CAPACITY UTILIZATION AND ITS IMPACT ON AIRLINE'S FINANCIAL PERFORMANCE

– Evidence from the U.S. Airline Industry

ABSTRACT

Capacity utilization is crucial to airline's financial performance and passenger load factor is the most studied capacity utilization measure in the airline industry. However, previous studies on the relationship between load factor and airline financial performance yield mixed results, leaving practitioners and researchers without clear guidance on this matter. Leveraging the concept of slack and the theory of Performance Frontier, we conduct a two-step between-within theorization to hypothesize an inverted-U shaped relationship between load factor and financial performance. As our 16-year panel data also demonstrates both cross-sectional between-carrier differences and longitudinal within-carrier changes, we accordingly adopt a recent econometrics advancement – between-within specification – to model and test our hypothesis. Our results indeed verify our hypothesized inverted-U shaped relationship with a profit-maximizing load factor to be at 88%, indicating that airline practitioners should not target a higher than 88% load factor, in contrary to the popular practice of overbooking which generally leads to higher load factors. Our study contributes to theory development in the operations management field by introducing a two-step between-within theorization that aligns with the nature of the longitudinal data. Our results also corroborate the theory of Performance Frontier and extend the understanding of this theory into the airline industry. In a model fit comparison, our between-within specification demonstrates the best model fit and best forecasting accuracy, reminding researchers to adopt this latest methodological advancement to increase research rigor when modeling secondary data.

Key words: Load Factor, Airline Industry, Financial Performance

CAPACITY UTILIZATION AND ITS IMPACT ON AIRLINE'S FINANCIAL PERFORMANCE

– Evidence from the U.S. Airline Industry

1. INTRODUCTION

The airline industry is critical to the U.S. economy, contributing 5.1% of the U.S. gross domestic product (directly and indirectly); supporting 4.3 million jobs; and generating $1.9 trillion in total economic activity in 2018 (Federal Aviation Administration 2022). The financial health of carriers in this industry is thus crucial for a strong and robust economy. In order to maintain their financial health, airlines focus on different ways to increase efficiency. Capacity utilization, a frequently used measure to increase efficiency in the airline industry, is considered to be "the most significant aspect of efficiency-oriented competition in the airline industry" (Ramaswamy et al. 1994, p. 72).

Capacity utilization, also known as passenger load factor (Dana and Orlov 2014), refers to the ratio of number of passengers to total available seats (Department of Transportation 2024). As load factor is a critical antecedent to airline's financial health, extant airline research has examined the impact of load factor on financial performance from various perspectives (Appendix A). However, there still remains a need for further investigation on this topic. First, the extant research presents mixed findings. Some studies suggest load factor positively contributes to airline financial performance due to increased capacity utilization (Behn and Riley 1999; Tsikriktsis 2007), while other studies show load factor having either no impact (Belobaba 2005) or a negative impact on airline financial performance (Collins et al. 2011). Second, much of the previous research have investigated the relationships using traditional econometrics approaches such as ordinary least squares (Behn and Riley 1999; Shaffer et al. 2000), fixed effects (Ramdas and Williams 2008; Sim et al. 2010; Atkinson et al. 2013), and random effects models (Saranga and Nagpal 2016; Zou and Chen 2017). However, these afore-mentioned methods, when applied on longitudinal data, have been proven to produce biased statistical estimates (Bells and Jones 2015). In the wake of the post-Covid urgency to improve airline financial performance (Cole 2023), providing correct statistical causal inference for airline practitioners becomes paramount. As such, we adopt a recent methodological advancement – a between-within modeling specification – to delineate the relationship between load factor and airline financial performance. The between-within specification adopted in the current study distinguishes *cross-sectional between-carrier differences* from *longitudinal within-carrier changes* (Bell and Jones 2015; Hoffman 2015). This approach is critical as failing to consider both between and within effects in longitudinal data "can lead to biased results and potentially incorrect conclusions" (Hoffman and Stawski 2009, p. 119). To this end, our first research objective is to utilize a recent methodological development to re-examine the relationship between load factor and airline financial performance, hoping to provide more accurate findings for airline practitioners and policymakers as well as introducing the new method to airline researchers to increase research rigor.

In reviewing related airline research, we observe that all previous research dive directly into econometrics analysis without providing significant theorization for the respective econometrics approaches adopted. We attempt to bridge this gap in the current study by conducting two thought experiments (Haans et al. 2016) that correspond to our proposed between-within methodological specification. We first provide a "between theorization" that corresponds to the between-effect in our longitudinal data. Our "between theorization" draws on the concept of slack and explores how different airlines with different levels of load factors should expect different financial outcomes. We then conduct a "within theorization" that corresponds to the within-effect in our data. Our "within theorization" draws on the theory of Performance Frontier (Schmenner and Swink 1998) to examine the impact of load factor on financial performance when a single carrier changes its load factor over time. As a result, both thought experiments lead to the same prediction: the relationship between load factor and airline financial performance is an inverted U shaped relationship. This is our second research objective – bridging the gap between theorization and statistical method in airline research, hoping to reconcile the mixed findings in the extant literature as well as call for airline researchers to conduct more relevant thought experiments to solve operations problems in the airline industry.

To test our hypothesized relationship, we collect data from Department of Transportation (DOT 2024). Compiling 16 years of DOT data from 2004 to 2019 and using operating profit over operating revenue to proxy airline financial performance, we find that the effect of load factor on airline financial performance indeed demonstrates an inverted U-shaped relationship, and the profit-maximizing load factor is 88%. To make sure we provide pertinent and accurate managerial insights, we conduct several robustness tests. First, we specify our model using different covariance structures to test our model fit. Next, we establish the causality between load factor and financial performance using Granger Causality test. Finally, we test the inverted-U shaped curve following a three-step process proposed by Haans et al. (2016). All our robustness tests still support the inverted U-shaped relationship between load factor and airline financial performance.

By re-examining the relationship between load factor and airline financial performance, our research makes distinct theoretical, methodological, and managerial contributions. Theoretically, we align between-within theorization with the empirical data that shows both between effects and within effects (more in Section 4.2), answering the call of Haans et al. (2016) that theoretical reasoning should be aligned with empirical data. To this end, our between-within thought experiments contributes to expanding the scope of theorizing using longitudinal data in operations management field. In addition, our between-within theorization also aligns with our between-within methodological specification. To this end, we echo Haans et al.'s (2016) call that theorizing and model selection should be a seamless integrated process rather than two separated processes. Next, our between-within specification uncovers the inverted U-shaped relationship between load factor and financial performance, providing more nuanced findings and partially explaining the mixed findings in the extant literature. Lastly, in delineating the relationship between load factor and financial performance, we rely on the concept of slack and the theory of Performance Frontier to explore our thought experiments. Our results subsequently support these two operations concepts. Therefore, our study also extends the understanding and application of these two operations concepts into the airline industry, contributing to knowledge accumulation in related airline literature.

Methodologically, we adopt a between-within model specification to analyze the relationship between load factor and airline financial performance. To the best of our knowledge, our study is the first in airline research to implement between-within modeling to test hypotheses. In a model fit comparison (more in

Section 5.1), between-within specification shows higher explanatory power (measured by R^2) as well as higher forecasting accuracy (measured by MSE) compared with fixed effects and random effects models that are widely used in airline research. In this regard, we remind researchers who conduct longitudinal analysis to use fixed effects and random effects models with caution as these two methods are prone to produce biased estimates for longitudinal analysis (Bells and Jones 2015). Therefore, we call for researchers to consider adopting between-within specification to increase research rigor as well as providing more accurate managerial insights for practitioners. Our study is also the first in airline research to adopt the three-step approach recommended by Lind and Mehlum (2010) and Haans et al. (2016) to test for the presence of an inverted U-shaped curve. We also encourage researchers to follow the three-step process to test for U-shaped relationships for other operations and supply chain studies.

For airline practitioners, our findings of the inverted U-shaped curve between load factor and airline financial performance provides a key insight for managers to design operations strategies. As financial performance starts to deteriorate for a load factor above 88%, we do not recommend mangers to target a load factor above 88%. Our findings also compliment another intensively researched topic – overbooking (Rothstein 1985; Suzuki 2002; He 2019). Overbooking generally leads to a higher load factor and airlines practice overbooking extensively to increase revenues. However, our analysis shows that too much overbooking, if leading to a load factor beyond 88%, may not be a good practice after all as overbooking normally results in rejecting passengers with confirmed seats by paying higher compensation to them – the effect of which may partially explain why financial performance declines beyond the load factor of 88%.

2. THEORY AND HYPOTHESES DEVELOPMENT

We first define terminologies for our current study. Following extant airline literature, we use passenger load factor to proxy capacity utilization (Dana and Orlov 2014). DOT defines load factor as "revenue passenger miles divided by available seat miles"; whereas revenue passenger miles are the "summation of the products of the revenue aircraft miles on each inter-airport segment multiplied by the number of revenue passengers carried on that segment" and available seat miles are "the aircraft miles flown in each inter-airport segment multiplied by the number of seats available for revenue passenger use on that segment" (DOT 2024). As such, when we refer to load factor, we are specifically focused on passenger load factor.

2.1 Load Factor and Financial Performance

Financial performance in the airline literature has been operationalized in a variety of ways, such as return on assets (Ramaswamy et al. 1994), operating profit over operating revenue (OPOR) (Tsikriktsis 2007), and profitability (Collins et al. 2011; Zou and Chen 2017). There are different findings regarding the impact of load factor on financial performance. Some research reveals that load factor positively contributes to airline financial performance. For example, Wyckoff and Maister (1977) found a 1% difference in load factors could lead to as high as a 5% difference in profitability. Ramaswamy et al. (1994) found 5% greater load factor translated into a 7% greater return on assets. This positive impact is generally explained by increased capacity utilization as Behn and Riley (1999) found that load factor is positively associated with operating income. Further, Tsikriktsis (2007) concluded a 1% increase in passenger load factor resulted in a 0.63% increase in OPOR. Zou and Chen (2017) also found that higher capacity utilization had positive effects on carrier's profitability. Alternatively, other research has found different, even contradictory, results. Belobaba (2005) analyzed DOT data from 2001 to 2004 and concluded that high load factors did not improve revenues. Collins et al. (2011), analyzing 14 top carriers from 1996 to 2008, found that load factor negatively contributed to carrier's profit margin using both quarterly and annual data.

To debunk the mixed findings of the relationship between load factor and financial performance as well as to "bring further clarity and transparency to theoretical development" (Haans et al. 2016, p.1180), we employ two thought experiments (Hempel 1965; Weick 1989; Deephouse 1999). The first thought experiment applies to different carriers who have high, moderate, and low level of load factors respectively. This thought experiment is termed "between theorization" by Haans et al. (2016). The second thought experiment, termed "within theorization" (Haans et al. 2016), applies to a single carrier who seeks to increase its load factors over time. Theoretically, "the two thought experiments should lead to the same prediction" (Haans et al. 2016, p.1180) if the underlying theories are solid.

We leverage the concept of slack to explore the first thought experiment of between theorization. Slack "conveys the notion of a cushion of excess resources available in an organization" that helps solve internal problems as well as pursue external goals (Bourgeois 1981, p. 29). The relationship between slack and firm performance has been explored in different disciplines. Mishina et al. (2004) investigated 112 manufacturing firms and concluded that more slack is "not necessarily better for growth" (p. 21) and can sometimes even hinder firm growth. Yu (2016) explored the impact of physical capacity utilization on actual and long run minimum costs in 13 international low-cost airlines, concluding that it is better for carriers to bear some idle capacity rather than to operate at full capacity. Tan and Peng (2003) examined the relationship between slack and firm financial performance across 120 firms, finding that this relationship is curvilinear such that too much or too little slack will negatively impact a firm's financial performance. Further, Tan and Peng (2003) suggest "the right question to ask is not whether slack is uniformly good or bad for performance, but rather, what range of slack is optimal for performance" (p. 1260).

In the operations management field, operational slack is measured in a variety of ways, including excess capacity (Steele and Papke-Shields 1993; Bourland and Yano 1994), days of inventory (Hendricks et al. 2009; Azadegan et al. 2013; Kovach et al. 2015), the ratio of sales to property, plant and equipment (Hendricks et al. 2009; Kovach et al. 2015), and cash to cash cycle time (Hendricks et al. 2009; Kovach et al. 2015). In our research, we consider excess capacity (i.e., the percentage of empty seats in an aircraft) to represent slack (i.e., high load factor indicates less slack while low load factor means more slack).

Applying the findings from extant literature to the relationship between slack and financial performance among the three different types of carriers, we argue that carriers with high load factors need more resources to cope with the greater number of passengers, thus, incurring additional costs. As a result, carriers with high load factors may expect relatively low financial performance because the additional costs may outweigh the benefits of additional passenger revenues. On the other hand, carriers with low load factors suffer from plenty of unsold empty seats, which translate to lost revenues. Accordingly, carriers with low load factors may also expect relatively low financial performance due to lost revenues. The third type of carriers have moderate level of load factors. Compared with carriers with high load factors, this type of carriers faces reduced number of passengers (hence reduced costs) while still maintaining relatively high revenues. So, its financial performance should be higher than carriers with high load factors. Compared with carriers with low load factors, this type of carriers harnesses higher revenues from higher percentage of seats sold while still maintaining similar operating costs. So, its financial performance should also be higher than carriers with low load factors. In sum, carriers with moderate level of load factors are expected to have high financial performance because they benefit from increased revenues while still maintaining their regular costs.

The second thought experiment within theorization applies to a single carrier as it seeks to increase its load factors. We turn our attention to the concept of asset frontier to support our discussion of this thought experiment. In examining why some manufacturing plants outperform others, Schmenner and Swink (1998) developed the Theory of Performance Frontier, defined as "the maximum performance that can be achieved by a manufacturing unit given a set of operating choices" (p. 108). As indicated by Schmenner and Swink (1998), a performance frontier is made up of an operating frontier (i.e., frontiers formed by choices in plant operations) and an asset frontier (i.e., frontiers formed by choices in plant design and investment). Schmenner and Swink (1998) proposed that if a firm operates close to its asset frontier, the firm will be likely to operate under the law of trade-offs, but if a firm operates away from its asset frontier, the firm will be likely to operate under the law of cumulative capabilities. When a firm operates close to its asset frontier under the law of trade-offs, "no single plant can provide superior performance in all dimensions simultaneously" (Schmenner and Swink 1998; p. 110). When a firm operates away from its asset frontier under the law of cumulative capabilities, it may simultaneously improve its performance in several dimensions (Schmenner and Swink 1998).

We apply the concept of asset frontier to our second thought experiment of within theorization. We define asset frontier as the optimum state of seat utilization and operating frontier as the actual operating load factor. Following the practice of Lapré and Scudder (2004), carriers with higher operating load factors are assumed to operate closer to their asset frontiers while carriers with low operating load factors are assumed to operate away from their asset frontiers. Therefore, if a carrier operates at a high load factor, it will operate under the law of trade-offs, i.e., it cannot increase both its load factor and financial performance simultaneously. In addition, this carrier has to spend "more and more resources … in order to achieve each additional increment of benefit" (Schmenner and Swink 1998, p. 110) to further improve its financial performance by increasing its load factor. At some point, the resources spent may eventually outweigh the synergistic effects garnered from these resources. Hence, increasing load factor will only dampen its financial performance. However, if a carrier operates away from its asset frontier at a lower operating load factor, it will operate under the law of cumulative capabilities. In this case, this carrier will be able to improve both dimensions simultaneously, i.e., increasing load factor will also increase its financial performance. Both thought experiments suggest that carriers with moderate level of load factors can expect higher financial performance compared with carriers having high and low level of load factors. Hence:

H1: The effect of increasing load factor on carrier's financial performance demonstrates an inverted U-shaped relationship.

2.2 The Differentiating Role of Low-Cost and Legacy Carriers

Airlines are typically classified into two groups: low-cost carriers (LCC) and legacy carriers (Rupp et al. 2006; Atkinson et al. 2013; Yimga 2017). LCCs are also referred to as focused carriers (Mellat-Parast et al. 2015) or geographic specialists (Lapré and Scudder 2004). LCCs generally target price sensitive customers by flying point to point within limited geographic areas with fewer aircraft types (Mellat-Parast et al. 2015). Legacy carriers are sometimes referred to as network carriers (Collins et al. 2011), full-service carriers (Tsikriktsis 2007), non-focused carriers (Mellat-Parast et al. 2015), and geographic generalists (Lapré and Scudder 2004). Legacy carriers operate a wider variety of flights by both geography and aircraft type.

The research of LCC's role on financial performance has also yielded mixed findings in the literature. Collins et al. (2011) showed that legacy carriers tend to achieve more persistent profit margins and asset turnover ratios than LCCs. Mantin and Wang (2012) additionally observe that the profitability of legacy carriers improved faster than that of LCCs after 9/11. However, Tsikriktsis (2007) concluded that LCCs outperformed the rest of the industry in terms of profitability by having more focused resources through limited point to point network operations. Given the mixed findings in the literature, we continue to explore the differentiating role of LCCs and legacy carriers regarding the proposed inverted U-shaped relationship between load factor and airline financial performance. We draw on another management concept of "dynamic capability" to theorize the differentiating effect between the two groups of carriers.

Airline financial performance is highly susceptible to external shocks, such as 9/11 event, fuel cost fluctuations, and the recent pandemic (Mantin and Wang, 2012; DOT 2024). To manage the impact of external shocks, airlines have developed idiosyncratic coping schemes where we lean on the concept of "dynamic capability" to explain our reasoning. Teece et al. (1997) define the concept of dynamic capabilities as "the firm's ability to integrate, build, and reconfigure internal and external competencies to address rapidly changing environments" (p. 516). At the core of dynamic capability is the organization's ability to reconfigure its existing resources to deal with external shocks. Since low-cost carriers have a point-to-point network and highly rigid resources, such as very limited types of aircrafts and short-haul routes, we argue that when external shocks are presented, low-cost carriers will find it more difficult to rapidly reconfigure their resources to address the new challenges, resulting in inferior organizational capabilities. In contrast, legacy carriers, utilizing hub-and-spoke networks, boast a wide variety of highly fungible resources, such as different types of aircrafts and routes, which may give them more flexibility to rapidly reconfigure their resources to overcome the shocks, leading to better organizational capabilities. It has also been established in supply chain management field (Morash 2001; Daugherty et al. 2009) that better organizational capabilities lead to better organizational performance. Additionally, compared with the highly resource-fungible legacy carriers, resource-rigid low-cost carriers may experience slower pace when redeploying their resources to enhance their organizational capabilities due to the resource rigidity – i.e., the learning curve will be flatter for low-cost carriers (Lieberman 1987). Combined, we posit:

H2: Low-cost carriers will demonstrate a more flattened inverted U-shaped relationship between load factor and financial performance.

3. DATA

We collect data from DOT which requires U.S. carriers with "at least one percent of total domestic scheduled-service passenger revenues" to report a variety of performance measures. Research using DOT airline data has been published in various disciplines, including operations management (Lapré and Scudder 2004), management (Schefczyk 1993), and economics (Atkinson et al. 2013).

Financial performance measures and other relevant measures were retrieved from various DOT schedules and DOT Air Travel Monthly Consumer Report. DOT financial reports are a conglomerate of six regions: Atlantic, Domestic, International, Latin America, Pacific, and System while Monthly Consumer Report consists of U.S. Domestic only. Only domestic financial figures were included in our dataset to match the domestic data in the Monthly Consumer Report. Financial measures are reported to DOT quarterly while other measures are reported to DOT either monthly or quarterly. For those measures reported monthly, we aggregate them to a quarterly format to match financial measures.

3.1 Airlines

To avoid the impact of DOT's report format change in October 2003, the impacts of 9/11, and the most recent impact of the global pandemic, we elect to choose our data starting point as the first quarter of 2004 and our data ending point as the last quarter of 2019. After merging data and tracking the name changes of some airlines, our data consists of 28 carriers from 2004Q1 to 2019Q4. Some carriers span the entire 64 quarters while others report fewer quarters either due to their revenues falling below the one percent reporting threshold or due to merger and acquisition. A detailed summary of airlines used in our analysis is presented in Table 1.

Table 1 Airlines in the Dataset

No.	Airline	First quarter in the sample	Last quarter in the sample	Total quarters in the sample
1	AIRTRAN	2004 Q1	2011 Q1	33
2	ALASKA	2004 Q1	2019 Q4	64
3	ALEEGIANT	2018 Q1	2019 Q4	8
4	ALOHA	2006 Q2	2008 Q1	8
5	AMERICA WEST	2004 Q1	2005 Q4	8
6	AMERICAN	2004 Q1	2019 Q4	64
7	ATA	2004 Q1	2006 Q4	12
8	ATLANTIC SOUTHEAST	2004 Q1	2011 Q4	32
9	COMAIR	2004 Q1	2010 Q4	28
10	CONTINENTAL	2004 Q1	2011 Q4	32
11	DELTA	2004 Q1	2019 Q4	64
12	ENDEAVOR	2010 Q4	2019 Q4	13
13	ENVOY	2004 Q1	2019 Q4	56
14	EXPRESSJET	2004 Q1	2019 Q4	62
15	FRONTIER	2005 Q2	2019 Q4	59
16	HAWAIIAN	2004 Q1	2019 Q4	64
17	INDEPENDENCE	2004 Q1	2005 Q4	8
18	JETBLUE	2004 Q1	2019 Q4	64
19	MESA	2006 Q1	2019 Q4	40
20	NORTHWEST	2004 Q1	2009 Q4	24
21	PSA	2018 Q1	2019 Q4	8
22	REPLUBLIC	2018 Q1	2019 Q4	8
23	SKYWEST	2004 Q1	2019 Q4	64
24	SOUTHWEST	2004 Q1	2019 Q4	64
25	SPIRIT	2015 Q1	2019 Q4	20
26	UNITED	2004 Q1	2019 Q4	64
27	US AIRWAYS	2004 Q1	2015 Q2	41
28	VIRGIN AMERICA	2012 Q1	2017 Q4	24

Notes:

1. *RU was used from October 2003 to June 2006 by DOT to code ExpressJet. Effective July 2006, ExpressJet changed in DOT report from RU to XE. In our dataset, RU was changed to XE.*
2. *American Eagle Airlines changed to Envoy effective April 2014 in DOT report. Both Envoy and American Eagle were coded as ENVOY in our data.*
3. *Atlantic Coast Airlines changed to Independence Airline since 2004 November in DOT report. Both airlines were coded as Independence in our data.*
4. *Endeavor Air, formerly Pinnacle Airlines, was ranked for the first time in January 2013 DOT report. Both Pinnacle and Endeavor were coded as Endeavor in our data.*
5. *Atlantic Southeast (EV) was acquired by ExpressJet and changed to XE since January 2012 in DOT report.*
6. *Low-Cost Carriers: Allegiant, Frontier, JetBlue, Southwest, Spirit, and Virgin America.*

3.2 Dependent Variable

Our dependent variable is financial performance. In the airline literature, three categories of measures are commonly used to assess airline financial performance: absolute measures, predicted values, and relative measures. Absolute measures take the form of profitability (Kalemba and Campa-Planas 2017). Predicted values are termed "abnormal returns" in a variety of disciplines (Ramdas et al. 2013), while relative measures are calculated either as operating profit over operating cost (Steven et al. 2012) or as operating profit over operating revenue (Tsikriktsis 2007; Mellat-Parast et al. 2015).

We elect to use operating profit over operating revenue (OPOR) instead of profitability as our financial performance measure for two reasons. First, profitability varies across years and is sometimes negative. When natural logarithms are calculated, those negative profitability values become missing data points, which is not a true reflection of airline financial status. Second, the excessive variance of profitability comes from different sized carriers. Ratio measures like OPOR, in this case, can better account for the size differences among carriers in comparison to other financial measures (Dresner and Xu 1995). In addition, ratio measures can also overcome the difficulty in measures associated with carriers that own aircraft versus carriers that lease aircraft (Tsikriktsis 2007; Mellat-Parast et al. 2015). Operating profit and operating revenue data were retrieved from DOT Schedule P1.2.

3.3 Independent Variables

Our main independent variable is load factor. DOT's definition of load factor (i.e., revenue passenger miles divided by available seat miles) was adopted. Relevant data were retrieved from DOT Schedule T1. The data were in a monthly format and subsequently collapsed into a quarterly format.

There are three different forms of load factors in our models: the between-effect load factor, the quadratic term of the between-effect load factor, and the within-effect of load factor. We denote each carrier by *i* and the measurement occasions by *t*. The between-effect load factor was calculated by taking the group mean of load factor for each carrier, denoted as $\overline{Load\ Factor}_i$ in our models. The square term of the between-effect of load factor is constructed to investigate the non-linear relationship, denoted as $\overline{Load\ Factor^2}_i$. The within-effect of load factor was calculated by subtracting each carrier's load factor from its mean (i.e., $Load\ Factor_{it} - \overline{Load\ Factor}_i$). This variable is denoted as $Load\ Factor\ Within_{it}$ in our models. More explanation of the between-within variable construction follows in Section 4.1.

We include LCC in our models as another independent variable. In the latest DOT report, six carries were classified as LCCs: Allegiant Air, Frontier, JetBlue, Southwest, Spirit, and Virgin America. These six carriers were assigned a value of 1 and the remaining carriers a value of 0.

3.4 Control Variables

Fuel Cost

Fuel is an important control variable for airline research given it is a key factor affecting airlines' financial performance. Typically, fuel has been either examined as a price or a cost (Ramdas et al. 2013; Dana and Orlov 2014). Since some carriers hedged their fuel requirements in order to leverage a fuel price below the market rate, we consider fuel cost to be a more appropriate control variable. To demonstrate this difference, the average fuel price fluctuated from \$111.8 per barrel (2012) to \$44.6 per barrel (2016) in our sample data period (IATA 2018). In the second half of 2005, Southwest hedged 85% of its fuel requirements at the equivalent of \$26 while the industry average was \$72.35 (Alexander 2006). Fuel cost was reported on a monthly basis in Schedule P12(a), which was aggregated to quarterly format.

Number of Enplaned Passengers

The number of enplaned passengers was included as a control variable because carriers can expect increased revenue as the number of passengers increases, which may accordingly help boost profit. We retrieve the number of passengers from DOT Monthly Consumer Report.

Market Share

Market share is a common control variable in airline research, and it has been operationalized in different ways (Behn and Riley 1999; Suzuki 2000; Rupp et al. 2006). We adopt Shaffer et al.'s (2000) method because this operationalization calculates market share at the carrier level, consistent with our unit of analysis. We derive a quarterly measure by calculating the ratio of a carrier's quarterly revenue passenger miles to the sum of revenue passenger miles of all 28 carriers in that quarter. Revenue passenger miles were retrieved from DOT Schedule T1.

Firm Size

As larger firms are expected to "have higher levels of resources and more developed market positions, it is important to control for the size of the firm" (Mishina et al. 2004, p. 1189). Because our dependent variable is operating profit over operating revenue, and our load factor is revenue passenger miles over available seat miles, we use the number of employees as a proxy for firm size to avoid collinearity. More employees add costs and negatively impact operating profit as employee expenses are the second largest element impacting carrier's profit after fuel cost (IATA 2018). The number of employees was taken from DOT Schedule P1(a).

Delays

Delays, either airborne or ground, result in significant costs to carriers, impacting carriers' financial performance (Hansen et al. 2001; Cook et al. 2012). As such, we include delays as a control variable in our models. Delay data were compiled from DOT Air Travel Monthly Consumer Report.

Yield

Yield management in the airline industry refers to the techniques used to allocate limited seats to different customer segments by offering differential airfares (Netessine and Shumsky 2002). Using proper yield management systems, American Airlines credited a revenue increase of $500 million per year and Delta Airlines generated additional revenues of $300 million per year (Boyd 1998). Following current airline literature (Alan and Lapré 2018), we calculate yield as passenger revenues divided by revenue passenger miles. Yield data were obtained from DOT Schedule T1 and P1.2.

Fleet Utilization

Higher fleet utilization was considered a key driver to Southwest's success (Gittell 2003). Southwest Airline boasts an impressive 20-30 minutes turnaround time while the industry average is 1.5-2 hours (Mantin and Wang 2012). Belobaba (2009) also recommends airlines that improve fleet utilization achieve better financial performance. In addition, Baltagi et al. (1998) indicate that load factor only measures occupied seats relative to total miles flown, so it ignores the utilization of the aircraft itself. Accordingly, we include fleet utilization as another control variable. We calculate fleet utilization as block hours divided by aircraft days following Alan and Lapré (2018). Relevant data were retrieved from DOT Schedule P52.

Macro-Economic Factors

One of the most challenging problems empirical studies face in economic research is endogeneity bias (Scotti and Dresner 2015; Wooldridge 2010). Changes in financial performance could be driven by other macro-economic factors, which, if not controlled for, could bias the results (Wooldridge 2010). Therefore, we also include two macro-economic variables previously used in airline research to further address the concern of endogeneity bias. The first variable is the Smoothed U.S. Recession Probabilities (Piger and Chauvet 2019), retrieved from Federal Reserve Economic Data. The second variable is percentage change of GDP (Bureau of Economic Analysis 2024), also retrieved from Federal Reserve Economic Data. Lastly, we retrieve and compile data from DOT to calculate the quarter-over-quarter change of averaged domestic fares from 2004 to 2019.

Year and Quarter Effects

Finally, we add year and quarter effects to avoid any additional omitted variable bias as well as to capture the time trends in our observed data period.

3.5 Summary Information and Statistics

Table 2 provides a summary of the variables used in our analysis, including variable names, a description of how variables were constructed, and data sources. Table 3 provides the correlation matrix for OPOR with load factor and other control variables. We report correlations for each variable constructed in both between and within effects in Table 3, as explained in Section 4.1.

Table 2: Variables Used in Analysis

Variable	Formula	Data Source
OPOR	Operating profit divided by operating revenue at quarterly level	DOT Schedule P1.2
Load Factor	Quarterly revenue passenger miles divided by available seat miles	DOT Schedule T1
Load Factor Between	Group mean of load factor for each carrier	DOT Schedule T1
Load Factor Within	Subtracting each carrier’s load factor from its mean	DOT Schedule T1
Fuel Cost Between	Group mean of quarterly fuel cost for each carrier	DOT Schedule P12(a)
Fuel Cost Within	Subtracting each carrier’s fuel cost from its mean	DOT Schedule P12(a)
Enplaned Passengers Between	Group mean of quarterly enplaned passengers for each carrier	DOT Air Travel Monthly Consumer Report
Enplaned Passengers Within	Subtracting each carrier’s enplaned passengers from its mean	DOT Air Travel Monthly Consumer Report
Market Share Between	The ratio of a carrier's quarterly revenue passenger miles to the sum of revenue passenger miles of the total carriers in that quarter. Take the group mean of each carrier to construct between variables.	DOT Schedule T1
Market Share Within	Subtracting each carrier’s market share from its mean	DOT Schedule T1
Number of Employees Between	Group mean of quarterly number of FTEs for each carrier	DOT Schedule P1(a)
Number of Employees Within	Subtracting each carrier’s number of employees from its mean	DOT Schedule P1(a)
Delay Between	Quarterly sum of delays reported to DOT. Take the group mean of each carrier to construct between variables.	DOT Air Travel Monthly Consumer Report
Delay Within	Subtracting each carrier’s total delay from its mean	DOT Air Travel Monthly Consumer Report
Yield Between	Group mean of passenger revenues divided by revenue passenger miles (RPMs) for each carrier	DOT Schedule T1 and P1.2
Yield Within	Subtracting each carrier’s yield from its mean	DOT Schedule T1 and P1.2
Fleet Utilization Between	Group mean of Block Aircraft Hours divided by Aircraft Days for each carrier	DOT Schedule P52
Fleet Utilization Within	Subtracting each carrier’s Fleet Utilization from its mean	DOT Schedule P52
GDP % Change	Quarter over quarter change in GDP	Federal Reserve Economic Data
Fare % Change	Quarter over quarter change in average domestic fares	DOT DB1B
Recession	Smoothed U.S. Recession Probabilities	Federal Reserve Economic Data
LCC	Low-cost carrier defined by DOT	DOT Air Travel Monthly Consumer Report

Table 3: Summary Statistics and Correlation Matrix

	Panel A			Panel B																							
	N	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1 OPOR	1035	0.04	0.15	1.00																							
2 Load Factor Between	1035	(0.00)	0.04	0.30	0.25	1.00																					
3 Load Factor Between.SQ	1035	(0.00)	0.05	0.29	0.25	1.00	1.00																				
4 Load Factor Within	1035	(0.00)	0.04	0.24	0.13	0.03	0.03	1.00																			
5 Load Factor Between*LCC	1035	0.00	0.01	0.05	(0.17)	0.45	0.47	0.01	1.00																		
6 Fuel Cost Between	1014	0.00	1.43	0.08	0.02	0.32	0.33	0.02	(0.15)	1.00																	
7 Fuel Cost Within	974	0.00	0.69	0.04	(0.05)	-	-	0.04	-	-	1.00																
8 Enplaned Passengers Between	1035	0.00	0.91	0.09	(0.07)	0.23	0.22	0.02	(0.37)	0.75	-	1.00															
9 Enplaned Passengers Within	1035	(0.00)	0.26	0.13	0.02	(0.03)	(0.03)	0.42	(0.01)	(0.05)	0.18	(0.01)	1.00														
10 Market Share Between	1035	(0.00)	0.07	(0.01)	0.00	0.29	0.29	0.02	(0.30)	0.77	-	0.89	(0.02)	1.00													
11 Market Share Within	1035	0.00	0.02	0.16	0.11	(0.00)	(0.00)	0.32	(0.00)	(0.01)	0.19	(0.00)	0.60	(0.00)	1.00												
12 Employees Between	1035	(0.00)	1.09	0.01	(0.05)	0.19	0.19	0.01	(0.39)	0.79	-	0.96	(0.00)	0.94	(0.00)	1.00											
13 Employees Within	1035	(0.00)	0.22	0.04	0.02	0.01	0.01	0.15	0.00	(0.04)	0.32	0.01	0.80	0.00	0.62	0.00	1.00										
14 Carrier Delay Between	1035	0.00	0.93	0.08	(0.28)	(0.21)	(0.23)	(0.01)	(0.47)	0.39	-	0.79	0.02	0.56	0.00	0.70	0.00	1.00									
15 Carrier Delay Within	1035	(0.00)	0.37	0.02	(0.49)	(0.04)	(0.04)	0.02	(0.01)	(0.04)	0.13	(0.02)	0.58	(0.02)	0.24	(0.01)	0.47	0.02	1.00								
16 Yield Between	1035	0.00	0.28	0.17	(0.00)	(0.54)	(0.54)	(0.03)	(0.29)	(0.35)	-	(0.40)	(0.00)	(0.55)	0.00	(0.45)	(0.04)	0.06	0.02	1.00							
17 Yield Withiin	1035	(0.00)	0.19	0.15	0.00	(0.08)	(0.08)	0.12	(0.02)	(0.02)	0.50	(0.05)	0.01	(0.06)	0.17	(0.05)	0.21	0.02	(0.00)	0.11	1.00						
18 Fleet Utilization Between	1035	(0.00)	0.21	(0.00)	(0.01)	0.46	0.47	0.06	0.38	0.46	-	0.24	(0.05)	0.31	(0.01)	0.23	0.01	(0.17)	(0.07)	(0.50)	(0.14)	1.00					
19 Fleet Utilization Within	1035	(0.00)	0.18	0.02	(0.11)	(0.13)	(0.14)	(0.10)	(0.03)	(0.05)	0.17	(0.05)	0.01	(0.07)	(0.10)	(0.03)	(0.03)	0.05	0.09	0.12	0.08	(0.22)	1.00				
20 GDP % change	1035	0.28	0.58	0.03	0.08	0.01	0.01	0.10	0.03	(0.00)	(0.06)	(0.00)	0.14	0.01	0.10	(0.00)	0.07	(0.02)	0.03	(0.04)	(0.07)	0.05	0.03	1.00			
21 Fare % change	1035	(0.00)	0.03	(0.12)	(0.03)	(0.02)	(0.02)	0.04	(0.02)	(0.00)	0.07	(0.00)	(0.05)	(0.01)	(0.03)	(0.00)	(0.06)	0.01	(0.03)	0.03	(0.06)	(0.02)	0.02	0.31	1.00		
22 Recession	1035	0.11	0.31	(0.10)	(0.10)	(0.04)	(0.04)	(0.15)	(0.04)	0.00	0.11	(0.00)	(0.16)	(0.01)	(0.15)	0.00	(0.08)	0.03	(0.05)	0.05	0.09	(0.08)	0.01	(0.68)	(0.15)	1.00	
23 LCC	1008	0.23	0.42	0.15	(0.14)	0.28	0.29	0.01	0.55	0.13	-	(0.00)	(0.01)	(0.13)	(0.00)	(0.14)	0.00	(0.06)	(0.02)	(0.03)	(0.04)	0.52	(0.05)	0.03	(0.02)	(0.07)	1.00

4. ANALYSIS AND RESULTS

Before we present the steps taken to build our models and test hypotheses, we briefly discuss the statistical basis for our between-within specification.

4.1 Between-within Specification

Estimation methods in the extant airline literature have evolved from OLS to fixed effects approaches, then to random effects approaches for estimating parameters using panel data. While a fixed effects approach had been described as the gold standard for modelling panel data (Schurer and Yong 2012), Bell and Jones (2015) contend that controlling heterogeneity "cuts out much of what is going on in the data" causing misleading results interpretations (p. 134). To solve the problems associated with fixed effects modeling, random effects modeling (also referred to as multilevel modeling, hierarchical linear modeling, or mixed models) is preferable. Random effects modeling has the following advantages: 1) it accounts for differences between groups (carriers in our case) by partitioning variance between them; 2) slopes of different groups are allowed to vary at different magnitudes; and 3) variances at measurement occasion level can also be modelled, allowing specifics of occasion level to be retained in the model while still having the ability for generalization (Bell and Jones 2015; Hoffman 2015).

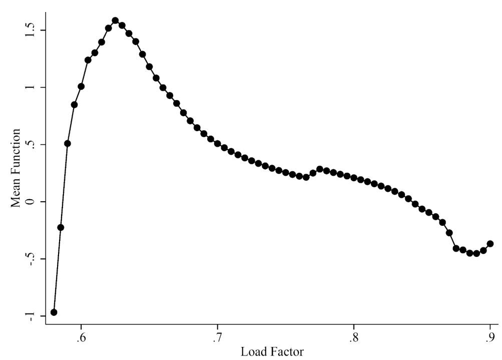
To correctly specify a random effects model in a panel dataset, two new variables for a time varying covariate should be constructed: one variable accounting for the between effect using the mean (Mundlak 1978) and another variable accounting for the within effect using the deviation from the mean (Berlin et al. 1999; Bells and Jones 2015). Time varying covariates (i.e., load factor in our case) contain two parts: one part specific to higher-level entities (i.e., carriers in our case) which does not change between measurement occasions and a second part that changes over time representing the differences between measurement occasions (Bell and Jones 2015; Hoffman 2015). These two parts, called "between" and "within" effects, respectively, have different effects in a model (Bell and Jones 2015). Correctly specifying a random effects model using between and within effects is critical because "[f]ailure to explicitly consider separate between- and within-person sources of variation when modelling repeated measures data can lead to biased results and potentially incorrect conclusions about within-person relationships over time" (Hoffman and Stawski 2009, p. 119).

Reviewing the random effects models adopted in extant airline research, we see two deficiencies here: 1) between and within effects were not constructed separately (Saranga and Nagpal 2016; Zou and Chen 2017); 2) Hausman test was used to conclude that random effects is preferred over fixed effects (Saranga and Nagpal 2016). But "Hausman test is not a test of fixed effect versus random effect; it is a test of the similarity of within and between effects" (Bell and Jones 2015, p. 144). Our research utilizes a between-within specification to further improve on the random effects approach in extant airline research. Since our panel data is hierarchically constructed (i.e., it consists of repeated measures over time t nested within multiple carriers i), between-within specification can be readily applied. We follow Bell and Jones (2015) and Hoffman (2015) to construct our variables in their respective between and within effects (Table 2) to fit our models. The between variable corresponds to our between theorization while the within variable corresponds to our within theorization in our hypotheses development section.

4.2 Methodology

Before fitting the data with a between-within specification, we performed two pre-modeling robustness checks as suggested in the literature (Haans et al. 2016). First, given we hypothesize an inverted U-shaped relationship between load factor and carrier's financial performance, we fit a nonparametric regression following Haans et al. (2016) and Aghion et al. (2005). Nonparametric regression allows "researchers to explore the shape of the relationship by imposing as little structure on the functional form as possible, before proceeding to test their hypotheses in a confirmatory manner with a restrictive quadratic specification" (Haans et al. 2016, p. 1183). We plot the nonparametric regression result in Figure 1. From the plot, we observe an inverted U-shaped relationship between load factor and OPOR, providing an initial support to the presence of an inverted U-shaped relationship.

Figure 1: Plot of Predictive Margins of Load Factor from Nonparametric Regression with 95% CI



Second, we calculate Intra Class Correlation (ICC) by fitting a random intercept model. This was conducted to confirm that the data is longitudinal in nature (Hoffman 2015) as well as to ensure the variations we observe over the 16 years are not random fluctuations but indeed "meaningful individual differences" (Bliese and Ployhart 2002, p. 368). ICC results are summarized in Table 4 with the ICC being 0.62, indicating that 62% of the variation in financial performance (OPOR) is between carriers and the remaining 38% of the variation is within carriers. The ICC strongly indicates the nature of longitudinal data, validating our between-within theorization.

Table 4: ICC Calculation

	Parameter	OPOR
Fixed Effects		
Intercept	β_0	0.004
		(0.13)
Random Effects		
Level 2: Carriers	σ^2_{u0}	0.0214
Level 1: Occasion	σ^2_{e0}	0.0129
Measures of Fit		
-2 Log Likelihood		-1456.20
ICC		0.62

*Notes: † = p < 0.10; * = p < 0.05; ** = p < 0.01 (two-tailed).*
Z-tests are reported in parentheses for the fixed effects parameters.

The full specification of our model is illustrated in Equation 1, following Bell et al. (2018), while the R square calculation (MVP) is presented in Equation 2 following Nakagawa and Schielzeth (2013). Individual carriers are denoted as *i* (level 2) which are measured/reported on multiple occasions *t* (level 1). This specification is "able to model both within and between individual effects concurrently, and also explicitly models heterogeneity in the effect of predictor variables at the individual level" (Bell et al. 2018, p. 5).

Equation 1 Full Model Specification

$$Y_{it} = \beta_0 + \beta_1 \overline{\text{Load Factor}}_i + \beta_2 \overline{\text{Load Factor}^2}_i + \sum \mathbf{B}\text{Controls}_{it} + \mu_{it} + \sum \text{Year}_i * \nu_t + \sum \text{Quarter}_i * \nu_t + e_{it}$$

Equation 2 R square (MVP) calculation

$$R^2(\text{MVP}) = \frac{\text{Var}(\hat{Y}_{it})}{\text{Var}(\hat{Y}_{it}) + \sigma_u^2 + \sigma_e^2}$$

4.3 Results

To test our hypotheses, we fit a series of models based on our full model specification in Equation 1 using *Stata/BE 17*. The results are summarized in Table 5. Maximum Likelihood (MLE in *Stata*) was used to estimate and report all model parameters. The significance of fixed effects was evaluated with individual Wald Tests while the significance of random effects was evaluated with likelihood ratio tests. Effect size was evaluated with MVP (i.e., the proportion of model predicted variance to the overall variances) as well as with the traditional Total R^2 (i.e., the squared correlation between the actual dependent variable and the dependent variable predicted by the fixed effects).

Table 5: Results for fitting the mixed effects models

Variables	Parameter	Model 1	Model 2	Model 3	Model 4
Fixed Effect					
Intercept	β_0	-0.03 (-1.21)	-0.02 (-0.71)	-0.015 (-0.70)	-0.01 (-0.47)
$\overline{\text{Load Factor}}$	β_1		**11.87** (2.80)**	**13.14** (3.08)**	12.80** (2.70)
$\overline{\text{Load Factor}^2}$	β_2		**-6.74* (-2.40)**	**-7.83** (-2.75)**	-7.30* (-2.31)
Load Factor_within	β_3		0.20 (0.77)	0.25 (0.95)	0.06 (0.42)
$\overline{\text{Load Factor}}$* LCC	β_4			**12.42 (0.65)**	
$\overline{\text{Load Factor}^2}$* LCC	β_5			**-6.46 (-0.54)**	
$\overline{\text{Fuel Cost}}$	β_6	0.01 (0.69)	0.01 (0.98)	0.005 (0.64)	0.003 (0.31)
Fuel Cost_within	β_7	0.004 (0.70)	0.01 (0.97)	0.006 (0.91)	-0.004 (-0.42)
$\overline{\text{Enplaned Passengers}}$	β_8	0.31** (5.28)	0.14* (2.39)	0.21** (3.28)	0.14* (2.46)
Enplaned Passengers_within	β_9	0.01 (0.35)	0.001 (0.04)	0.003 (0.07)	0.11* (2.51)
$\overline{\text{Market Share}}$	β_{10}	-0.88† (-1.77)	-0.65† (-1.78)	-0.63† (-1.83)	-0.51 (-1.47)
Market Share_within	β_{11}	0.72** (2.74)	0.77** (2.91)	0.76* (2.84)	0.87* (2.46)
$\overline{\text{Total Employees}}$	β_{12}	-0.12* (-2.28)	-0.06 (-1.45)	-0.07* (-1.93)	-0.06 (-1.42)
Total Employees_within	β_{13}	-0.17** (-5.45)	-0.16** (-5.06)	-0.16** (-4.91)	-0.25** (-5.53)
$\overline{\text{Delay}}$	β_{14}	-0.12** (-3.70)	-0.03 (-0.87)	-0.05* (-1.76)	-0.03 (-1.05)
Delay_within	β_{15}	0.01 (1.04)	0.01 (1.10)	0.01 (1.06)	-0.01 (-0.80)
$\overline{\text{Yield}}$	β_{16}	0.17** (2.97)	0.20** (4.71)	0.26** (5.56)	0.21** (4.36)
Yield_within	β_{17}	0.14** (5.69)	0.14** (5.61)	0.14** (5.52)	0.24** (7.29)
$\overline{\text{Fleet Utilization}}$	β_{18}	0.08 (0.88)	0.05 (0.74)	0.03 (0.43)	-0.02 (-0.25)
Fleet Utilization_within	β_{19}	0.04 (1.51)	0.04 (1.48)	0.04 (1.50)	0.02 (1.00)
GDP % change	β_{20}	-0.01 (-1.28)	-0.01 (-1.32)	-0.01 (-1.33)	-0.02** (-2.80)
Fare % change	β_{21}	-0.32* (-2.40)	-0.33* (-2.48)	-0.33* (-2.48)	-0.28* (-2.40)
Recession	β_{22}	-0.02 (-0.90)	-0.02 (-0.83)	-0.02 (-0.81)	-0.02 (-0.63)
LCC	β_{23}	-0.09* (-2.37)	**-0.05† (-1.93)**	-0.08** (-2.74)	-0.02* (-2.40)
Dummy 1	β_{24}				-0.08 (-1.11)
Random Effect					
Level 2: Carriers	σ_{u0}^2	0.0088	0.0023	0.0017	0.0006
Variance (Load Factor)	σ_{u1}^2	1.1010	1.1239	1.0908	
Covariance	σ_{u0u1}	-0.0912	-0.0389	-0.0319	
Level 1: Occasion	σ_{e0}^2	0.0079	0.0079	0.0079	0.0097
	AR(1)				0.4235
Measures of Fit					
-2 Log Likelihood		-1836.02	-1854.59	-1858.73	-1918.11
AIC		-1754.02	-1766.59	-1766.73	-1830.11
BIC		-1553.88	-1571.81	-1542.18	-1615.33
Total R^2		64.38%	64.53%	64.51%	63.92%
R^2 (MVP)		44.15%	56.52%	57.83%	56.29%
ΔR^2 (MVP)			12.37%	1.31%	
$\Delta\chi^2$ (LRT)			18.58**	4.14	

Notes: † = $p < 0.10$; * = $p < 0.05$; ** = $p < 0.01$ (two-tailed).
Z-tests are reported in parentheses for the fixed effects parameters.

Model 1 is our baseline model where we only include the control variables. A likelihood ratio test was conducted between a random intercept model and a random intercept and random slope model, the result of which shows that adding a random slope is a better model fit ($\chi^2 = 92.35$, $p = 0.000$). We report the results from running the random intercept and random slope model.

In Model 2, we add the main variables of interest ($\overline{\text{Load Factor}}$ and $\overline{\text{Load Factor}^2}$). A likelihood ratio test again confirms that allowing the slope of each carrier to vary results in a better model fit (χ^2= 92.35, p=0.000). Since our model consists of two levels, we examined the residual distribution of Model 2 at both Level 2 and Level 1. Figure 2 and Figure 3, respectively, show the Kernel density estimate for Level 2 residuals and Level 1 residuals for model 2. From these plots, we see that the residuals are approximately normally distributed. Therefore, we can be confident that Model 2 is a good fit to the data.

Figure 2: Plot of Kernel Density Estimate of Level 2 Residuals of Model 2

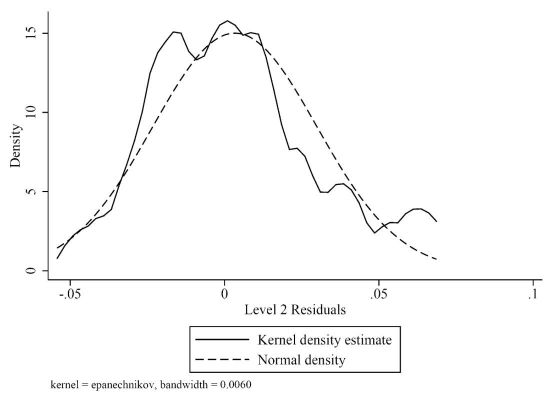
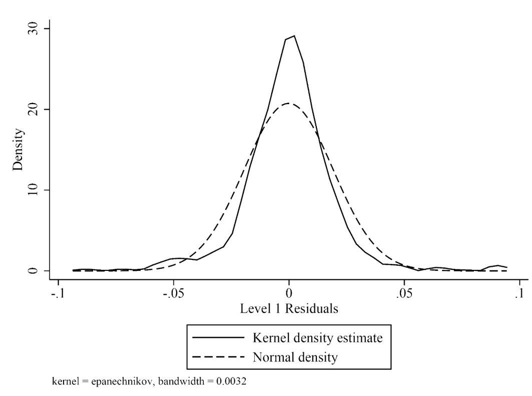


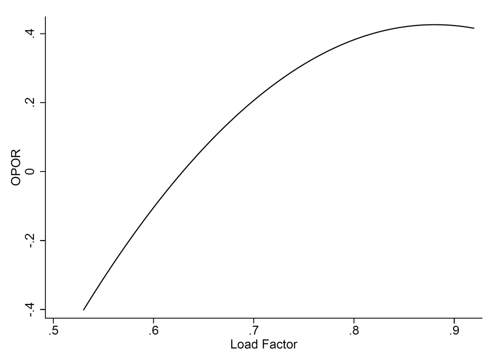
Figure 3: Plot of Kernel Density Estimate of Level 1 Residuals of Model 2



Turning our attention to the model fit, we see that Model 2 explains 56.52% of variation in OPOR, a significant 12.37% increase in R^2 (MVP) compared with Model 1. The reduction of the value of –2Loglikelihood and smaller AIC and BIC values in Model 2 also indicate improved model fit.

Model 2 tests Hypothesis 1. Hypothesis 1 predicts an inverted U-shaped relationship between load factor and airline financial performance. H1 is jointly tested by the coefficients of $\overline{\text{Load Factor}}$ and $\overline{\text{Load Factor}^2}$. The coefficient of $\overline{\text{Load Factor}}$ is positive (β_1=11.87*, p*=0.005) while that of $\overline{\text{Load Factor}^2}$ is negative (β_2=–6.74, *p*=0.016), both of which are statistically significant. To test a true presence of an inverted U-shaped relationship, a three-step procedure was conducted following Haans et al. (2016) and Lind and Mehlum (2010). First, the coefficient of the quadratic term must be "significant and of the expected sign" (Haans et al., 2016, p. 1181). Looking at the coefficient of $\overline{\text{Load Factor}^2}$ (β_2=–6.74, *p*=0.016), we see that this test was passed. Second, "the slope must be sufficiently steep at both ends of the data range" (Haans et al., 2016, p. 1181-1182). The slope for the lower end of our data range (load factor of 52.94%) is 4.06 (*p*=0.000), while the slope for the higher end of our data range (load factor of 92.75%) is –1.05 (*p*=0.09). The second test was also passed. Third, "the turning point needs to be located well within the data range" (Haans et al., 2016, p. 1182). The calculated turning point 88.06% is well within our load factor range (52.94% to 92.75%). With all three tests passed, we can be reasonably confident that there exists an inverted U-shaped relationship, which is plotted in Figure 4. In this plot, all significant control variables were taken at their mean values, and load factor was taken at the range of 52.94% to 92.75% according to our data range. In sum, H1 is supported.

Figure 4: The Effect of Load Factor on OPOR Without Moderation Terms



Hypothesis 2 suggests that the status of being a low-cost carrier moderates the effect of load factor on financial performance. To test the moderation effect in H2, we add two interaction terms in Model 3 following Aiken and West (1991) and Haans et al. (2016): one interaction of load factor ($\overline{\text{Load Factor}}$) with low-cost carrier (LCC), and the other interaction of the square term of load factor ($\overline{\text{Load Factor}^2}$) with low-cost carrier (LCC). This effect was jointly tested by the coefficients of $\overline{Load\ Factor}$, $\overline{\text{Load Factor}^2}$, LCC, and their interaction terms $\overline{Load\ Factor}$*LCC and $\overline{\text{Load Factor}^2}$*$LCC$. We find the coefficients for the main effects of $\overline{\text{Load Factor}}$ (β_1=13.14, p=0.002), $\overline{\text{Load Factor}^2}$ (β_2=–7.83, p=0.006), and LCC (β_{23}=–0.08, p=0.006) still remain significant but the coefficients of their interaction terms $\overline{\text{Load Factor}}$*LCC ($\beta_4$=12.42, p=0.516) and $\overline{\text{Load Factor}^2}$*$LCC$ (β_5=–6.46, p=0.587) are not significant. Hence, Hypothesis 2 was not supported. In addition, the likelihood ratio test between Model 2 and Model 3 also reveals that adding the two interaction terms does not improve Model 3 fit (χ^2=4.14, p=0.126).

After adding the interaction terms, we see that the coefficients of $\overline{\text{Load Factor}}$ and $\overline{\text{Load Factor}^2}$ still remain significant with the expected sign. Hence, there is a need to test if there is a shift in the turning point as the turning point now depends on the moderation (Haans et al. 2016, p. 1187). This was conducted following the procedures of Haans et al. (2016) by assigning the mean value of the variables to the required equation (Equation 11, Haans et al. 2016, p. 1187). Test results show that the equation is not significantly different from zero (p=0.525), indicating that the turning point did not shift after adding the two interactions terms.

4.4 Robustness Test

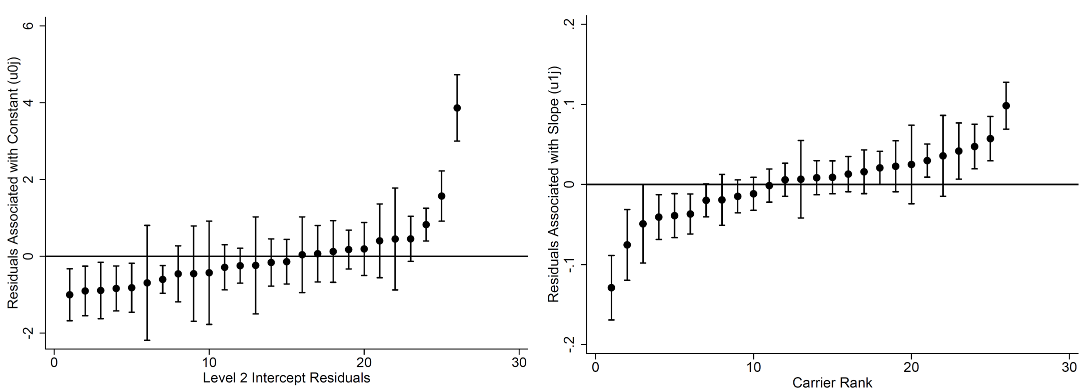
Model Fit Test

Given our data shows within-carrier changes (38% in OPOR), we examine a variety of alternative covariance structures and conduct likelihood ratio tests to see which alternatives fit better (results shown in Table 6). The model that offers better fit uses a random intercept to control for individual carrier differences (unstructured in G matrix) and enables the residuals of one measurement occasion apart to covary (AR1 in R matrix). In addition, one of the key interest areas of airline longitudinal data is how carriers operate differently from each other over time. To this end, we plot the shrunken high-level (level 2) residuals from Model 2 to consider variation between carriers. Left chart of Figure 5 shows one carrier (ATA) having higher differential positive *intercept* than other carriers, while right chart of Figure 5 shows one carrier (ATA again) having lower negative differential *slope* than other carriers. Hence, ATA cannot be considered part of the overall distribution of carriers. Accordingly, ATA was dummied out to have its own differential intercept and slope, while preserving the assumption of normal residuals for those carriers that remain in the random effects as part of a common distribution. We then run our model again with the alternative covariance structure set to AR1 and adding the ATA dummy(Model 4 in Table 5). Judging by the R^2 (MVP) and the Total R^2, Model 4 does not fit better than Model 2. However, we see that Hypothesis 1 still remains valid in Model 4 despite a small change in the coefficients.

Table 6: Testing Alternative Residual Structure Based on Model 2

Random Term	Model	*df*	AIC	BIC	-2 LogLik	Test	L.Ratio	*p* Value
Independent	1	42	-1688.57	-1483.55	-1772.57			
AR (1)	2	43	-1830.96	-1621.06	-1916.96	1 vs 2	144.39	0.000
AR (2)	3	44	-1829.43	-1614.65	-1917.43	2 vs 3	0.47	0.493
AR (3)	4	45	-1831.36	-1611.69	-1921.36	3 vs 4	3.93	0.047
AR (4)	5	46	-1843.30	-1618.75	-1935.30	4 vs 5	13.95	0.000

Figure 5: Plot of Shrunken Residuals (Intercept and Slope) from Model 2 with 95% Confidence Intervals



Granger Causality Test Among Profitability, Load Factor, and Yield

Garvett and Hilton (2002) posit that airlines that seek to maximize financial outcomes should strive to achieve high load factors and high yield at the same time. On one hand, it is easy to achieve very high load factors if airlines keep their yield very low by offering low airfares to attract consumers. On the other hand, if airlines keep their yield very high by excessively raising their fares, they will drive consumers away thus more empty seats would be expected. Neither of these two strategies will lead to profitability (Garvett and Hilton, 2002).

Since our dependent variable is a profitability measure and we have included both load factor and yield in our model, our robustness test boils down to two levels of tests. First, as we are trying to find the optimal profit-maximizing load factor, we need to unveil the causal relationships between 1) load factor and profitability; and 2) yield and profitability. High load factors may lead to higher profitability but if airlines witnessed low profitability, this may trigger them to take actions to increase load factors. The same also holds true between yield and profitability. Second, we need to unveil the causal relationships between load factor and yield to make sure including both variables in our model simultaneously does not interfere with our statistical inferences of our profit-maximizing load factor.

Given the nature of our panel data, we adopt panel vector autoregression (VAR) method to estimate the relevant models. Following the literature (Abrigo and Love 2016; Yetkiner and Beyzatlar 2020), we conduct this test in the following sequential steps. First, we run unit root test for the three variables of interest (i.e., OPOR, load factor, and yield). Since we have unbalanced panels and some of our panels have gaps, the most appropriate method is the Fisher type test. We perform augmented Dickey–Fuller tests (Dickey and Fuller 1979) and all four test results, particularly the inverse normal Z statistic (Choi 2001) rejected the null hypothesis that all panels contain unit roots: OPOR (Z=–15.73, p=0.000), load factor (Z=–12.24, p=0.000), and yield (Z=–8.67, p=0.000). Once the stationarity test was passed, we move on to the next step where we conduct Granger causality test using panel vector autoregression.

To justify load factor and yield as two eligible "explanatory" variables for profitability, we run two separate Granger causality tests without adding any exogenous variables. Before drawing causality inferences, we check the stability condition of our panel VAR estimates using eigenvalue of the fitted model (Abrigo and Love 2016). A VAR model is stable if all moduli eigenvalues are less than one (Hamilton 1994; Lütkepohl 2005). From Figure 6 and Figure 7, we see all eigenvalues in the companion matrix are less than one.

Therefore, we can proceed to draw causality inferences. The result from the stabilized model shows that 1) load factor Granger causes profitability (bottom left graph in Figure 8 where 95% confidence interval does not fully contain the zero line), but profitability does not Granger cause load factor (upper right graph in Figure 8 where 95% confidence interval does contain the zero line); and 2) yield Granger causes profitability (lower left graph in Figure 9), but profitability does not Granger cause yield (upper right graph in Figure 9). Hence, both load factor and yield are perfectly justified as "explanatory" variables for profitability.

Figure 6: Roots of the Companion Matrix for Load Factor and Profitability Model

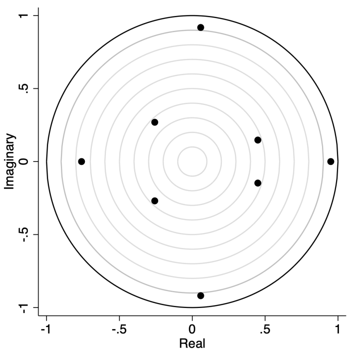


Figure 7: Roots of the Companion Matrix for Yield and Profitability Model

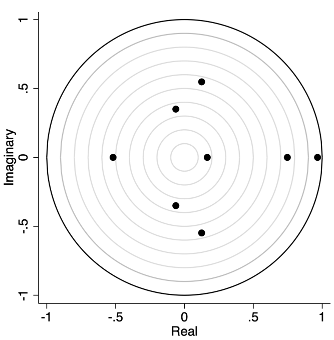


Figure 8: Orthogonalized IRFs Between Load Factor and Profitability

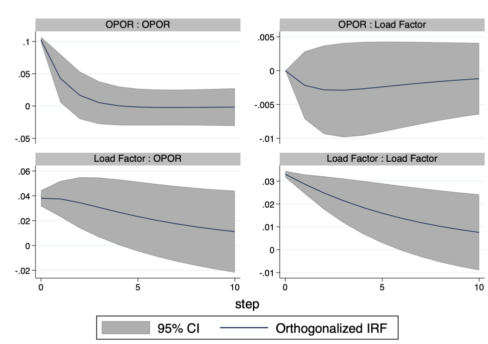
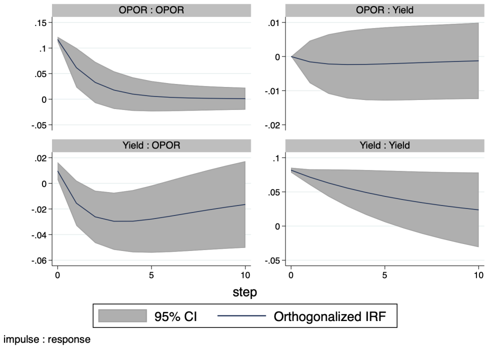


Figure 9: Orthogonalized IRFs Between Yield and Profitability



To test that including both load factor and yield simultaneously in the same model will not interfere with our statistical inference from Model 2, we run Granger causality test by including all other relevant variables from our Model 2. The Granger causality test result from the stabilized VAR model (Figure 10) shows that by controlling for other variables in Model 2, load factor does not Granger cause yield and yield does not Granger cause load factor (Figure 11). This reinforces the assertion that including both variables simultaneously in our models does not affect our inference of the profit-maximizing turning point. As an additional step in this robustness test, we run all Granger causality tests in different lags up to 10, all producing the same conclusions.

Figure 10: Roots of the Companion Matrix for Yield and Load Factor Model

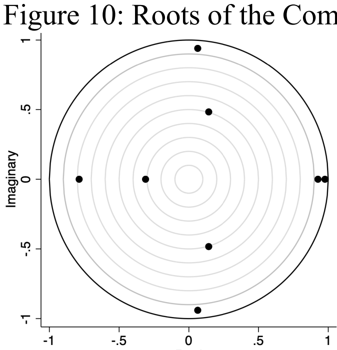
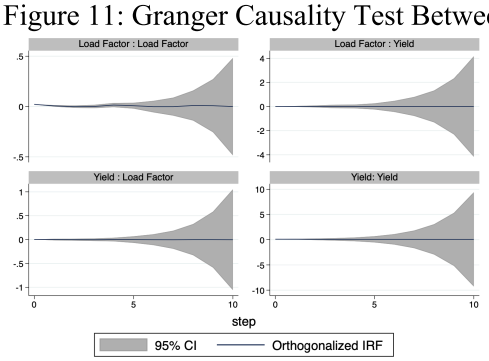


Figure 11: Granger Causality Test Between Load Factor and Yield



U Curve Test

Since our key hypothesis (H1) concerns an inverted U-shaped curve between load factor and financial performance, we conduct additional robustness tests that center around validating the inverted U-shaped curve following the recommendations of Haans et al. (2016). First, we split our data into two different sets at the turning point. Then we run two separate regressions (using Model 2 in Table 5) on the two separate data sets. For the data set having load factors less than the turning point, the coefficient of $\overline{\text{Load Factor}}$ is 3.25 (p=0.000), indicating a positive relationship between load factor and carrier's financial performance. For the data set having load factors greater than the turning point, the coefficient of $\overline{\text{Load Factor}}$ is −1.62 (p=0.031), indicating a negative relationship between load factor and carrier's financial performance. The results from these two separate models confirm the inverted U-shaped relationship. Second, as observed in Figure 5, ATA was an outlier in intercept and slope. We accordingly delete ATA from our data set and re-estimate Model 2. The associated coefficients with $\overline{\text{Load Factor}}$ and $\overline{\text{Load Factor}}^2$ are 12.02 (p=0.010) and –6.70 (p=0.034) respectively, again confirming the inverted U-shaped curve. Third, to rule out an alternative model specification, we add a cubic term of $\overline{\text{Load Factor}}^3$ to Model 2 to test if the relationship should be S-shaped rather than an inverted U-shaped. The coefficients for $\overline{\text{Load Factor}}$ (74.14, p=0.321), $\overline{\text{Load Factor}}^2$ (–89.28, p=0.384) and $\overline{\text{Load Factor}}^3$ (36.02, p=0.438) all become non-significant. A likelihood ratio test also confirmed that the model with the cubic specification fits significantly worse than the model with quadratic specification (χ^2=0.59, p=0.443). Hence, the alternative cubic specification was also ruled out. In sum, we can conclude that there exists an inverted U-shaped relationship between load factor and carrier's financial performance.

5. THEORETICAL AND MANAGERIAL CONTRIBUTIONS

5.1 Theoretical and Methodological Contributions

Our research demonstrates both theoretical and methodological contributions to knowledge advancement and accumulation in airline research. We contribute to theory development in the following ways. First, to the best of our knowledge, this is the first study in airline research to adopt a between-within theorization

(Haans et al. 2016) to theorize an inverted U-shaped relationship. Our study sheds light on future operation research that utilizes longitudinal data to conduct similar thought experiments. Second, our two step between-within theorization aligns with the empirical data we collected. The ICC calculation discussed in section 4.2 reveals that OPOR clearly demonstrate between-carrier cross sectional differences and within-carrier longitudinal changes. Hence, our study corresponds to the recommendation of Haans et al. (2016) advocating that the thought experiment in theoretical reasoning should be in line with the data used in empirical analysis. Third, we build our theorization on existing operations theory (i.e., Theory of Performance Frontiers) and the concept of slack to re-investigate the relationship between load factor and financial performance. The strong support for our hypothesized relationship between load factor and carrier financial performance reinforces the importance of adopting existing theories to explain operations phenomenon. Accordingly, our research provides references for similar future research to extend the application of theories to other service sectors.

Methodologically, while related airline research has compiled panel data from DOT, the research methodologies adopted by past work are distinctively different, evolving from OLS to fixed effects and further to random effects. Fixed effects methods – although indisputably popular – have shortcomings when it comes to estimating panel data, especially when researchers are interested in between-group differences (Singer and Willett 2003; Hoffman 2015). Random effect – with its increasing exposure in recent research – needs to be *correctly* specified in order to capture both cross-sectional differences across groups and longitudinal changes within groups (Bell and Jones 2015).

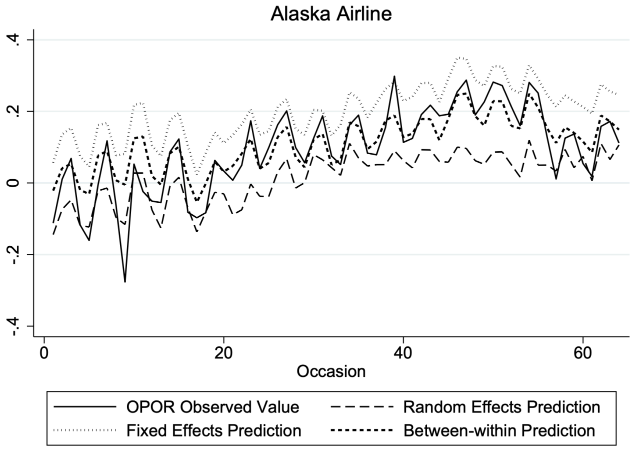
We illustrate our methodological contributions in the following perspectives. First, our between-within specification reveals more nuanced relationships than either fixed effects or traditional random effects models can provide. Take the variable of market share as an example: in fixed effects model, its coefficient is 0.84 (p=0.001) while in random effects model, its coefficient is 0.25 (p=0.253). A Hausman test concludes that fixed effects model is preferred (χ^2=205.86, p=0.000). So, we will use the result from fixed effects model to illustrate the nuances. Since fixed effects model can only measure within-carrier effects – the positive and statistically significant coefficient suggests that an individual carrier *increasing* its market share leads to *higher* profitability. In our between-within specification, the coefficient of the between-effect market share (β_{10}=–0.63, p=0.074) and the coefficient of the within-effect market share (β_{11}=0.76, p=0.004) tell a slightly different story: *increasing* an individual carrier's market share (β_{11} =0.76) still leads to *higher* profitability, but given the 16 years data we observed, consolidating market share in the industry ($\beta_{10=}$ –0.63) leads to *lower* profitability. Fixed effects model is not able to capture this subtleness in statistical causal inferences.

Second, the explanatory power of our between-within specification is higher than fixed effects and traditional random effects models. We use the exact same set of variables from Model 2 in Table 5 to run fixed effects and traditional random effects models and we see that both fixed effects and traditional random effects models validated the inverted U-shaped relationship albeit the turning points are slightly different: $\beta_{load\,factor}$=5.60 (p=0.000) and $\beta_{load\,factor2}$=–3.55 (p=0.000) for fixed effects; $\beta_{load\,factor}$=8.29 (p=0.000) and $\beta_{load\,factor2}$ =–4.95 (p=0.000) for random effects. Total R^2 for fixed effects (8.9%), random effects (41.47%), and between-within specification (64.53%) confirms that between-within specification has the highest explanatory power. In addition, we predict fitted OPOR values for the three models and we calculate both MAPE and MSE between fitted OPOR values and observed OPOR values for the entire 28 carriers (Table 7), we see that between-within specification has the smallest MAPE and MSE among the three models, indicating the best fit. We use Alaska as an example to illustrate this comparison (Figure 12). From Figure 12, we see that the fitted values from between-within specification are the closest to the actual values almost across all occasions. The comparisons among the Total R^2 and the fitted values also reinforce the fact that Hausman test should not be used to test between fixed effects and random effects (Bell and Jones 2015). In our analysis, Hausman test prefers the fixed effects model which actually has the worst fit among the three.

Table 7: MAPE and MSE of Fitted Values from the Three Models

Model	No. of Observations	Sum of APE	MAPE	Sum of SE	MSE
Fixed Effect	1039	8,294.87	798%	66.78	0.0643
Random Effect	1039	1,670.96	161%	11.66	0.0112
Between-Within	1039	1,500.20	144%	10.07	0.0097

Figure 12: Comparison between Fitted values of the Three Models

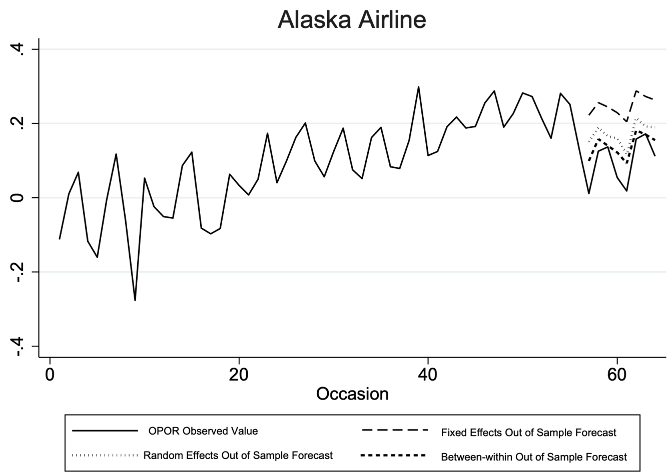


Third, our between-within specification has the best forecasting accuracy among the different model specifications. We split our data into training and test set. We keep the last 8 observations (last 8 quarters) of each carrier as the test set and the remaining as the training set (carriers with only 8 quarter observations are therefore dropped). We run fixed effects, random effects, and between-within specification models on the training data set. Then, we forecast OPOR 8 steps ahead using the three model outputs. We again calculate MAPE and MSE to compare forecasted OPOR values with observed OPOR values for these 8 quarters. We report the result of 12 carriers who span the entire 64 quarters for a reasonable comparison (Table 8), and we see that between-within specification yields the highest accuracy (smallest MAPE and MSE). Using Alaska Airline again as an example, we plot the forecasted values in Figure 13, where between-within specification is shown to have the best forecasting accuracy (closest to the observed values).

Table 8: MAPE and MSE of Forecasted Values from the Three Models

Model	No. of Observations	Sum of APE	MAPE	Sum of SE	MSE
Fixed Effect	96	245.61	256%	6.25	0.0651
Random Effect	96	76.31	79%	0.44	0.0046
Between-Within	96	59.31	62%	0.23	0.0024

Figure 13: Comparison between Forecasted values of the Three Models

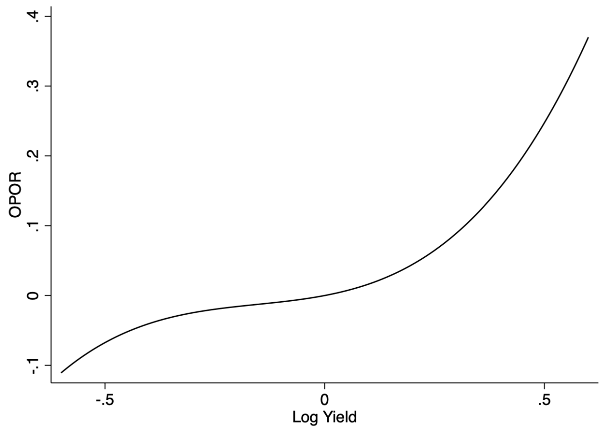


Fourth, to the best of our knowledge, our study is also the first study in airline research to adopt the three steps recommended by Lind and Mehlum (2010) and Haans et al. (2016) to test for the presence of an inverted U-shaped curve. Our results reveal that there exists an inverted U-shaped relationship between load factor and carrier financial performance. To this end, our three-step test procedure contributes to future airline research aiming to test curvilinear relationships by illustrating more stringent methodological approaches.

5.2 Managerial Insights

Responding to Garvett and Hilton's (2002) call that carriers who seek to maximize their financial performance should "obtain the best possible yields and load factors given the circumstances", we further investigate the relationship between yield and financial performance. Using Model 2 in Table 5, we found a significant S-shaped relationship between yield and financial performance: Yield (β=0.13, p=0.024), Yield^2 (β=0.37, p=0.000), Yield^3 (β=0.78, p=0.000). We plot this relationship in Figure 15 and see that the higher the yield, the better the financial performance. Recall that the profit-maximizing load factor is 88%, our recommendation to all carriers, therefore, is that to maximize financial performance, carriers should increase their yields as high as possible but should not increase their load factors beyond 88%. This recommendation is especially important coming out of the global pandemic where we foresee that carriers could be in desperate need to increase both their yields and load factors as a means to recover financially.

Figure 15: Relationship between Yield and OPOR



Carriers that are operating higher than 88% of load factors should expect detrimental rather than beneficial results. One potential explanation is that when carriers operate close to their asset frontier, they have reached the stage where "structural, technological changes" (Schmenner and Swink 1998, p. 111) are required to achieve better financial results. However, these structural and technological changes "usually require a large investment of resources in order to move the frontier or change its shape" (Schmenner and Swink 1998, p. 111). The airline industry, due to its asset-heavy and resource-rigid nature, may be unable to accommodate the required resources for structural and technological changes.

Despite the popular belief that airlines should increase their fleet utilization to achieve better financial performance (Gittell 2003; Belobaba 2009), we did not find a significant relationship between fleet utilization and financial performance. This finding is in line with Mantin and Wang (2012). Since Gittell's (2003) findings were drawn from a single airline (Southwest), we further refine our model by creating an interaction term between Southwest and the within fleet utilization, allowing Southwest to have its own slope. The refined model shows that neither the interaction term (β=0.07, p=0.911) nor the two fleet utilization terms (β_{18}=–0.02, p=0.754; β_{19}=0.11, p=0.413) are statistically significant, although with the expected signs. 19 years have passed since Gittell's (2003) research, possibly the airline industry has experienced some underlying changes such that the importance of fleet utilization has faded away or possibly that the authorities have restricted the ability of airlines to maneuver their fleet utilization. Therefore, to achieve better financial performance, our recommendation to the airline decision makers is that upon recovering from the pandemic, the focus should be on yields and load factors, rather than fleet utilization.

For policy makers, our results suggest that allowing more carriers to consolidate and increase their market size, possibly through merger and acquisition, leads to negative financial performance for the airline industry. This may be the result of market cannibalization following merger and acquisition. The recommendation here is that when policy makers approve airline merger and acquisition, industry wise financial performance may be another factor to assess in addition to service quality evaluations and anti-trust considerations.

6. LIMITATION AND FUTURE RESEARCH

Our research focused on load factor (i.e., revenue passenger miles divided by available seat miles), but some research suggests considering other factors beyond capacity measures. From an input-output analysis perspective, Schefczyk (1993) argued that load factor does not adequately capture and reflect an airline's overall operational performance because it ignores non-passenger inputs, outputs, and differences in factor costs, and it does not consider other inputs beyond capacity. Given these factors, we suggest these factors to be considered in future research.

Airline research has typically focused on one of three units of analysis: 1) firm level (Spiller 1983; Ramaswamy et al. 1994; Behn and Riley 1999; Shaffer et al. 2000; Mishina et al. 2004; Collins et al. 2011; Atkinson et al. 2013); 2) individual flight level (Bratu and Barnhart 2006; Ramdas and Williams 2008); and 3) airport-flight level (Rupp and Holmes 2006; Dana and Orlov 2014). Since the financial measures were only reported to DOT at the firm and quarter level, we accordingly aggregate all our data at a quarterly level to match the financial measures. However, additional insights might be drawn if using flight level data. For example, the insignificant findings between fleet utilization and financial performance may be better explained by flight level analysis if flight level financials are available. While DOT data does not allow us to explore potential nuances at the flight level, we encourage future researchers to use flight level data, if possible, to further delineate this relationship.

Our research confirms Schmenner and Swink's (1998) Theory of Performance Frontiers, supporting the assertion that when operating close to the asset frontier, a company will operate under the law of trade-offs while when operating away from the asset frontier, a company will operate under the law of cumulative capabilities. Airline operations fall under the classic categorization of service operations; hence, we believe that our findings can be generalized to other service operations to explore similar relationships, offering another fruitful future research path.

REFERENCES

Abrigo, M.R. and Love, I. 2016. Estimation of panel vector autoregression in Stata. The Stata Journal, 16(3), 778-804.

Aghion P, Bloom N, Blundell R, Griffith R, Howitt P. 2005. Competition and innovation: an inverted-U relationship. Quarterly Journal of Economics, 120(2), 701–728.

Aiken, L.S. and West S.G. 1991. Multiple regression: Testing and interpreting interactions. Sage.

Alan, Y. and Lapré, M.A., 2018. Investigating operational predictors of future financial distress in the US airline industry. Production and Operations Management, 27(4), 734-755.

Alexander, K. 2006. Southwest Pushes Fares Higher, Blaming Fuel Costs. Retrieved from http://www.washingtonpost.com. (Accessed on January 10, 2024).

Atkinson, S., Ramdas, K. and Williams, J.W. 2013. The costs of inefficient robust scheduling practices in the US airline industry. Retrieved from http://faculty.london.edu/kramdas. (Accessed on January 10, 2024).

Azadegan, A., Patel, P.C. and Parida, V. 2013. Operational slack and venture survival. *Production and Operations Management*, 22(1), 1-18.

Baltagi, B.H., Griffin, J.M. and Vadali, S.R. 1998. Excess capacity: a permanent characteristic of US airlines?. *Journal of Applied Econometrics*, 13(6), 645-657.

Behn, B.K. and Riley Jr, R.A. 1999. Using nonfinancial information to predict financial performance: The case of the US airline industry. *Journal of Accounting, Auditing and Finance*, 14(1), 29-56.

Bell, A. and Jones, K. 2015. Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods*, 3(1), 133-153.

Bell, A., Fairbrother, M. and Jones, K. 2018. Fixed and random effects models: making an informed choice. *Quality and Quantity*, 1-24.

Belobaba, P. 2005, November. Impacts of 9/11 on us airline performance. MIT Research and Development Conference.

Belobaba, P. 2009. Overview of airline economics, markets and demand. *The Global Airline Industry*, 47-72.

Berlin, Jesse A., Stephen E. Kimmel, Thomas R. Ten Have, and Mary D. Sammel. 1999. An Empirical Comparison of Several Clustered Data Approaches under Confounding due to Cluster Effects in the Analysis of Complications of Coronary Angioplasty. *Biometrics*, 55(2), 470–476.

Bliese, P.D. and Ployhart, R.E. 2002. Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, 5(4), 362-387.

Bourgeois III, L.J. 1981. On the measurement of organizational slack. *Academy of Management Review*, 6(1), 29-39.

Bourland, K.E. and Yano, C.A. 1994. The strategic use of capacity slack in the economic lot scheduling problem with random demand. *Management Science*, 40(12), 1690-1704.

Boyd, A. 1998. Airline Alliance Revenue Management: Global alliances within the airline industry add complexity to the yield management problem. *Or Ms Today*, 25, 28-31.

Bratu, S. and Barnhart, C. 2006. Flight operations recovery: New approaches considering passenger recovery. *Journal of Scheduling*, 9(3), 279-298.

Bureau of Economic Analysis 2024. Gross Domestic Product [GDP], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GDP. (Accessed on April 10, 2024).

Choi, I. 2001. Unit root tests for panel data. *Journal of International Money and Finance*, 20, 249-272.

Cole, F. 2023. Despite Recent Challenges, Major U.S. Airlines Report Post-Pandemic Profit. https://businesstravelerusa.com/news/despite-recent-challenges-major-u-s-airlines-report-post-pandemic-profit/. (Accessed April 10, 2024)

Collins, D.L., Román, F.J. and Chan, H.C. 2011. An empirical investigation of the relationship between profitability persistence and firms' choice of business model: Evidence from the US airline industry. *Journal of Management Accounting Research*, 23(1), 37-70.

Cook, A., Tanner, G. and Lawes, A. 2012. The hidden cost of airline unpunctuality. *Journal of Transport Economics and Policy*, 46(2), 157-173.

Dana Jr, J.D. and Orlov, E. 2014. Internet penetration and capacity utilization in the US airline industry. *American Economic Journal: Microeconomics*, 6(4), 106-37.

Daugherty, P.J., Chen, H., Mattioda, D.D. and Grawe, S.J. 2009. Marketing/logistics relationships: influence on capabilities and performance. *Journal of Business Logistics*, 30(1), 1-18.

Deephouse DL. 1999. To be different, or to be the same? It's a question (and theory) of strategic balance. *Strategic Management Journal,* 20(2), 147–166.

Department of Transportation (DOT). 2024. https://www.transtats.bts.gov. (Accessed on January 10, 2024).

Dickey, D.A. and Fuller, W.A. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427-431.

Dresner, M. and Xu, K. 1995. Customer service, customer satisfaction, and corporate performance. *Journal of Business logistics*, 16(1), 23-40.

Federal Aviation Administration, January 2022. The Economic Impact Report of Civil Aviation on the U.S. Economy. Retrieved from https://www.faa.gov/. (Accessed on April 10, 2024).

Garvett, D. and Hilton, K. 2002. What drives airline profits? A first look. *Handbook of Airline* Economics,169-181.

Gittell, J. H. 2003. The Southwest airlines Way: Using the power of relationships of high performance. New York: McGraw-Hill.

Haans, R.F., Pieters, C. and He, Z.L. 2016. Thinking about U: Theorizing and testing U-and inverted U-shaped relationships in strategy research. Strategic Management Journal, 37(7), 1177-1195.

Hamilton, J. D. 1994. Time Series Analysis. Princeton, NJ: Princeton University Press.

Hansen, M.M., Gillen, D. and Djafarian-Tehrani, R. 2001. Aviation infrastructure performance and airline cost: a statistical cost estimation approach. *Transportation Research Part E: Logistics and Transportation Review*, 37(1), 1-23.

He, W. 2019. Integrating overbooking with capacity planning: Static model and application to airlines. Production and Operations Management, 28(8), 1972-1989.

Hempel, C.G., 1965. Aspects of scientific explanation (Vol. 3). New York: Free Press.

Hendricks, K.B., Singhal, V.R. and Zhang, R. 2009. The effect of operational slack, diversification, and vertical relatedness on the stock market reaction to supply chain disruptions. Journal of Operations Management, 27(3), 233-246.

Hoffman, L. 2015. Longitudinal analysis: Modeling within-person fluctuation and change. Routledge.

Hoffman, L. and Stawski, R.S. 2009. Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development*, 6(2-3), 97-120.

IATA, 2018. Fuel Fact Sheet. Retrieved from https://www.iata.org. (Accessed on April 10, 2024).

Kalemba, N. and Campa-Planas, F. 2017. An overview of the quality concept in the air transportation business: a systematic literature review. *International Journal for Quality Research*, 11(1), 51-70.

Kovach, J.J., Hora, M., Manikas, A. and Patel, P.C. 2015. Firm performance in dynamic environments: The role of operational slack and operational scope. *Journal of Operations Management*, 37, 1-12.

Lapré, M.A. and Scudder, G.D. 2004. Performance improvement paths in the US airline industry: Linking trade-offs to asset frontiers. *Production and Operations Management*, 13(2), 123-134.

Lieberman, M.B. 1987. The learning curve, diffusion, and competitive strategy. *Strategic Management Journal*, 8(5), 441-452.

Lind JT, Mehlum H. 2010. With or without U? The appropriate test for a U-shaped relationship. *Oxford Bulletin of Economics and Statistics*, 72(1), 109–118.

Lütkepohl, H. 2005. New Introduction to Multiple Time Series Analysis. Heidelberg: Springer.

Mantin, B. and Wang, J.H.E. 2012. Determinants of profitability and recovery from system-wide shocks: The case of the airline industry. *Journal of Airline and Airport Management*, 2(1), 1-21.

Mellat-Parast, M., Golmohammadi, D., McFadden, K.L. and Miller, J.W. 2015. Linking business strategy to service failures and financial performance: Empirical evidence from the US domestic airline industry. *Journal of Operations Management*, 38, 14-24.

Mishina, Y., Pollock, T.G. and Porac, J.F. 2004. Are more resources always better for growth? Resource stickiness in market and product expansion. *Strategic Management Journal*, 25(12), 1179-1197.

Morash, E.A. 2001. Supply chain strategies, capabilities, and performance. *Transportation journal*, 37-54.

Mundlak, Y. 1978. On the pooling of time series and cross section data. *Econometrica: Journal of the Econometric Society*, 69-85.

Nakagawa, S. and Schielzeth, H. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133-142.

Netessine, S. and Shumsky, R. 2002. Introduction to the theory and practice of yield management. *INFORMS Transactions on Education*, 3(1), 34-44.

Piger, J. M. and Chauvet, M., Smoothed U.S. Recession Probabilities [RECPROUSM156N], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/RECPROUSM156N (Accessed on April 10, 2024).

Ramaswamy, K., Thomas, A.S. and Litschert, R.J. 1994. Organizational performance in a regulated environment: the role of strategic orientation. *Strategic Management Journal*, 15(1), 63-74.

Ramdas, K. and Williams, J. 2008. An empirical investigation into the tradeoffs that impact on-time performance in the airline industry. *Washington Post*, 1-32.

Ramdas, K., Williams, J. and Lipson, M. 2013. Can financial markets inform operational improvement efforts? Evidence from the airline industry. *Manufacturing and Service Operations Management*, 15(3), 405-422.

Rothstein, M. 1985. OR Forum—OR and the airline overbooking problem. *Operations Research*, *33*(2), 237-248.

Rupp, N. and Holmes, M. 2006. An investigation into the determinants of flight cancellations. *Economica,* 73(292), 749-783.

Rupp, N., Owens, D. and Plumly, L. 2006. Does competition influence airline on-time performance? *Advances in Airline Economics*, 1, 251-272.

Saranga, H. and Nagpal, R. 2016. Drivers of operational efficiency and its impact on market performance in the Indian Airline industry. *Journal of Air Transport Management*, 53, 165-176.

Schefczyk, M. 1993. Operational performance of airlines: an extension of traditional measurement paradigms. *Strategic Management Journal*, 14(4), 301-317.

Schmenner, R.W. and Swink, M.L. 1998. On theory in operations management. *Journal of Operations Management*, 17(1), 97-113.

Schurer, S. and Yong, J. 2012. Personality, well-being and the marginal utility of income: What can we learn from random coefficient models? Research paper.

Scotti, D. and Dresner, M. 2015. The impact of baggage fees on passenger demand on US air routes. *Transport Policy*, 43, 4-10.

Shaffer, B., Quasney, T.J. and Grimm, C.M. 2000. Firm level performance implications of nonmarket actions. *Business and Society*, 39(2), 126-143.

Sim, K.L, Joon Song, C. J. and Killough, L.N. 2010. Service quality, service recovery, and financial performance: An analysis of the US airline industry. In *Advances in Management Accounting* (27-53). Emerald Group Publishing Limited.

Singer, J.D. and Willett, J.B. 2003. Applied longitudinal data analysis: Modeling change and event occurrence. Oxford University Press.

Spiller, P.T. 1983. The differential impact of airline regulation on individual firms and markets: An empirical analysis. *The Journal of Law and Economics*, 26(3), 655-689.

Steele, D.C. and Papke-Shields, K.E., 1993. Capacity slack: strategic alternative to lead time. *Production and Inventory Management Journal*, 34(4), 1-5.

Steven, A.B., Dong, Y. and Dresner, M. 2012. Linkages between customer service, customer satisfaction and performance in the airline industry: Investigation of non-linearities and moderating effects. *Transportation Research Part E: Logistics and Transportation Review*, 48(4), 743-754.

Suzuki, Y. 2000. The relationship between on-time performance and airline market share: a new approach. *Transportation Research Part E: Logistics and Transportation Review*, 36(2), 139-154.

Suzuki, Y. 2002. An empirical analysis of the optimal overbooking policies for US major airlines. *Transportation Research Part E: Logistics and Transportation Review*, *38*(2), 135-149.

Tan, J. and Peng, M.W. 2003. Organizational slack and firm performance during economic transitions: Two studies from an emerging economy. *Strategic Management Journal*, 24(13), 1249-1263.

Teece, D.J., Pisano, G. and Shuen, A. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7), 509-533.

Tsikriktsis, N. 2007. The effect of operational performance and focus on profitability: A longitudinal study of the US airline industry. *Manufacturing and Service Operations Management*, 9(4), 506-517.

Weick KE. 1989. Theory construction as disciplined imagination. *Academy of Management Review,* 14(4), 516 – 531.

Wooldridge, J.M. 2010. Econometric analysis of cross section and panel data. MIT press.

Wyckoff, D. D. and D. H. Maister. 1977. The Domestic Airline Industry, Lexington Books, Lexington, MA.

Yetkiner, H. and Beyzatlar, M.A. 2020. The Granger-causality between wealth and transportation: A panel data approach. *Transport Policy*, 97, 19-25.

Yimga, J. 2017. Airline on-time performance and its effects on consumer choice behavior. *Research in Transportation Economics*, 66, 12-25.

Yu, C. 2016. Airline productivity and efficiency: concept, measurement, and applications. In Airline Efficiency (11-53). Emerald Group Publishing Limited.

Zou, L. and Chen, X. 2017. The effect of code-sharing alliances on airline profitability. *Journal of Air Transport Management*, 58, 50-57.

Appendix A Comparison of Current Research with Selective Literature

Research	Load Factor as Main Predictor	Between Carrier Effect	Within Carrier Effect	Curvilinear Relationship	Level of Analysis
Atkinson et al. (2013)	No	No	Yes	No	Carrier
Behn and Riley (1999)	Yes	No	Yes	No	Carrier
Collins et al. (2011)	No	No	Yes	No	Carrier
Ramdas and Williams (2008)	No	No	Yes	No	Flight
Sim et al. (2010)	Yes	No	Yes	No	Carrier
Current Research	Yes	Yes	Yes	Yes	Carrier