground delay given an aircraft type, often based upon aircraft block hour cost data. However, estimates of this kind may be unduly influenced by accounting conventions that often have little empirical basis. A surrogate is to conduct interviews, especially when even aggregate data is not available. The value of cost factors in this case depends upon the knowledge and understanding of the subject matter experts, and therefore is inherently subjective. In effect, interviewees usually tend to incorporate cost impacts that are obvious to them while neglecting those that are not directly visible.

In addition to the issues that are tied individually to delay and cost factor, further concerns pertain to the intrinsic shortcomings of the cost factor approach. One is the assumption of linearity. The cost factor approach assumes that cost is additive function of individual events, and that the cost of each event is a linear function of the duration of delay. It is well recognized that a single four-hour delay is likely cost more than 24 ten-minute delays, for at least two reasons. First, some cost items only appear when delay exceeds a certain duration threshold (M2P Consulting, 2006). Second, the delay propagation effect increases non-linearly with the size of the initial delay (Beatty et al., 1998). Longer delays propagate to more flights, and are more likely to disrupt ground operations, gate assignments, crew schedules, and passenger itineraries to a greater extent.

[1] To our knowledge the only database that allows this distinction is the European Central Office for Delay Analysis (CODA) database.

Taking this argument one step further, the relationship between delay duration and cost can even be non-monotonic. Airlines sometimes add delays to flights in order, for example, to avoid having a flight arrive at a hub in the middle of a departure rank. The interaction of delays for many flights is particularly evident in a hub-and-spoke network in which flights are scheduled in connecting banks. If all the flights in an inbound bank are delayed by the same amount, then the effect may be far less severe than if half the flights are delayed by a larger (or even the same) amount (Hansen et al., 2001). Such effects, however, cannot be captured using the cost factor approach.

The prevalence of delay also generates sizable indirect effect that may be difficult to quantify with the cost factor approach. Carriers may take a variety of measures in flight scheduling to make their operations more robust to delay. One routine practice is padding extra minutes into the schedule. This implies the necessity of accounting for the buffer side when evaluating the whole delay cost impact. Airlines may also require spare aircraft, flight crew, and ground personnel, and load flights with more fuel. These adaptations entail overhead costs that are not accounted in determining cost factors. The cost of delay may therefore permeate throughout the entire cost structure of the airline in ways that are not tied to individual delay events. It would be extremely difficult to represent such effects using cost factors.

The weaknesses in the cost-factor method suggest alternative approaches should be considered. The second avenue, which we term the aggregate cost approach, is built upon firm or industry level relationships between total operating cost and delay. One simple version of this approach assumes that airline operating costs are proportional to the total aircraft operating time, and estimate delay cost as the fraction of total aircraft operating time that results from delay, multiplied by the total airline operating cost. This avoids the difficult (if done carefully) task of determining cost factors, and only requires straightforward calculation of the aggregate delay time and determination of the total operating cost. The approach is also likely to account for the overhead cost, as a consequence producing higher delay cost estimates than the cost factor approach (JEC, 2008).

Nonetheless, the validity of this approach is very questionable. First, the underlying assumption is still a linear relationship between delay cost and time. Second, the approach ignores differences between delay and total aircraft hours; for example, a much larger proportion of delay time occurs on the ground. Moreover, the total operating cost includes not only fuel, crew salaries, maintenance, and depreciation, but also advertising, ticket agents, landing fees, legal fees, and other items that may be relatively insensitive to delays. Therefore, cost numbers generated by this approach are almost certainly too high.

This paper focuses on a second version of the aggregate approach, which rests on the idea that airlines providing air transportation service to passengers is a production process. By incorporating operational performance variables into airline cost models and statistically observing how these variables influence airline expenses, we establish a direct empirical basis for translating airline operational performance into monetary terms. Because relationships are derived from observed co-variation between performance variables and cost using flexible functional forms, the results entails a minimum of assumptions about the delay-cost interaction mechanisms involved. We extend the current delay-centric characterization of airline operational performance by explicitly considering schedule buffers, and developing an alternative, novel set of performance metrics. Use of this alternative scheme for representing operational performance allows for an alternative estimate of total delay cost, and therefore an assessment of the sensitivity of the cost estimate to the operational performance metrics used in the cost models.

CHARACTERIZING AIRLINE OPERATIONAL PERFORMANCE

We employ the Bureau of Transportation Statistics (BTS) Airline On-time Performance database to characterize airline operational performance. This database contains detailed activity information for each non-stop scheduled-service flight by major US air carriers between points within the United States.[2] These individual flight records are used to construct various flight-level operational performance metrics, which are then aggregated for the use of airline cost models. In this study the aggregation level is by airline-quarter, a level that is consistent with the minimum observation unit for other operational and cost information.

To characterize airline operational performance two sets of metrics are developed. The first set, which we term “delay-buffer” metrics, considers delay against schedule and schedule buffer. Both of these represent the excess travel time that would not exist in an operationally perfect air transportation system. While buffer is predetermined for a particular flight, delay against schedule varies from day to day and flight to flight, and depends on the amount of buffer in the schedule. Many metrics may be used to characterize the delay against schedule. Common ones include: departure delay, positive departure delay,[3] arrival delay, positive arrival delay, and on-time performance—the fraction of flights whose delay is less than 15 minutes. Other potentially relevant metrics include standard deviation of delay and average negative delay, which both reflect degree of variability in flight operations. We performed factor analysis on these performance metrics using the 52 quarters extending from the spring of 1995 through the winter of 2007 of nine US major air carriers.[4] Appendix A provides the details. We observe similar variation patterns for these delay metrics. Exceptions are metrics describing negative delays, and, consequently, metrics based on both positive and negative delays. The results further indicate that high delays are often accompanied by high delay fluctuations (i.e. standard deviation). It may therefore suffice to use one representative metric to capture the delay against schedule side. We follow the convention of focusing on the arrival end, and choose average positive arrival delay per airline-quarter as the representative metric. This metric is preferable to average arrival delay, in that early arrivals are widely believed to save little in the way of resources as compared to on-time flights, and therefore have little cost impact.

[2] Currently, Airline on-time data are reported each month to the US Department of Transportation and BTS by the 16 U.S. air carriers that have at least 1 percent of total domestic scheduled-service passenger revenues, plus two other carriers that report voluntarily.

[3] Positive departure delay measures the difference between the actual and scheduled departure time, truncated so that delay of early departures are counted as zero. The positive arrival delay likewise.

[4] The 9 US major air carriers included in the study are: American (AA), Alaska (AS), Continental (CO), Delta (DL), American West (HP), Northwestern (NW), United (UA), US Airways (US), and Southwest (WN).

Schedule buffer refers to the extra time padded to the unimpeded flight time in a flight schedule. Unfortunately, defining the unimpeded flight time—which in reality varies from day-to-day and flight-to-flight—for purposes of measuring schedule padding is difficult, and there is little consensus on how to do this. ITA (2000) treats a certain portion of a flight's schedule time as buffer. Mayer and Sinai (2003a) define their delay as the increase in travel time relative to the minimum flight feasible flight time on a route; JEC (2008) calculates delay against the 5th percentile of all observed flight times for a given segment and month (JEC, 2008). In this study, buffer for each flight is defined as the difference between the scheduled flight time and the 5th, 10th, and 20th percentiles of all observed travel time, for a given segment (directional), airline, and quarter. We distinguish by directions to allow for the effects of prevailing winds, and by airlines to account for potential aircraft/equipment difference across carriers. We consider these percentile flight times as useful benchmarks for what travel time would be if airport/airspace were sufficiently uncongested. Netting out these unimpeded times by quarter controls for possible changes over time in the types of routes flown or in the performance of the air traffic control system that could affect average flying times (Mayer and Sinai, 2003a). Not choosing the minimum travel time makes the calculation more robust to measurement error, and reduces the influence of unusually favorable conditions, such as strong tailwinds. The average buffer is obtained by averaging the calculated buffer time over all flights for each airline and quarter. Having multiple buffer measures enables the sensitivity of cost models to the measures employed to be assessed and allows cross-validation of results.

One might argue that cancellation represents another dimension of operational performance. We do not include it into our cost model, however, for several reasons. First, cancellation accounts for a very small portion—usually less than four percent—of total flight operations. Second, cancellation is closely intertwined with delay and schedule buffer. As pointed out by Mayer and Sinai (2003b), cancellation reflects the volatility of flight time and airlines' operational strategies for coping with this volatility. Third, as found by Xiong (2010), cancellation behaviour varies significantly across airlines, reflecting different corporate policies on the trade-off between avoiding passenger and downstream disruption, maintaining schedule integrity, and "sacrificing" less important flights to reduce delays of more important ones. Finally, many delays are a consequence of airline internal factors rather than congestion and delay in the National Airspace System. Thus, while earlier studies interpret cancellation as an indicator of highly degraded conditions in which there are many flights with long delays (Hansen et al., 2000; 2001), we conjecture that such effects are better incorporated by the delay and buffer metrics. Including cancellation as an additional metric could compromise our ability to discern the impact of these variables, while also confounding our results.

The aggregate behavior of delay and buffer for the nine carriers over time is illustrated in figures 1 and 2. Seasonal variations are reflected in delay and buffer, but more prominent in the former. The positive arrival delay stays at approximately the same level during the second half of the 1990s, and is reduced substantially because of the air travel slump after the 9/11 terrorist attacks. The recovery of air travel progressively resumes the delay level, with a new peak reached in the summer of 2007. A different story occurs to the buffer side however. Over the years airlines have progressively added more buffers to their flight schedule, and this trend does not seem to be significantly affected by the 9/11 event. Unlike delay, changes in buffer are embodied in the airline schedule planning process, which often extends over a year with updates every three months (Ramdas and Williams, 2008). Some slight decline of buffer around 2003 may reflect such response to the 9/11 aftermath. Less responsive to external, short-term traffic variations, the amount of buffer built in schedule is likely to suggest airlines' longer-term expectations of delay, as evidenced by its more consistent growth trend.

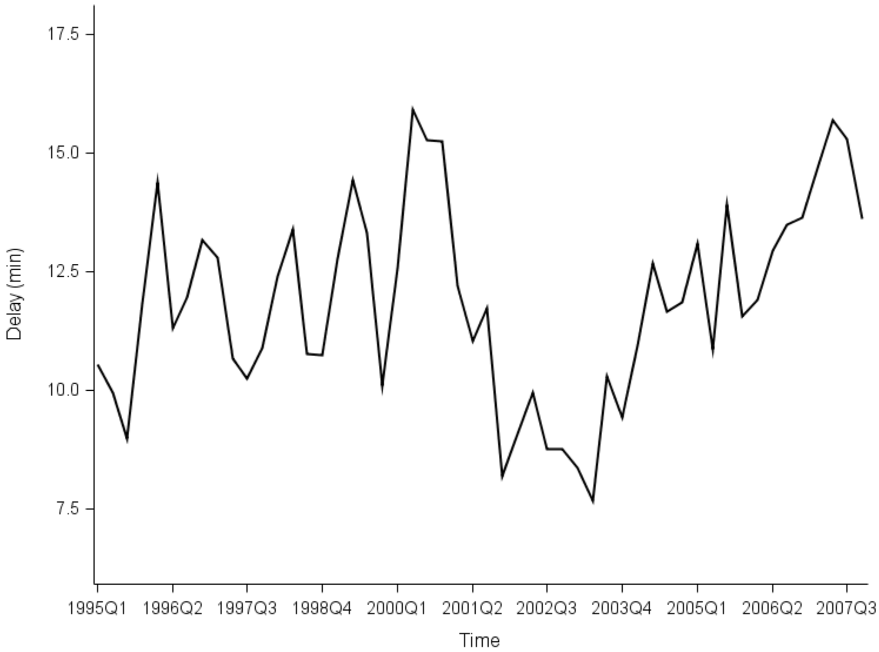

Figure 1 – Average Positive Arrival Delay across Carriers over Time

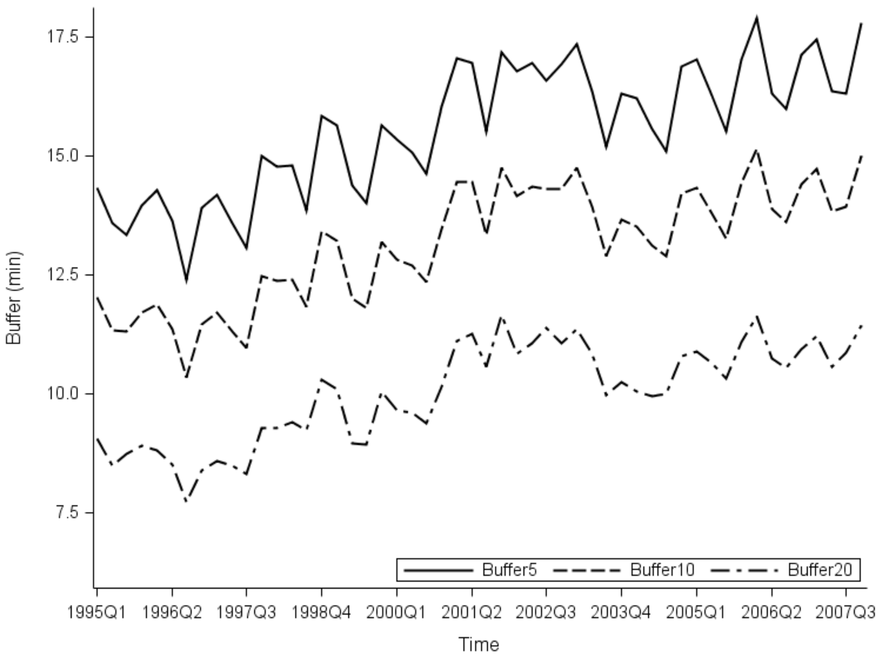

Figure 2 – Average Buffer across Carriers over Time

Carrier wise, American and Continental build the most buffers into their schedules, while the least is seen in Alaska Airlines (Figure 3). The buffer levels may reflect different airline strategies for coping with the volatility of flight time, but can also be attributed to other factors. For example, the majority of Alaska Airlines' business is outside of the most congested areas in the US, therefore less prone to delay. United incurs the largest average delays, while also having a relative large amount of schedule padding—the third most among the carriers being compared. In contrast, Southwest has low levels of both buffer and delay.

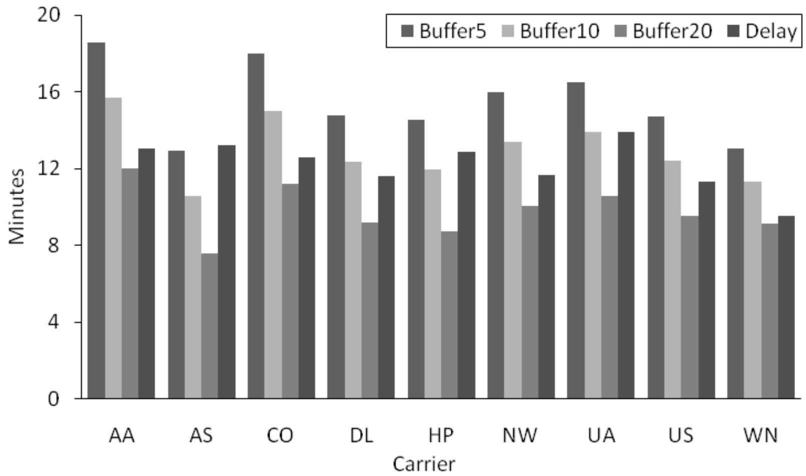

Figure 3 – Delay and Buffer by Carrier

While the above delay and buffer metrics are well-accepted in measuring airline operational performance, we also develop a second set of metrics which looks at the delay phenomenon from a different perspective. Recognizing that quantifying flight operational performance essentially rests on two elements, the planned flight time (schedule) and actual flight time, in the following we explicitly characterize the relationship between the times when a given flight is scheduled to be, and actually is, active. The metrics developed below describe operational performance based on this idea.

We first introduce three new time measures: total absorbed time, scheduled time, and actual flight time. The total absorbed time (TAT) of a flight is defined as the time interval between the earlier of scheduled and actual departure times, and the later of the scheduled and actual arrival times. Scheduled time (S) is a subset of TAT, defined as the time between the scheduled departure and scheduled arrival. Actual flight time (A) denotes the time from the actual departure to the actual arrival, and is also a subset of TAT.

Using these three measures, the TAT for any flight can be categorized into the following subsets: scheduled-active time ($S \cap A$), scheduled-non-active time ($S \cap \sim A$), active-non-scheduled time ($\sim S \cap A$) time, and non-scheduled-non-active time ($\sim S \cap \sim A$). $S \cap A$ denotes the time falling into both the scheduled and actual flight time intervals. $S \cap \sim A$ is the time within the scheduled flight time but outside the actual flight time. It can be caused by either late departures or early arrivals. $\sim S \cap A$ represents the converse, which results from early departures and late arrivals. In the rare events of extremely early or late departures, time between the actual arrival and scheduled departure, or between the scheduled arrival and actual departure, is $\sim S \cap \sim A$. Theoretically there are six possible situations, as illustrated in Figure 4. For each situation, the solid and dashed arrow lines represent respectively the scheduled and actual flight time. For example, if the scheduled departure time of a flight was 7:00am and it actually left the gate at 7:30am, then the total duration of $S\cap\sim A$, denoted by $T_{S\cap\sim A}$, equals 30 min. At the arrival end, the scheduled arrival time is 9:00am but the flight pulled up to the gate at 9:20am. Following the same notation, the total duration of $\sim S\cap A$, $T_{\sim S\cap A}$, is just the arrival delay, equal to 20min. The time between the actual departure and the scheduled arrival is $S\cap A$, amounting to 90min (i.e. $T_{S\cap A}$ = 90min). The case described here corresponds to the top-left situation, i.e. late-departure-late-arrival. The other five situations can be described as: early-departure-early-arrival (top right), late-departure-early-arrival (middle left), early-departure-late-arrival (middle right), extremely-late-departure (bottom left), and extremely-early-departure (bottom right).

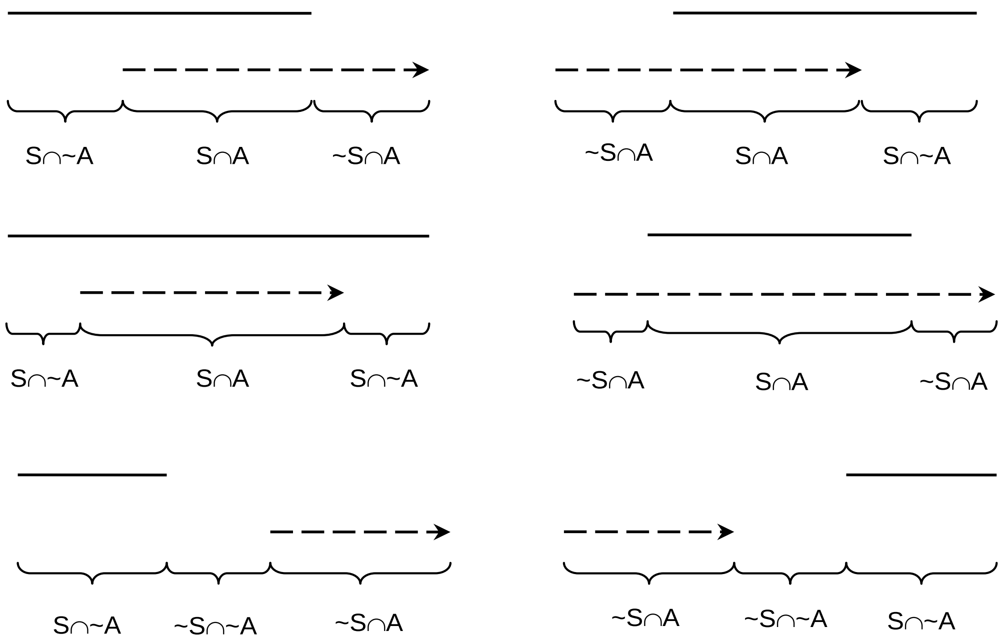

Figure 4 – Identification of Time Components in Six Possible Situations

Based on the above time categorization, three new metrics, which we term "time-based" metrics, are constructed: the duration of TAT, denoted T_{tot}, the fraction of time in $S\cap\sim A$ (i.e. $T_{S\cap\sim A}/T_{tot}$), which we denote $P_{S\cap\sim A}$, and the fraction that is in $\sim S\cap A$ (i.e. $T_{\sim S\cap A}/T_{tot}$), denoted $P_{\sim S\cap A}$. The aggregate portions of $S\cap\sim A$, $\sim S\cap A$, and $\sim S\cap\sim A$, by quarter and carrier, are plotted in Figures 5 and 6. The sum of the three components, equal to the portion of $\overline{S\cap A}$ in TAT (i.e. $P_{\overline{S\cap A}}$), reflects the degree to which actual operations deviate from scheduled ones. Such deviation accounts for 15 to more than 20 percent in TAT, with $S\cap\sim A$ and $\sim S\cap A$ being the dominant components. By contrast, $\sim S\cap\sim A$ accounts for little more than 1 percent in most cases. Among the airlines, Southwest has the highest $P_{\overline{S\cap A}}$, in part a consequence of its high $P_{S\cap\sim A}$, which can be attributed not only to a significant fraction of late departures and early arrivals, but also a short average stage length. *Ceteris paribus*, the latter increases the fraction of delay minutes in TAT. US Airways and Alaska hold the 2^{nd} and 3^{rd} highest $P_{\overline{S\cap A}}$'s, presumably for similar reasons. For American and Continental, which perform substantial padding while still maintaining a certain level of delay, their relatively long stage length leads to low $P_{\overline{S\cap A}}$'s.

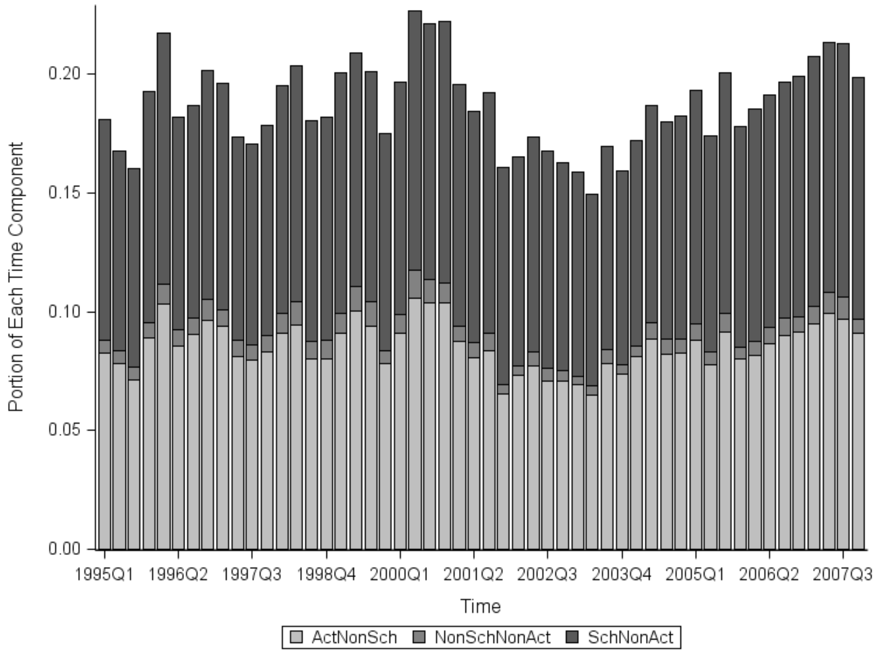

Figure 5 – Portion of the Time Components by Quarter
(SchNonAct (S∩~A), ActNonSch (~S∩A), and NonSchNonAct (~S∩~A))

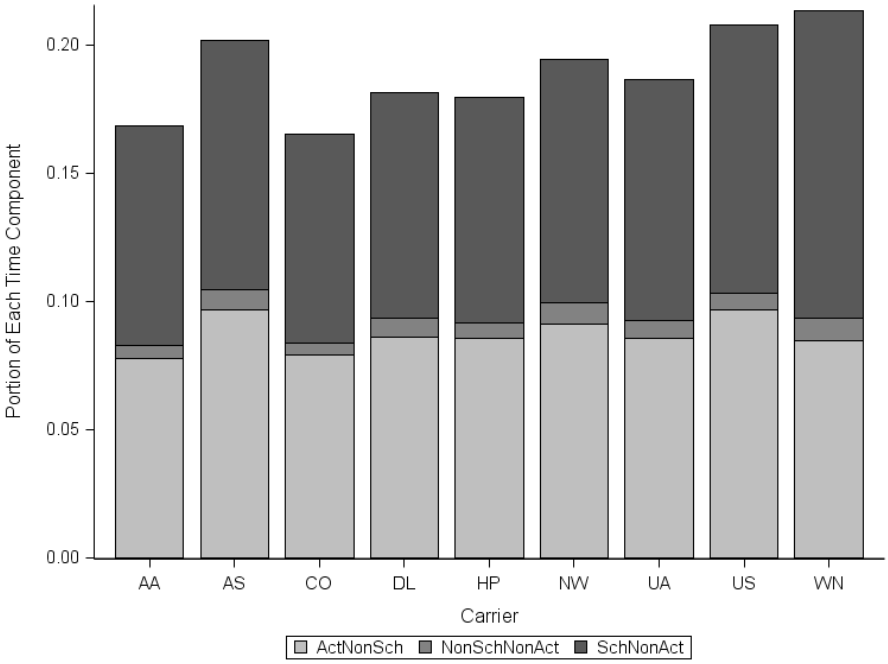

Figure 6 – Portion of the Time Components by Carrier
(SchNonAct (S∩~A), ActNonSch (~S∩A), and NonSchNonAct (~S∩~A))

COST MODEL SPECIFICATION

In addition to the above operational performance metrics, other explanators as well as the functional form must be specified in order to develop an airline cost model. Economic theory suggests that in the production process, each firm minimizes the cost C_{it} at which it can produce a given amount of output Y_{it}, given the price it pays for inputs $\vec{W}_{it}$, where subscript *i* denotes a particular firm (airline), and *t* identifies the time period. In our study, the selected airlines are all passenger service focused, with only a small portion of their traffic undertaking cargo, mail, and other types of business. We therefore use a single measure, revenue-ton-mile (RTM), to represent the aggregate output. Typical inputs considered include labor, fuel, capital, and materials. In reality, however, capital inputs cannot be adjusted to the optimal level instantaneously (Cave, Christensen and Tretheway, 1984; Gillen, Oum and Tretheway, 1990). We relax the assumption of optimal capital stock by treating capital input, denoted by *S*, as quasi-fixed and employing a variable cost function to reflect the short-run cost minimization process. The airline variable cost function can then be written as a function of its output Y_{it}, the price of the three variable inputs (fuel, labor, and materials) $\vec{W}_{it}$, and capital input S_{it}, i.e. $VC_{it} = f(Y_{it}, \vec{W}_{it}, S_{it})$.

In the airline cost literature, it has long been recognized that costs depend on the nature and quality of airlines' output as well as the quantity. Because the nature and quality of output also vary over time and across carriers, the specification of the airline cost function above needs to take these into account. A set of additional variables $\vec{Z}_{it}$ describing the nature of the output are introduced. Variables of this kind that often appear in literature include a measure of the size of the airline's network (the number of points served) and the average stage length. In this study we further add a vector $\vec{N}_{it}$ which captures airline *i*'s operational experience during time period *t*. In a more general sense, $\vec{N}_{it}$ can be regarded as part of the output characteristics. As we have developed two sets of airline operational performance metrics, there are more than one specification of the airline cost model. The new general functional form becomes $VC_{it} = f(Y_{it}, \vec{W}_{it}, \vec{Z}_{it}, S_{it}, \vec{N}_{it})$.

We employ a translog functional form due to its flexibility. The dependent and all continuous explanatory variables are de-meaned; therefore the translog form can be regarded as a second-order Taylor expansion to approximate any arbitrary cost function about the mean values in the dataset. Since our dataset is a panel, it is also important to consider the use of fixed effects. Note that substantial inter-period variation persists in the operational performance metrics. Including time period fixed effects would affect our model's ability to discern the impact of operational performance on airline cost. We instead employ a time trend variable taking the value 1 in the first quarter of our data set, 2 in the second quarter, etc. On the other hand, recognizing that different airlines may have systematic cost differences, we include firm fixed effect dummies. This helps preclude potential bias in the coefficients – bias that might arise from the omission of unmeasurable variables that vary by airline but are constant over time. Although some might argue models without airline fixed effect capture the long-run impact of persistent inter-airline differences in operational performance, we do not pursue this here. To some extent, including firm fixed effect is more consistent with the choice of the variable cost functional form which reflects airlines' short-run behavior.

The conventional translog functional form implicitly assumes the dependent and all continuous explanatory variables are in logarithmic form. As the true relationship between cost and operational performance has rarely been tested, for comparison reasons we estimate models with operational performance variables taking both the logarithmic and level forms. If logarithmic values are chosen, the translog form for the variable cost function is:

$$\begin{aligned}\ln(VC_{it}) = {} & \alpha_i + \alpha_T t + \beta\ln(Y_{it}) + \sum_j \gamma_j \ln(W_{jit}) + \sum_j \delta_j \ln(Z_{jit}) + \sum_j \omega_j \ln(N_{jit}) + \kappa\ln(S_{it}) \\ & + \frac{1}{2}\eta_{YY}[\ln(Y_{it})]^2 + \frac{1}{2}\sum_j\sum_k \phi_{jk}\ln(W_{jit})\ln(W_{kit}) + \frac{1}{2}\sum_j\sum_k \mu_{jk}\ln(Z_{jit})\ln(Z_{kit}) \\ & + \frac{1}{2}\sum_j\sum_k \varphi_{jk}\ln(N_{jit})\ln(N_{kit}) + \frac{1}{2}\lambda_{SS}[\ln(S_{it})]^2 + \sum_k \theta_{Yk}\ln(Y_{it})\ln(W_{kit}) \\ & + \sum_j\sum_k \rho_k \ln(Y_{it})\ln(Z_{kit}) + \sum_j\sum_k \sigma_{jk}\ln(Y_{it})\ln(N_{kit}) + \varsigma_{YS}\ln(Y_{it})\ln(S_{it}) \\ & + \sum_j\sum_k \tau_{jk}\ln(W_{jit})\ln(Z_{kit}) + \sum_j\sum_k \upsilon_{jk}\ln(W_{jit})\ln(N_{kit}) + \sum_j \xi_{jS}\ln(W_{jit})\ln(S_{it}) \\ & + \sum_j\sum_k \psi_{jk}\ln(Z_{jit})\ln(N_{kit}) + \sum_j \zeta_{jS}\ln(Z_{jit})\ln(S_{it}) + \sum_j d_{jS}\ln(N_{jit})\ln(S_{it}) + \varepsilon_{it}\end{aligned} \qquad (2)$$

where VC_{it} is the variable cost for airline *i* in time period *t*; *t* a time trend variable. The other variables follow the specification in the beginning of the section, with subscript *j* denoting the *j*th component in the relevant argument vectors (i.e. $\vec{W}_{it}, \vec{Z}_{it}, \vec{N}_{it}$). ε_{it} is a stochastic error term.

The symmetry of coefficients in the cost function requires $\phi_{jk} = \phi_{kj} \quad \mu_{jk} = \mu_{kj} \quad \varphi_{jk} = \varphi_{kj}$. In addition, a cost function must be linearly homogeneous in the input prices, requiring the following restrictions to be imposed on the translog cost function:

$$\begin{aligned}&\sum_j \gamma_j = 1 \quad \sum_k \theta_{Yk} = 0 \quad \sum_j \xi_{jS} = 0 \\ &\sum_j \phi_{jk} = 0 \quad \sum_j \tau_{jk} = 0 \quad \sum_j \upsilon_{jk} = 0 \quad \forall k\end{aligned} \qquad (3)$$

Shephard's lemma implies the input shares be equated to the logarithmic partial derivatives of the cost function with respect to the input prices:

$$\frac{\partial\ln(VC_{it})}{\partial\ln(W_j)} = S_{jit} = \gamma_j + \sum_k \phi_{jk}\ln(W_{kit}) + \theta_{Yj}\ln(Y_{it}) + \sum_k \tau_{jk}\ln(Z_{kit}) + \sum_k \upsilon_{jk}\ln(N_{kit}) + \xi_{jS}\ln(S_{it}) \qquad (4)$$

To further increase estimation efficiency, we follow Oum and Zhang (1991) and Oum and Yu (1998) by imposing an additional equation for the shadow value of capital input:

$$-\frac{C_{it}^k}{VC_{it}} = \frac{\partial\ln(VC_{it})}{\partial\ln(S_{it})} = \kappa + \lambda_{SS}\ln(S_{it}) + \varsigma_{YS}\ln(Y_{it}) + \sum_j \xi_{jS}\ln(W_{jit}) + \sum_j \zeta_{jS}\ln(Z_{jit}) + \sum_j d_{jS}\ln(N_{jit}) \qquad (5)$$

where C_{it}^k is the depreciated capital cost. Equation (5) is basically the first order condition for the short-run total cost minimization which endogenizes the capacity utilization (Oum and Zhang, 1991; Oum and Yu, 1998).

The translog cost function (2), the cost share Equations (4), and the shadow price of capital input Equation (5) are jointly estimated using Zellner's method of seemingly unrelated regression, which takes into account contemporaneous correlation among the error terms across equations. Because the input shares add to unity, one of the share functions is excluded from the model (in our case, fuel and labor equations are retained). Since the models are estimated in deviation form, the first order coefficients in the cost function can be read as elasticities (if in logarithmic form) at the mean value of the data.

We estimate the models on quarterly panel data of nine US major airlines spanning from the first quarter of 1995 to the fourth quarter of 2007. Because there is some time inconsistency between Form 41 and the Airline On-time Performance database for American West and US Airways due to their merger, observations of the two airlines starting from 2005 are excluded. Aside from the operational performance variables, most of the data are either directly available from the airline balance sheet, traffic, and expenditure information published in the BTS Form 41 database, or can be obtained using simple calculations. We focus on domestic data, since the BTS Airline On-time Performance records are only for domestic flights.

As discussed above we use a single RTM as a measure for output. Fuel and labor input prices are calculated using fuel expenses per gallon and labor expense per employee per quarter. To account for the difference between full- and part-time employees, we use a weighted sum of employment based on the hours paid to employees. As a proxy for materials price, we choose the producer price index (PPI), which varies by quarter but not by airline. The index data are collected from the US Bureau of Labor Statistics. Following Oum and Yu (1998), capital input is obtained by multiplying the capital stock with the utilization rate, for which load factor is used as a proxy. We do not pursue the perpetual inventory method to construct capital stock series because of the rather arbitrary depreciation rules used by airlines. Instead our measure of capital stock consists of the asset values plus investment for each airline-quarter. Four types of assets are included: flight equipment, ground property and equipment, capital leases, and land. For variables in vector $\bar{Z}_{it}$, we divide the total distance flown by the total number of departures performed to obtain the average stage length. Information regarding the number of points served is not directly attainable, but extracted from the BTS Airline On-time Performance database. The summary statistics of the sample data are presented in Table 1.

Table 1 – Descriptive Statistics of Key Variables

	Mean	Std. Dev.	Minimum	Maximum
Revenue-ton-miles (million)	1266.5	662.5	176.6	2541.9
Fuel price ($/gallon)	0.94	0.52	0.36	2.68
Labor price ($/employee)	17800.7	4111.1	8688.8	30729.4
Materials price (PPI)	147.9	22.3	109.3	187.9
Capital input (million $)	11314.7	8524.8	589.4	29127.7
Load factor (%)	72.0	5.7	55.3	87.4
Stage length (miles)	815.9	187.2	396.5	1167.9
Number of points served	80.4	26.0	34.0	130.0
Variable cost (million $)	1548.3	864.3	183.2	3513.6
Positive arrival delay (min)	12.2	3.2	5.5	28.8
Average buffer5*	15.5	2.9	7.1	22.4
Average buffer10	13.0	2.6	5.0	19.3
Average buffer20	9.8	2.3	2.7	15.5
$T_{S\cap A}$ (million min)	15.7	7.1	3.2	32.0
$T_{S\cap\sim A}$ (million min)	1.8	0.9	0.2	4.6
$T_{\sim S\cap A}$ (million min)	1.6	0.8	0.3	4.9

* Average buffer5/10/20 denote buffer are measured against the 5th/10th/20th percentiles of the observed travel time, as defined in the previous section.

ESTIMATION RESULTS

Incorporating the aforementioned performance metrics sets into the cost function suggests two different types of models, which we term delay-buffer and time-based models. Table 2 reports the estimation results of delay-buffer models, with operational performance variables taking the level form. The estimation results for models with the delay and buffer variables in log form are provided in Appendix B. To conserve space, only the estimates and standard errors of first-order coefficients are presented. Table 2 contains three models, corresponding to the three buffer measures discussed in section 2. All of them have very high goodness-of-fit. In order to be consistent with the economic theory, the curvature conditions of the cost models are further checked. Our results show that, for all three models about 70 percent of the data points in the sample satisfy the curvature conditions, which compares favorably to other airline cost studies in which such a statistic is reported (e.g. Cave, Christensen and Tretheway, 1984; Chua et al., 2005).

The individual coefficients reflect the sensitivity of cost to various regressors at the sample mean. The first-order coefficients for input prices indicate that at the sample mean, fuel and labor inputs account for about 20% and 38%, respectively, of the total variable cost. This leaves the materials input to account for 41% of the total variable cost. The first-order coefficient for capital input is negative, implying a positive shadow value of capital input. The coefficient for average stage length, around -0.2, indicates that a 1 percent increase in average stage length, output held constant, causes a decrease in variable cost of about 0.2 percent. This should be interpreted as the effect on cost of flying fewer passengers over a longer distance each to obtain the same level of output. The coefficients for points served, about 0.67, suggest a 1 percent increase in network size leads to an increase in variable cost of about 0.6 percent at the sample mean. Perhaps surprisingly, the time trend variable has a positive coefficient, suggesting a trend toward diminishing productivity. However, the effect is quite small.

The traditional measure of returns to scale (RTS), proposed by Cave, Christensen and Swanson (1981), is defined as the percent increase in output and points served made possible by a one percent increase in variable cost and capital input, at the sample mean:

$$RTS = (1-\kappa)/(\beta + \delta_{\text{points}}) \tag{6}$$

The calculated RTS and standard errors are reported at the bottom in Tables 2. It is intriguing to examine whether there exist constant returns to scale, as found in many previous studies (Cave, Christensen and Tretheway, 1984; Gillen, Oum and Tretheway, 1990; Oum and Zhang, 1991). For Models 1-3, we fail to reject the hypothesis of constant returns to scale at 5% level. The underlying assumption in using this measure is that variables other than output and network size are kept unchanged. While an alternative measure of RTS exist (e.g. Xu, Windle et al., 1994; Oum and Zhang, 1997), we do not attempt this here, but note that, Oum and Zhang (1997) find a slight increase in the value of RTS under the alternative measure.

Table 2 – Delay-buffer Cost Function Estimation Results (Level Form for Delay and Buffer)

	Model 1		Model 2		Model 3	
	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Output (RTM)	0.4875***	0.0369	0.4831***	0.0362	0.4783***	0.0356
Fuel price	0.2016***	0.0016	0.2014***	0.0016	0.2010***	0.0016
Labor price	0.3858***	0.0022	0.3856***	0.0022	0.3853***	0.0022
Materials price	0.4126***	0.0032	0.4130***	0.0032	0.4136***	0.0032
Capital input	-0.0547***	0.0009	-0.0546***	0.0009	-0.0544***	0.0009
Stage length	-0.2172***	0.0837	-0.2071**	0.0838	-0.1913**	0.0835
Points served	0.6650***	0.0573	0.6685***	0.0571	0.6720***	0.0569
Avg. pos. delay	0.0060***	0.0014	0.0059***	0.0014	0.0057***	0.0015
Avg. buffer5	0.0070***	0.0027				
Avg. buffer10			0.0066**	0.0029		
Avg. buffer20					0.0057*	0.0031
Time trend	0.0012**	0.0006	0.0013**	0.0006	0.0013**	0.0006
R^2		0.9902		0.9901		0.9900
RTS	0.9152	0.0439	0.9157	0.0441	0.9167	0.0444

*** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Turning now to the variables that are of particular interest to this study, the estimates of the operational performance variables support our hypothesis that delay and buffer affect airline cost. In all models, delay proves to be consistently significant. In Models 1-3, the coefficients suggest that, at the sample mean, one minute increase in delay would cause 0.6% increase in variable cost. The models with operational performance variables taking the log form indicate that, the cost elasticity with respect to delay is about 0.7 at the sample mean. For the buffer variables, the significance of the coefficients diminishes with the increase in the percentiles chosen. Under the linear-form (Models 1-3), the coefficients suggest a 0.6-0.7%

cost increase would occur if there is a one-minute increase in buffer from the sample mean point. Doing this requires a concurrent shortening of flight schedule and actual flight time (in order to keep delay level unchanged). The estimates appear inconclusive when the log form is employed. This may imply the level form be a better approximation to the relationship between airline cost and operational performance. Of the three buffer measures considered, the "Avg. buffer5" is preferred because of its highest coefficient significance.

Table 2 shows that the magnitudes of first-order delay and buffer coefficients are very close under the level form specification. Hypothesis tests fail to reject the null that the two coefficients are equal for Models 1 and 2, at the 0.05 level. On the other hand, there exist several studies which use a single measure of delay against their defined unimpeded flight time (e.g. Mayer and Sinai, 2003a; Dresner et al., 2010). Accordingly, we also re-estimate Models 1-3 with a single delay variable. This variable equals the sum of the average positive arrival delay and schedule buffer. Estimation results are presented in Table 3, where "New delay5/10/20" denote delays that are measured against $5^{th}/10^{th}/20^{th}$ percentiles in the travel time defined before. The coefficients for the new delay variable do not significantly differ from those for the delay and buffer variables in Models 1-3. Neither do the remaining estimates, to some extent justifying the use of a single delay variable in the airline cost models.

Table 3 – Delay-based Cost Function Estimation Results

	Model 4		Model 5		Model 6	
	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Output (RTM)	0.4847***	0.0338	0.4798***	0.0339	0.4743***	0.0342
Fuel price	0.2012***	0.0016	0.2011***	0.0016	0.2009***	0.0016
Labor price	0.3862***	0.0022	0.3861***	0.0022	0.3859***	0.0022
Materials price	0.4125***	0.0032	0.4128***	0.0032	0.4132***	0.0032
Capital input	-0.0543***	0.0009	-0.0542***	0.0009	-0.0541***	0.0009
Stage length	-0.1873**	0.0774	-0.1749**	0.0775	-0.1571**	0.0776
Points served	0.6544***	0.0555	0.6596***	0.0556	0.6658***	0.0558
New delay5	0.0067***	0.0014				
New delay10			0.0065***	0.0014		
New delay20					0.0061***	0.0015
Time trend	0.0009**	0.0005	0.0009**	0.0005	0.0010**	0.0005
R^2		0.9901		0.9900		0.9889

*** significant at 1% level; ** significant at 5% level; * significant at 10% level.

We observe that the coefficients for some higher order delay terms are not significant in the above models. Since the airlines cost models will be used for the subsequent delay cost estimation, keeping these variables in the model will jeopardize the robustness of the cost estimates. As a consequence insignificant delay terms (these are delay*delay for Model 4, and delay*delay and delay*stage length for Models 5 and 6) are removed and the above models are re-estimated. It should be noticed that, even after dropping out the higher-order terms, the first-order effect is non-linear by construction, since each additional minute of delay/buffer has the same percentage impact on cost. The estimates for the remaining coefficients are almost unchanged. All the terms that involve the delay variable now have statistically significant coefficients. Table 4 documents the first-order coefficient estimates for these new models.

Table 4 – Delay-based Cost Function Estimation Results (with Insignificant Higher Order Delay Terms Dropped)

	Model 7		Model 8		Model 9	
	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Output (RTM)	0.4814***	0.0338	0.4840***	0.0339	0.4793***	0.0342
Fuel price	0.2013***	0.0016	0.2012***	0.0016	0.2010***	0.0016
Labor price	0.3862***	0.0022	0.3861***	0.0022	0.3859***	0.0022
Materials price	0.4125***	0.0032	0.4127***	0.0032	0.4131***	0.0032
Capital input	-0.0543***	0.0009	-0.0542***	0.0009	-0.0541***	0.0009
Stage length	-0.1746**	0.0771	-0.1753**	0.0771	-0.1603**	0.0771
Points served	0.6594***	0.0556	0.6628***	0.0558	0.6672***	0.0559
New delay5	0.0059	0.0012				
New delay10			0.0061***	0.0013		
New delay20					0.0058***	0.0013
Time trend	0.0009*	0.0005	0.0010*	0.0005	0.0010*	0.0005
R^2		0.9900		0.9899		0.9898

*** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Turning now to the time-based models, T_{tot}, $P_{S\cap\sim A}$, and $P_{\sim S\cap A}$ are the operational performance variables entering the airline cost model. T_{tot} measures the total amount of time the aircraft fleet and crews of an airline are dedicated, in either plan or execution, to performing flights. Since T_{tot} is integral to airline production, we keep this variable in logarithmic form. The other two variables are included in level form since they can, in principle, be eliminated under ideal operating conditions.[5] The portion of ~S∩~A in TAT, $P_{\sim S\cap\sim A}$, is not considered because it is very small in total operations. Estimation results are shown in Table 5 (Model 10). In Model 10, we observe the coefficients for the majority of the higher-order time variables not involving input prices are insignificant.[6] Again, to ensure more robust cost estimates, we re-estimate a "quasi" version of Model 10, by dropping out the higher-order time variables that do not involve input prices. Estimation results are reported in the 3rd and 4th columns of Table 5 (Model 11).

The results from Model 10 and 11 are consistent, but somewhat different from those of delay-buffer models in several aspects. The output coefficient and its significance are substantially lower, due to the inclusion of the T_{tot} variable. As a consequence the conventional formula is not applicable to calculate RTS. Stage length is no longer significant and has a seemingly counter-intuitive sign, its effect captured by the T_{tot} variable, since longer average stage length allows the same output to be produced with less flight time.

[5] We also estimated a time-based model with both $P_{S\cap\sim A}$ and $P_{\sim S\cap A}$ taking the logarithmic form. Results are documented in Appendix B. Compared to Model 10, the model implications largely remain the same.
[6] Only one among the 15 such variables has a coefficient estimate that is significant at 5% level.

Table 5 – Time-based Model Estimation Results

	Model 10 (Full)		Model 11 (Quasi)	
	Est.	Std. Err.	Est.	Std. Err.
Output (RTM)	0.2102***	0.0561	0.2424***	0.0531
Fuel price	0.1997***	0.0016	0.1995***	0.0016
Labor price	0.3860***	0.0021	0.3858***	0.0021
Materials price	0.4143***	0.0031	0.4147***	0.0031
Capital input	-0.0537***	0.0009	-0.0536***	0.0009
Stage length	0.0918	0.0880	0.0979	0.0783
Points served	0.5111***	0.0718	0.4901***	0.0590
T_{tot}	0.4368***	0.0725	0.4424***	0.0687
$P_{\sim S \cap A}$	1.0875***	0.3740	0.7111**	0.3201
$P_{S \cap \sim A}$	-0.4211	0.5167	-0.0492	0.4383
Time trend	0.0006	0.0006	0.0003	0.0005
R^2		0.9902		0.9896

*** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Focusing on the operational performance variables, the coefficient for T_{tot} has the expected positive coefficient and is highly significant. The $P_{\sim S \cap A}$ variable has a significantly positive coefficient, suggesting everything else held equal, flight activity outside the schedule window results in additional cost. The coefficient for $P_{S \cap \sim A}$ implies that flight inactivity during the schedule window—either because of departing late or arriving early—does not significantly reduce costs. There can be several interpretations for late departures not reducing cost. First, compared to a late departure, an aircraft may push back "on-time" from the gate but just stay on the ramp/tarmac so as to gain a superficial good on-time departure record. Little cost difference would result between this "on-time" departure and an actual late push-back from the gate. It is also likely that, when departing late, to catch up with the schedule a flight will need to fly at a faster speed and thus increase fuel consumption, resulting in a cost penalty. Furthermore, many resources used for airline production are allotted in advance, and often not adaptive to favorable flight time situations, such as arriving ahead of schedule. Therefore, an early arrival may yield little cost savings vis-à-vis an on-time flight. Compared to the full model, the coefficient for $P_{S \cap \sim A}$ in the quasi model remains insignificant and is much smaller. The $P_{\sim S \cap A}$ coefficient is also somewhat smaller (but still significant), apparently as a result of absorbing the effect of higher-order terms in the full model.

COST IMPACT OF OPERATIONAL PERFORMANCE ON AIRLINES

In this section, the previously estimated cost models are used to gauge the potential cost impact of operational performance on airlines. The year under investigation is 2007, the worst on record in terms of flight delays. We choose the more robust Models 7-9 and 11 to perform the cost estimation. Using Models 7-9, two scenarios are considered. In the first scenario, delay against schedule is entirely eliminated, without changing the amount of schedule buffer; in the second scenario, we further reduce schedule buffer to zero. The new operating costs for each airline-quarter observation are predicted under the two scenarios, which are compared to the predicted costs with delay and schedule buffer kept at their original level (baseline). The difference between the predicted new and baseline costs, under the two scenarios, is the estimated cost of delay against schedule and the total cost delay against schedule and schedule buffer, respectively. The difference between these two costs corresponds to the cost of schedule buffer. This schedule buffer cost should be interpreted as the potential savings by reducing flight schedule from the current level to the "minimum" (unimpeded) amount while maintaining delay at the zero level. Estimates for these costs appear in the first three rows of Table 6.

We also use Model 11 to investigate the cost impact under two scenarios. In the first one, T_{tot} is set to be the sum of $T_{S \cap A}$ and $T_{S \cap \sim A}$ over all flights, with $P_{\sim S \cap A}$ and $P_{S \cap \sim A}$ reduced to zero. Under this scenario, a flight's actual departure and arrival times exactly coincide with the current schedule, which contains some schedule buffer. In the second scenario, all flights not only fly strictly according to schedule, but the schedule now is based upon unimpeded flight times. Therefore, in addition to zero $P_{\sim S \cap A}$ and $P_{S \cap \sim A}$, we further reduce T_{tot} to the sum of unimpeded flight times across all flights. We perform this calculation by employing each of the three unimpeded flight time measures used to estimate the delay-buffer cost models. Similar to before, airline costs are predicted under the two scenarios, and compared to the predicted value of the baseline cost. By a rough analogy with the scenarios using the delay-buffer models, we consider the difference between the baseline cost and the cost in the first scenario as the cost of delay against schedule, and the difference between the baseline cost and the cost in the second scenario the total cost of operational performance. The difference between the two defines the cost of schedule buffer. Estimates are reported in rows 4-6 in Table 6.

We obtain somewhat larger estimates of both delay-against-schedule and total cost using the time-based model. This may be attributed to the counterfactual considered in the time-based model which entails perfect adherence to both arrival and departure time schedules, whereas the delay-based models focus on just the arrival end. In any case, the comparable magnitudes of the cost estimates provide some cross-validation between the models using different operational performance metrics sets. If we follow convention and divide the total delay-against-schedule cost by the total delay minutes using the delay-buffer cost estimates, an average delay cost factor of $66/hr will result,[7] which is comparable to existing cost factors in the literature. This cost factor is certainly a composite one containing overhead cost, and averaged over all airborne and ground delays. On the other hand, the buffer costs in the two models are obtained following the same idea of first eliminating delay and then reducing buffer. The two types of models produce quite similar numbers—$2-3 billion for the seven major airlines.

As a first-order industry-wide estimate, we extrapolate the above cost to the entire system based on the portion of available seat miles (ASM) provided by the 7 major airlines in all carriers reporting data to BTS (Table 6). Although this leaves out some regional and commuter airlines (those whose annual operating revenue is below $20 million), such airlines

[7] This is the averaged value of the cost factors obtained from Models 7-9.

account for a very small fraction of the total ASM, so excluding them will have little effect on the system-wide result.

Table 6 – Airline cost estimates ($ billions), for 2007

	Cost category		Buffer5**	Buffer10	Buffer20
7 major airlines*	Delay-based model	Delay against schedule	4.1	3.3	3.1
		Buffer	3.2	2.6	1.9
		Total	7.3	5.9	5.0
	Time-based model	Delay against schedule	6.7	6.7	6.7
		Buffer	2.8	2.4	1.8
		Total	9.5	9.1	8.5
Industry wide	Delay-based model	Delay against schedule	5.7	4.6	4.4
		Buffer	4.6	3.7	2.7
		Total	10.3	8.3	7.1
	Time-based model	Delay against schedule	9.4	9.4	9.4
		Buffer	4.1	3.4	2.7
		Total	13.5	12.8	12.1

* US Airways and American West are excluded due to merger.
** Buffer5 denotes that schedule buffer is measured against the 5th percentiles in the observed travel time for a given segment, airline, and quarter, and Buffer10 and Buffer20 likewise. When delay-based models are used, this further purports that the corresponding delay-based cost models (Models 7-9) are used.

Table 7 shows the cost estimates in previously published cost-of-delay studies. Among them, only JEC (2008) implicitly accounts for the buffer component. Except for Hansen et al. (2000, 2001), all others assume delay is entirely removed. Of particular interest is to compare our estimates with the JEC (2008) and ATA (2008) results, because of the same evaluation year. Under the cost factor approach, the JEC study yields lower estimates than the ATA one due to different cost factors applied, although delay in JEC is calculated against some nominal flight time whereas in ATA delay is against flight schedule. It is not surprising that the total time approach produces larger numbers than the cost factor approach, since this approach is prone to overestimating the true cost. Overall, our cost values are consistent with the range of existing estimates. As we employ a variable cost function with fixed effects and a trend variable, our figures are more conservative since they reflect the short-run cost response to the improved operational performance.

Table 7 – Cost of Delay to Airlines in the US: Estimates from Previous Studies

Sources	Evaluation year	System-wide cost ($ billions)	Approach
JEC (2008)	2007	3.6-6.1	Cost factor
		12.2-23.4	Total time
ATA (2008)	2007	6.9	Cost factor
Hansen et al. (2001)	1995-1997	1-4	Econometric
Hansen et al. (2000)	1995-1997	1.7-2.3	Econometric
Citrenbaum and Juliano (1998)	1996	1.2	Cost factor
Odoni (1995)	1993	2-4	Cost factor
Geisinger (1988)	1986	1.8	Cost factor

CONCLUSIONS

In this paper, we extend the conventional delay characterization by developing two distinct sets of operational performance metrics, which provide a more comprehensive perspective on flight delay phenomena. Different from many existing airline cost-of-delay studies which are based upon strong and untested assumptions about the cost-generating mechanisms, this study adopts a different approach to empirically investigate the impact of operational performance on airline cost structure. Our results from estimating various airline cost models suggest that, both delay and buffer represent important cost drivers. Using the alternative time-base metrics, we learn that while flight activity outside the schedule window increases airline cost, flight inactivity inside the schedule window does not exhibit significant cost impact. Our airline cost models with different operational performance metrics produce roughly comparable estimates of the “cost of delay”, or more precisely the cost of imperfect operational performance. These estimates provide a benchmark for assessing the accuracy of more traditional and simple methods for costing delay. Despite the fundamentally different methodologies employed, it is somewhat reassuring, that estimates of the total cost of delay in the US based on our airline cost modeling approach yield results of a similar magnitude to those from using other approaches.

The cost impact estimates here should be read as an upper bound on the cost savings from aviation infrastructure investments, since some flight delays are unavoidable in the real world. Even in the presence of ample capacity, certain causes of delay, e.g. aircraft mechanical problems, passenger-related aircraft loading delays, etc., are likely to remain with us. So is the flight time uncertainty. In addition, it rarely makes sense to size the air transportation system to fully accommodate peak period traffic flows, and there will almost always be some amount of queuing and delay during those periods. It is also important to note that there can be tradeoffs between throughput and delay. The uncertainty of flight times very often leads to the need to create buffers of arriving flights (airborne queues) in order to insure that arrival capacity to an airport is maximized. All these indicate total elimination of delays is neither practical nor desirable.

Moreover, when investment in capacity is made, both passengers and flight operators would no doubt react to capacity increase and delay reduction, leading to an equilibrium shift in the system. The extent of the shift is often very difficult to predict. The airline cost impact estimates presented here, along with other cost components such as monetized potential passenger delay savings, should be regarded as a frame of reference for decision makers to assess the magnitude of the flight delay problem and the need for initiatives to address it, rather than a precise estimate of the potential benefits from such initiatives. As noted in the beginning of this paper, he combined NextGen and airport capacity investment could reach $100 billion over the next 15 years. If we believe this can lead to a substantial improvement of airline operational performance and cost savings, e.g. 50% of the estimated total cost impact, then this together with the passenger delay saving benefits which previous estimates place at the same or even slightly higher magnitude (e.g. Ball et al, 2010), would well justify these investments. Certainly, further assessment of the linkages between system investments, airline operational performance, and economic benefits is warranted, and these efforts should aim at more aptly quantifying the achievable benefit, to ensure cost-effective investment decisions.

ACKNOWLEDGEMENT

The research was funded by the Federal Aviation Administration through a grant to the National Center of Excellence for Aviation Operations Research (NEXTOR) for “Total Delay Impact”. Earlier versions of this paper were presented at the 2009 INFORMS Annual Meeting in San Diego, and the 12th World Conference on Transport Research in Lisbon, Portugal. The first author would like to thank Professor Robert Windle for kindly sharing his expertise in airline cost modeling, and Professor Yimin Zhang for his helpful suggestions.

REFERENCE

Air Transport Association (ATA) (2008). Cost of Delays: Direct Cost. Available at: http://www.airlines.org/economics/specialtopics/ATC+Delay+Cost.htm, assessed on Jan 25, 2010.

Ball, M., C. Barnhart, M. Dresner, M. Hansen et al. (2010). Total Delay Impact Study. NEXTOR Draft Final Report prepared for the Federal Aviation Administration.

Barnett, A., R. Shumsky, M. Hansen, A. Odoni et al. (2001). Safe at home? An experiment in domestic airline security. Operations Research, 49, 181-195.

Caves, D. W., L. R. Christensen and J. A. Swanson (1981). Productivity growth, scale economies, and capacity utilization in U.S. railroads, 1955-1974. American Economic Review, 71, 994-1002.

Caves, D. W., L. R. Christensen and M. W. Tretheway (1984). Economies of density versus economies of scale: why trunk and local service airline costs differ. Rand Journal of Economics, 15, 471-489.

Chua, C. L., H. Kew and J. Yong (2005). Airline code-share alliances and costs: imposing concavity on Translog cost function estimation. Review of Industry Organization, 26, 461-487.

Citrenbaum, D. and R. Juliano (1998). A simplified approach to baselining delays and delay costs for the national airspace system. Federal Aviation Administration, Operations Research and Analysis Branch, Preliminary Report 12.

Civil Aviation Authority (CAA) (2000). Delay costs based on UK Airline actual P&L figures, SS4 Working Paper on Delay Cost Methodology, UK.

Cook, A., G. Tanner and S. Anderson (2004). Evaluating the True cost to airlines of one minute of airborne or ground delay. Report prepared by the University of Westminster. Performance review unit, EUROCONTORL.

Dresner, M., Britto, R. and Voltes, A. (2010). The impact of flight delay on passenger demand, yields and welfare. Proceedings of the 12th World Conference on Transport Research, Lisbon, Portugal.

Eurocontrol (2007). Standard inputs for EUROCONTORL cost benefit analyses. Available at: www.**eurocontrol**.int/eatm/gallery/content/public/.../CBA-**standard**-values.pdf, accessed on Feb 22, 2009.

Geisinger, K. (1988). Airline delay: 1976-1996. Federal Aviation Administration, Office of Aviation Policy and Plans.

Gillen, D. W., T. H. Oum and M. W. Tretheway (1990). Airline cost structure and policy implications: a multi-product approach for Canadian airlines. Journal of Transport Economics and Policy, 24, 9-34.

Hansen, M. M., D. W. Gillen and R. Djafarian-Tehrani (2000). Assessing the impact of Aviation System Performance using Airline Cost Functions. Transportation Research Record, 1073, 16-23.

Hansen, M. M., D. W. Gillen and R. Djafarian-Tehrani (2001). Aviation infrastructure performance and airline cost: a statistical cost estimation approach. J. Trans. Res., Part E, 37, 1-23.

Institut du Transport Aérien (ITA) (2000). Costs of air transport delay in Europe. Available at: www.eurocontrol.int/prc/gallery/content/public/Docs/stu2.pdf, accessed on Aug 2, 2008.

Janic, M. (1997). The flow management problem in air traffic control: a model of assigning priorities for landings at a congested airport. Journal of Transportation Planning and Technology, 20, 131-162.

Joint Economic Committee (JEC) (2008). Your flight has been delayed again: flight delays cost passengers, airlines, and the US economy billions. Available at: http://jec.senate.gov/index.cfm?FuseAction=Reports.Reports&ContentRecord_id=11116dd7-973c-61e2-4874a6a18790a81b &Region_id=&Issue_id, assessed on Nov 2, 2008.

Mayer, C. and T. Sinai (2003a). Network effects, congestion externalities, and air traffic delays: or why not all delays are evil. American Economic Review, 93, 1194-1215.

Mayer, C. and T. Sinai (2003b). Why do airlines systematically schedule their flights to arrive late? Working paper, the Wharton School of Business, University of Pennsylvania. Available at: http://real.wharton.upenn.edu/~sinai/papers/ Schedule-Mayer-Sinai-10-30-03-2.pdf, accessed on Aug 10, 2009.

Odoni, A. (1995). Research directions for improving air traffic management efficiency. Argo Research Corporation.

Oum, T. H., and C. Yu (1998). Cost competitiveness of major airlines: an international comparison. J. Trans. Res., Part A, 32, 407-422.

Oum, T. H. and Y. Zhang (1991). Utilization of quasi-fixed inputs and estimation of cost functions. Journal of Transport Economics and Policy, 25, 121-134.

Oum, T. H. and Y. Zhang (1997). A note on scale economies in transport. Journal of Transport Economics and Policy, 31, 309-315.

Ramdas, K. and J. Williams (2008). An empirical investigation into the tradeoffs that impact on-time performance in the airline industry. Working paper, University of Virginia, Charlottesville. Available at: http://www.gsb.stanford.edu/FACSEMINARS/events/oit/documents/oit_10_08_ramdas.pdf, accessed on Aug 8, 2009.

Richetta, O. and A. R. Odoni. (1993). Solving optimally the static ground-holding policy problem in air traffic control. Journal of Transportation Science, 27, 228-238.

Wingrove, E. R. et al. (2005). Extent and impact of future NAS capacity shortfalls in the United States: a socio-economic demand study. Proceedings of 6th US/Europe Air Traffic Management Research and Development Seminar, Baltimore, MD, USA, June 2005.

Wu, C. L. and R. E. Caves (2004). Modelling and Optimisation of Aircraft Turnaround Time at an Airport. Journal of Transportation Planning and Technology, 27, 47-66.

Xiong, J. (2010). Revealed preference of airlines' behavior under air traffic management initiatives. PhD dissertation, University of California, Berkeley.

Xu, K. and R. Windle (1994). Re-evaluating returns to scale in transport. Journal of Transport Economics and Policy, 28, 275-286.

APPENDIX A. RESULTS OF FACTOR ANALYSIS FOR DELAY VARIABLES

The idea of factor analysis is to use a smaller set of factors to represent the variation of the observed performance metrics as much as possible. In standard procedure the first two steps are to determine the number of factors to be extracted from the data, and then to rotate these factors so as to facilitate the interpretation of factors. Following the often-cited “eigenvalue greater than or equal to one” rule, two factors are retained. The rotated factor patterns are reported in Table A.

The ten delay variables are collapsed into two factors, with the first one highly correlated with positive delays. The second factor has the highest loadings on the negative delays. Because the departure and arrival delay variables involve both positive and negative parts of delay, their factor patterns become less evident, suggesting the disparities between the positive and negative parts. For reasons alike, the factor pattern of the on-time performance variable is analogous to those of the departure/arrival delay variables (but in a reverse sign since it counts the number of on-time flights). In addition, we observe that the standard deviation of the delay variables varies in a way similar to the positive delay variables, and the variation pattern does not vary by the departure and arrival ends. Overall, one might summarize the results by terming the first factor “positive delay”, and the second factor “negative delay”.

Table A – Rotated Factor Patterns for Delay Variables

Variables	Factor 1	Factor 2
Departure delay	0.79716	-0.48142
Positive departure delay	0.92585	-0.25617
Arrival delay	0.7835	-0.60812
Positive arrival delay	0.94963	-0.27466
Negative arrival delay	-0.02601	0.98761
On-time performance	-0.79374	0.49155
Std. dev. of departure delay	0.97469	0.16744
Std. dev. of positive departure delay	0.98107	0.12658
Std. dev. of arrival delay	0.95249	0.26664
Std. dev. of positive arrival delay	0.97179	0.14986
Std. dev. of negative arrival delay	0.15346	0.9582

APPENDIX B. FIRST-ORDER COEFFICIENTS FOR COST MODELS WITH OPERATIONAL PERFORMANCE VARIALBES TAKING THE LOGARITHMIC FORM

Table B1 – Delay-buffer Cost Function Estimation Results

	Model 1B		Model 2B		Model 3B	
	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Output (RTM)	0.4851***	0.0372	0.4796***	0.0366	0.4724***	0.0359
Fuel price	0.2004***	0.0016	0.2004***	0.0016	0.2003***	0.0016
Labor price	0.3847***	0.0021	0.3846***	0.0021	0.3846***	0.0021
Materials price	0.4149***	0.0031	0.4150***	0.0031	0.4151***	0.0031
Capital input	-0.0537***	0.0009	-0.0536***	0.0009	-0.0535***	0.0009
Stage length	-0.1926**	0.0849	-0.1886**	0.0850	-0.1802**	0.0848
Points served	0.6830***	0.0581	0.6862***	0.0577	0.6909***	0.0572
Avg. pos. delay	0.0682***	0.0160	0.0677***	0.0163	0.0657***	0.0167
Avg. buffer5	0.0858**	0.0410				
Avg. buffer10			0.0648*	0.0371		
Avg. buffer20					0.0367	0.0305
Time trend	0.0012**	0.0006	0.0013**	0.0006	0.0014**	0.0006
R^2	0.9902		0.9901		0.9901	
RTS	0.9021	0.0423	0.9038	0.0426	0.9057	0.0429

*** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table B2 – Time-based Cost Function Estimation Results

	Model 10B	
	Est.	Std. Err.
Output (RTM)	0.2006***	0.0564
Fuel price	0.1998***	0.0016
Labor price	0.3855***	0.0021
Materials price	0.4147***	0.0031
Capital input	-0.0535***	0.0009
Stage length	0.1147	0.0880
Points served	0.5068***	0.0724
T_{tot}	0.4485***	0.0725
$P_{\sim S \cap A}$	0.0942***	0.0325
$P_{S \cap \sim A}$	-0.0385	0.0504
Time trend	0.0005	0.0006
R^2	0.9903	

*** significant at 1% level; ** significant at 5% level; * significant at 10% level.