Transportation Research Part E 51 (2013) 95–108

Contents lists available at SciVerse ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

Economies of scale in the US airline industry

Ahren Johnston [a,*], John Ozment [b]

[a] Missouri State University, United States
[b] Oren Harris Chair of Transportation, University of Arkansas, United States

ARTICLE INFO

Article history:
Received 24 October 2011
Received in revised form 1 June 2012
Accepted 26 August 2012

Keywords:
Airline
Scale economies
Economics
Transportation
Modeling

ABSTRACT

This study investigates economies of scale in the US airline industry using annual data, from 1987 to 2009, on the largest airlines. The paper estimates both a translog and Cobb–Douglas model of both an economic and transportation definition of economies of scale. The study shows that the results from models based on the two definitions are remarkably similar except during rapid growth in output and that the largest US airlines operate under modest scale economies.

1. Introduction

Prior to passing the Airline Deregulation Act of 1978, research presented before Congress provided an almost unanimous conclusion that the airline industry was characterized by constant returns to scale, and, therefore, not a natural monopoly (Antoniou, 1991; Kyle and Phillips, 1985; White, 1979). Consequently, law makers believed that elimination of federal economic regulation would permit the development of a truly competitive airline industry in which a large number of competitors would provide more service at lower prices (Harper, 1982).

The early experience under airline deregulation was very much as expected; the industry became less concentrated with a rush of new carriers providing more service at highly competitive rates. However, many of the new carriers exited the market during the late 1980s and early 1990s, and the industry became more heavily concentrated than ever (Borenstein, 1992; Brueckner and Spiller, 1994; Dempsey, 1993; Goetz and Sutton, 1997; Kahn, 1988; Rakowski and Bejou, 1992). Studies of returns to scale (RTS) during this period continued to confirm the results of previous studies; i.e., constant returns to scale.[1] These studies continued to use data from the time period immediately before and after deregulation when the industry was in a state of flux before the departure of newer carriers and increased concentration of the late 1980s and early 1990s.

Consolidation and concentration in the face of such overwhelming evidence from the literature disputing the existence of economies of scale has led some authors to question the methods used in prior studies. Measures of RTS and output were called into question, leading to problems associated with including into cost models operational characteristic such as average stage length, average load factor, and network size (Ying, 1992; Jara-Diaz and Cortes, 1996; Oum and Zhang, 1997). While corrections for these differences have been applied to the results of previous studies (Jara-Diaz and Cortes, 1996;

* Corresponding author.
E-mail address: Ahren.Johnston@MissouriState.edu (A. Johnston).

[1] See Baltagi et al. (1995), Bauer (1990), Caves et al. (1984), Roy and Cofsky (1985), Formby et al. (1990), Gillen et al. (1985, 1990), Keeler and Formby (1994), Kirby (1986) and Kumbhakar (1992).

1366-5545/$ - see front matter
http://dx.doi.org/10.1016/j.tre.2012.09.002

Oum and Zhang, 1997; Basso and Jara-Diaz, 2005, 2006b), they have not been incorporated into models analyzing current or even more recent data.

The most recent airline data analyzed was from 1983 to 1989 (Creel and Farell, 2001). Creel and Farell did not correct for the methodological shortcomings of earlier models in that they used the same variables but applied a Fourier series rather than a Taylor series expansion, and the data analyzed may have been affected by the chaotic conditions of an unstable industry. Other studies analyzed data from before deregulation and/or through 1985 when the industry was behaving as if there were indeed no economies of scale. While the US airline industry may still be affected by the irregularities of bankruptcies and mergers, it has grown significantly over the past two decades; is a much different and far more mature environment than that of the 1980s; and is behaving as though economies of scale may exist. Thus, there are clear benefits to examining a larger, more recent data set.

Furthering the case that a new study is warranted is the fact that variables such as output, density, scope, stage length, load factor, and network size are likely to be correlated, and when several of these are included as independent variables in the same model, results are likely to be affected by multicollinearity, which makes interpretation of significance levels of certain variables problematic (Farrar and Glauber, 1967; Neter et al., 2000). The effects of multicollinearity may also explain some of the lack of evidence that has been reported with respect to RTS. In fact, there is evidence to suggest that early studies may have been affected by multicollinearity (Eads et al., 1969; Harbeson, 1970; CAB, 1972a,b; Sarndal and Statton, 1975), and given the nature of the variables used in more recent studies, it probably had effects on them as well, but the topic has not been addressed in the more recent studies. Although there are few ways of dealing with the effects of multicollinearity, one of the most appropriate is addition of new data (Neter et al., 2000). Therefore, this study incorporates more data and a simpler model than previous studies leading to lower correlations between the variables.

Analyses of RTS in transportation have departed significantly from the definition of RTS in the traditional economic literature (Varian, 1996), and the industry has changed substantially from that which is represented in the data used in previous analyses. Consequently, this study examines data from 1987 through 2009. This study is designed to reconcile some of the disparity between traditional economic definitions of economies of scale and those that have evolved in the transportation literature, while assessing the potential effects of multicollinearity.

2. Literature review

There have been at least 30 studies in the last 60 years that have dealt with economies of scale in the airline industry, although economies of scale were not always the main focus of these studies. Most of the studies have examined US carriers, with a few exceptions that looked at either Canadian or Australian carriers. The earliest study was conducted by Crane (1944) and the most recent was performed by Creel and Farell (2001). All of this literature uses data prior to 1989 when concentration was decreasing immediately following deregulation (Johnston and Ozment, 2011). This study uses more recent data to fill the void and also discusses other proposed measures of returns (e.g. returns to density, scope, etc.).

2.1. Pre-deregulation studies

Most of the research prior to deregulation concluded that economies of scale do not exist in the airline industry[2] or are exhausted after an airline achieves a relatively small size.[3] One pre-deregulation study found economies of scale (Eads et al., 1969), and one found diseconomies of scale (Reid and Mohrfeld, 1973), so the overall consensus was that the airline industry exhibited constant returns to scale. Collectively, these pre-deregulation studies were used to promote the deregulation of the airline industry in 1978. In the years that followed deregulation, there were several more studies looking at the effects of deregulation, and approximately half of these studies found increasing RTS[4] while the others found constant returns to scale.[5] However, most of these more recent studies have used data obtained from the period prior to and immediately following deregulation, and the most recent data used in any study was from 1983 to 1989 (Creel and Farell, 2001). Creel and Farell's may also be the only study found that did not use any pre-deregulation data.[6] The lack of studies using recent data was a major impetus to this study which uses data from 1985 to 2009.

With the exception of Proctor and Duncan (1954), the studies through 1965 used non-statistical methods (Crane, 1944; Koontz, 1951), simple correlations (Cherington, 1958; Koontz, 1952), or graphical analysis (Caves, 1962; Wheatcroft, 1956) to see if the average cost per unit of output was smaller for the larger carriers, and all of these studies concluded that RTS were constant after the very small or medium sized carriers. There have been a few more recent studies that used these types of non-statistical approaches (Harbeson, 1970; Jordan, 1970) and also reached the same conclusions. Proctor and Duncan (1954) performed a multiple regression analysis, but reached the same conclusion as the other studies of the time period through 1965.

[2] See CAB (1972a,b) Douglas and Miller (1974), Harbeson (1970), Keeler (1972), Murphy (1969), Sarndal and Statton (1975), Sarndal et al. (1978), Strazheim (1969) and US Congress (1975).

[3] See Crane (1944), Caves (1962), Cherington (1958), Gordon (1965), Jordan (1970), Koontz (1951, 1952), Proctor and Duncan (1954) and Wheatcroft (1956).

[4] See Creel and Farell (2001), Kumbhakar (1990), Sickles (1985), Sickles et al. (1986) and Viton (1986).

[5] See Caves et al. (1984), Formby et al. (1990), Gillen et al. (1985), Kirby (1986) and Roy and Cofsky (1985).

[6] For further review of the literature on economies of scale in the airline industry, see Antoniou (1991).

Beginning with a study by Eads et al. (1969), many authors began estimating the total cost curve or average cost curve and using elasticity of total costs or average cost with respect to output to determine if economies of scale existed. Eads et al. concluded that slightly increasing RTS existed for local service carriers, but this was consistent with previous studies that found economies of scale only in very small carriers (1969). Most of the other studies from 1965 through 1984 used a similar technique to conclude that RTS in the US airline industry were constant or constant after the smallest carriers. An exception to this was a study by Reid and Mohrfeld (1973), which used a simple regression between revenue and assets to conclude that decreasing RTS existed. Reid and Mohrfeld's study is notable only in the fact that it is the only study to conclude decreasing RTS.

2.2. Post-deregulation studies

The studies mentioned in Section 2.1 tended to estimate a total cost function by specifying a specific structural form such as the Cobb Douglas cost function, but beginning with a study by Caves et al. (1984), it became common practice to estimate the total cost function using a second order Taylor Series expansion known as the transcendental logarithmic (translog) total cost function. Caves et al. defined returns to scale (RTS), as the inverse of the sum of elasticity of total cost with respect to output and the elasticity of total cost with respect to network size and returns to density (RTD), as the inverse of elastic of total cost with respect to output. Thus RTS in transportation came to be the impact of increasing network size and output simultaneously and RTD came to be the impact of increasing output and leaving network size constant. The differences between these and the traditional definition of RTS are discussed in detail in Section 3. Between 1985 and 2001, more studies were conducted, most of which used a similar technique of estimating a total cost curve and looking at elasticity of total cost with respect to output (Sickles et al., 1986; Kumbhakar, 1990; Formby et al., 1990). There were two exceptions to this technique. Viton (1986) looked at average cost divided by marginal cost using data from the US in 1979 to see if fares had fallen to marginal costs after de-regulation. The focus of Viton's paper was not on RTS but on efficiency, but it concluded that increasing RTS existed. The other exception was the most recent study by Creel and Farell (2001). Creel and Farell used a Fourier Series approximation and compared this to the Taylor Series approximation. The Fourier Series approximation resulted in a total cost curve with a slightly different shape, but the conclusions were the same for both of the models that were tested, that only slight economies of scale existed in the airline industry.

2.3. Studies critical of Caves et al.

Although must studies conducted since Caves et al. (1984) have used the definitions of RTD and RTS used in that 1984 study, there have been multiple papers criticizing these definitions.[7] In one of the first of these papers, Panzar (1989) showed that RTD were equal to what had previously been defined as multiproduct RTS and showed an example where the Caves et al. (1984) definition of RTS is always equal to unity showing that this is not a particularly relevant measure.[8]

In response to this criticism, two additional measures of return were investigated by Basso and Jara-Diaz (2005, 2006b). In the first of these papers, Basso and Jara-Diaz (2005) argued that the calculation of economies of spatial scope would be more useful than calculations of economies of scale and go onto calculate this measure based on the results of a previous study by Gillen et al. (1990). They conclude that economies of spatial scope do exist and that it would be beneficial for a company operating multiple subsidiaries to merge those operations. This, however, would only be the case in the presence of economies of scale for non-overlapping networks or increasing RTD for overlapping networks.[9] Therefore, the estimation of the traditional economic definition of RTS and the transportation definition of RTD should be sufficient to determine whether returns to spatial scope exist.

The second measure investigated by Basso and Jara-Diaz (2006b) is a measure of multiproduct economies of scale, *S*, they developed as an alternative to RTD to describe how expenses could be lowered with fixed network size by varying the route structure. As opposed to RTD which describes a fixed network size and route structure. Kraus (2008) later showed that these two measures are actually identical, if an airline is operating in a cost-minimizing manner. Therefore, *S*, as defined by Basso and Jara-Diaz is useful in determining whether an airline is operating in a cost-minimizing manner with a fixed network size, but still differs from the traditional economic definition of economies of scale which allows one to see the impact on costs of allowing all factors other than factor input prices to vary as output is increased.

2.4. Rational for this paper

As stated in Section 1, there have been a large number of studies looking at the issue of returns to scale in the airline industry, but none have been conducted using data from the last 20 years. In light of the large-scale mergers of the last 10 years and the increased industry concentration since deregulation it seems reasonable to theorize that economies of scale

[7] For a more detailed description of these papers critical of Caves et al., see Basso and Jara-Diaz (2006a).

[8] Similarly, Hurdle et al. (1989), Filipini and Maggi (1992), Jara-Diaz and Cortes (1996), and Oum and Zhang (1997) have shown that either Caves et al. (1984) definition of RTS or an improved version of it is the same as multiproduct RTS.

[9] In the absence of increasing RTS, the merger of two non-overlapping airlines would result in either the same or higher costs based on the definitions of constant and diminishing RTS.

exist today in the US airline industry even though they may not have existed prior to 1985 (Johnston and Ozment, 2011). This study examines economies of scale using an updated data set and seeks to reconcile some of the disparity between the economic and transportation definitions of economies of scale. The analysis in this study focuses on replicating the Caves et al. study using more recent data and compares those measures of RTS and RTD to a traditional economic measure of RTS using the same data. While the studies discussed in Section 2.3 provide valuable insights, investigating all of these potential measures in beyond the scope of the current study.

3. Returns to scale

From the traditional economics perspective, returns to scale (RTS) are a characteristic of a particular production technology or production function. They describe the impact on output of scaling all inputs up or down in constant proportions. Increasing RTS refers to a more than proportional change in output for a given change in inputs, diminishing RTS refers to a less than proportional change in output for a given change in inputs, and constant RTS refers to a proportional change in output for a given change in inputs. In most instances, one would expect to see constant RTS, decreasing RTS are typically a short run phenomenon caused by one input being held fixed, and increasing RTS are sometimes possible because an increase in output does not require an increase in fixed costs. It is also possible for a technology to have RTS vary with levels of output and other factors (Varian, 1996).

Economies/diseconomies of scale, is a related concept but is a characteristic of a particular cost function. This concept describes the impact on total costs of scaling output up or down. Economies of scale refers to a less than proportional change in total cost for a given change in output, and diseconomies of scale refers to a more than proportional change in total cost for a given change in output. If costs go up proportionally with a change in output, neither economies nor diseconomies of scale exist.

"Koontz (1951) was the first to introduce a distinction between economies of size and of density" (Antoniou, 1991), but "after Caves et al. (1984), it became customary to analyze transport industry structure using two indices: (1) economies of density (RTD) and (2) economies of scale with variable network size (RTS)" (Basso and Jara-Diaz, 2006a,b). RTD was defined as the impact of increasing output while holding network size, average length of haul, and load factor constant, and the definition of economies of scale in the transportation industry was adjusted to describe the impact on cost of changing both output and network size simultaneously and by the same percentage. Neither of these definitions coincides with the economic definition of economies of scale due to the inclusion of operating characteristics in the cost function. The traditional economic definition of economies of scale describes the impact on cost of increasing output while keeping input prices constant, and with management making decisions that result in the maximum profit. This entails adjusting or holding constant things such as load factor, network size, and average length of haul as appropriate. By including these network factors as control variables in a cost function, it is only possible to see the impact of increasing or decreasing output while holding all other independent variables constant or adjusting some simultaneously and by the same percentage and holding the others constant. Again, the traditional economic definition of RTS involves the impact on cost of increasing output, while holding the cost of inputs constant, and allowing all other factors to vary in the most efficient manner in order for the individual firm to maximize profits.

This study addresses the disparity between these definitions by estimating both the traditional economic and transportation definitions of RTS. An added benefit of estimating a model without the various control variables will be the elimination of almost all of the multicollinearity and corresponding increases in the variation of coefficient estimates. The differences in the Variance Inflation Factors (VIF) between the two are discussed further in Section 6.

4. Econometric model

RTS can be measured by elasticity of scale (E_S) or the percentage change in output given a one percent change in all inputs, which can be derived from a production function. However, due to the duality theory, it can also be derived from the cost function as the reciprocal of the elasticity of costs with respect to output, which can be derived from a cost function. An elasticity of costs less than one corresponds to economies of scale and increasing RTS (elasticity of scale greater than one), and an elasticity of costs of greater than one corresponds to diseconomies of scale and decreasing RTS (elasticity of scale less than one). Finally an elasticity of costs (and scale) equal to one corresponds to neither economies nor diseconomies of scale and constant RTS.

Since first introduced by Caves et al. in 1984, the majority of studies on economies of scale in the airline industry and transportation in general have attempted to derive a total cost function of the following form:

$$TC = f(W, Z, Y, T, F) \tag{1}$$

where TC is total operating cost, W is a vector of factor input prices, Z is a vector of control variables or operating characteristics, Y is a measure of output, T is a vector of time effects, and F is a vector of firm effects. This study uses the same basic model to estimate a transcendental logarithmic (translog) total cost function. The primary advantage of the translog function is that it is an approximation of any general function because it is based on a second order Taylor series expansion (Caves et al., 1984). Another advantage of the translog functional form over a Cobb–Douglas cost function, which is often used for cost and production functions, is that it allows for variances in economies of scale as the various factors change over time. This more closely fits economic theory than constant RTS, which exist in Cobb–Douglas production functions. This implies that a particular airline could exhibit increasing RTS at some times and constant RTS at others as the values of the input prices, control variables, and output level vary.

The problem with this functional form in Eq. (1) is that the economic definition of economies of scale does not include the effects of managerial decisions that would impact operational characteristic such as average stage length, average load factor, and network size. For example, if a factory doubles its output and costs increase by less than twice, that factory exhibits economies of scale whether it increased output by investing in better equipment, improving efficiency, doubling the size of the existing factory, or building a second, identical factory. These factors are simply assumed to vary in the most efficient way possible. However, including operational characteristics in the cost function means that a researcher can only use this model to see what will happen to costs as some (or all) of these operational characteristics are held fixed while the others vary proportionally with output. Given this limitation, the decision by Caves et al. (1984) to define RTS as the impact on costs while holding average stage length and load factor constant and allowing network size to vary proportionally with changes in output is logical. However a researcher using this model has no insight as to what happens if output increases, network size changes by a non-proportional amount, and average stage length and/or load factor change (a likely scenario). Therefore, in this study a second model was estimated to look at E_S with the following functional form:

$$TC = f(W, Y, T, F) \tag{2}$$

The same factor input prices (W) have been used in nearly every previous study. These have been fuel, labor, and capital (or materials and capital), which are arguably the most important factors affecting total costs for airlines. Therefore, this study uses fuel, labor, and capital prices as the factor input prices.

The various control variables (Z) used in previous studies have included average load factor (ALF), average stage length (ASL), number of points served, number of routes served, and number of city pairs served (CPS). For this study, ALF, ASL, and CPS will be used as control variables in Model 1. Number of points served was considered as an alternative to CPS but the estimated coefficient was not significant and the sum of squared errors was higher than for the model using CPS as the third control variable. CPS is a measure of direct connections that exist within an airline's network rather than connections through a hub because each leg of a journey would have similar expenses related to terminal operations. That is to say the cost of terminal operations related to a flight with one connection would be approximately double the cost of terminal operations related to a direct flight. These particular variables were chosen because they are some of the most commonly used in previous studies, and they do not have excessive correlation with other variables such as output or control variables. However, significant correlation does become an issue in the translog model with the addition of second order terms. These three control variables also capture how effectively the flight equipment is utilized and the size of the network. It is not necessary to include traffic density as a variable, which could be measured as ASM per CPS because this information is already included in the model by using ASM as a measure of output and CPS as a measure of network size.

There have been a variety of measures of output (Y) used in previous studies, and the most prevalent have been revenue seat miles (RSM), available seat miles (ASM), revenue ton-miles, available ton-miles, number of passengers and departures. However, all of these measures are highly correlated and for that reason, should not be used as multiple independent variables in the same model. Also, because of this high correlation, it seems unnecessary to develop an aggregate measure of output when just one could be used. Jara-Diaz and Basso (2003) suggest the use of both a flow-distance term and a pure flow term, such as number of passengers, in order to capture both flying and terminal expenses; however, with the data set used in this analysis, these two measures are highly correlated ($\rho = 0.89$), so either measure could be used as a variable to capture the same information. ASM is chosen as a measure of output because a cost function should depend upon the total amount of product produced (measured by ASM) not just the portion that is sold (measured by RSM) or the number of customers (measured by number of passengers). There is an assumption that a firm will operate efficiently, so that any large discrepancies between ASM and RSM should be a short lived phenomenon. Because of the nature of the industry and the fact that services cannot be stored for future use, small discrepancies will exist between the two in an attempt to capture all potential business.

Firm specific dummy variables (F) are included to capture any unobserved differences between firms. In addition to this, year specific dummy variables (T) are included to capture any yearly differences resulting from anything other than inflation. Finally, all dollar figures were converted to 1987 dollars using the producer price index for the airline industry (PPI) to account for differences resulting from inflation.

Including all of these variables results in Model 1 (Eq. (3)) which was proposed (with the restrictions) by Caves et al. (1984) and can be used to estimate the transportation definitions of RTS and RTD. The traditional economic definition can be estimated by Model 2 (Eq. (4)) if the restrictions and factor share equations are adjusted in the same manner.

$$\begin{aligned} \ddot{TC} = {} & \alpha_0 + \sum_T \alpha_T + \sum_F \alpha_F + \alpha_Y \ddot{Y} + \sum_i \beta_i \ddot{W}_i + \sum_i \emptyset_i \ddot{Z}_i + 1/2\delta_{YY}(\ddot{Y})^2 + 1/2\sum_i\sum_j \gamma_{ij}\ddot{W}_i\ddot{W}_j + 1/2\sum_i\sum_j \psi_{ij}\ddot{Z}_i\ddot{Z}_j \\ & + \sum_i \rho_{Yi}\ddot{Y}\ddot{W}_i + \sum_i \mu_{Yi}\ddot{Y}\ddot{Z}_i + \sum_i\sum_j \lambda_{ij}\ddot{W}_i\ddot{Z}_j \end{aligned} \tag{3}$$

$$\ddot{TC} = \alpha_0 + \sum_T \alpha_T + \sum_F \alpha_F + \alpha_Y \ddot{Y} + \sum_i \beta_i \ddot{W}_i + 1/2\delta_{YY}(\ddot{Y})^2 + 1/2\sum_i\sum_j \gamma_{ij}\ddot{W}_i\ddot{W}_j + \sum_i \rho_{Yi}\ddot{Y}\ddot{W}_i \tag{4}$$

where $\ddot{TC}_i = \left(\ln TC_i - \frac{\sum_{i=1}^{n} \ln TC_i}{n}\right), \ddot{Y}_i = \left(\ln Y_i - \frac{\sum_{i=1}^{n} \ln Y_i}{n}\right)$, etc. and Y = ASM, W = {fuel price, labor price, capital price}, Z = {ALF, ASL, CPS}. The mean of each variable was subtracted from each individual observation in order to center the data and make the first order coefficients directly interpretable as elasticities of cost. To ensure that the estimated cost function is homogenous of degree one in input prices and to ensure symmetric cross effects, the following standard restrictions were imposed on the parameters:

$$\begin{aligned} &\sum_i \beta_i = 1, \quad \sum_i \gamma_{ij} = 0 \ (\forall j), \quad \sum_i \rho_{Yi} = 0, \quad \sum_i \lambda_{ij} = 0 \ (\forall j). \\ &\gamma_{ij} = \gamma_{ji} \quad \psi_{ij} = \psi_{ji} \quad \lambda_{ij} = \lambda_{ji} \end{aligned} \tag{5}$$

Factor share equations were computed using Shephard's (1953) lemma. Specifically, each input share (C_i) can be equated to the derivative of the cost function with respect to input price i:

$$C_i = \beta_i + \sum_j \gamma_{ij} \ddot{W}_j + \rho_{Yi} \ddot{Y} + \sum_j \lambda_{ij} \ddot{Z}_j \tag{6}$$

Taking the derivative of the total cost function with respect to output results in an estimate of elasticity of total cost with respect to output (E_Y) or elasticity of costs. In a Cobb–Douglas function, this is simply the estimated coefficient on the measure of output, but it takes the following form for the translog function:

$$E_Y = \frac{\partial \ddot{TC}}{\partial \ddot{Y}} = \alpha + \delta_{YY} \ddot{Y} + \sum_i \rho_{Yi} \ddot{W}_i + \sum_i \mu_{Yi} \ddot{Z}_i \tag{7}$$

Similarly elasticity of total cost with respect to city pairs served (E_{CPS}) can be calculated at any point in the data. This means that both E_Y and E_{CPS} vary with changes in output level, factor input prices, and control factors. For Model 1, it is possible to calculate RTS and RTD as defined by Caves et al. (1984), and for Model 2, it is possible to calculate elasticity of scale (E_S) as the inverse of E_Y. For the sake of clarity, E_S will hereafter refer to the impact on cost of increasing output without regard to operational characteristics, RTS will refer to the impact on cost of increasing output and network size simultaneously while holding ALF and ASL constant, and RTD will refer to the impact on cost of increasing output while holding network size, ALF, and ASL constant.

$$E_S = 1/E_Y \quad \mathrm{RTS} = 1/(E_Y - E_{CPS}) \quad \mathrm{RTD} = 1/E_Y \tag{8}$$

Elasticity of scale, RTS, or RTD are then said to be increasing, constant, or decreasing when E_S is greater than one, equal to one, or less than one, respectively.

Model 1 is the transportation model popularized by Caves et al. (1984) and includes control variables. Model 1 is therefore used to estimate RTD and the transportation definition of RTS. That is: the impact on cost of increasing output while holding network size, average stage length, average load factor, and input prices constant (RTD) and the impact of increasing output and network size simultaneously and at the same rate while holding average stage length, average load factor, and input prices constant (RTS). Model 2 is the traditional economic model and can be used to estimate E_S or the impact on cost of increasing output while holding input prices constant and allowing all other factors to vary in the most efficient manner. A possible criticism of Model 1 is that output and the three control variables may be related to each other, which would result in endogenous variables appearing on the right hand side of the equations. This is a problem that is eliminated with Model 2 because output is determined from consumer demand and unobserved competitive forces, including the market determined price of output, not the price of inputs. Because this study seeks to reconcile the differences between a traditional definition of economies of scale and the transportation definition of economies of scale as first measured by Caves et al. (1984), it is important to use the same modeling technique rather than structural equation modeling or another technique to account for the potential endogenous right hand side variables. This same logic is also why the full translog functional form is estimated rather than eliminating some of the non-significant second order variables.[10]

5. Data

The data set used in this study is comprised of annual observations on what were the nine largest US airlines in 2006. The data consists of yearly observations on these airlines for every year between 1987 and 2009. These airlines present a wide range of output level and controlled 67.2% of the market in 1987 and 78.4% of the market in 2009 (ATA, 1976–2010). Data on America West Airlines was only available through 2006 because US Airways began reporting consolidated numbers in 2007 following their 2005 merger, and data for Northwest Airlines was only available through 2008 because consolidated data for Delta/Northwest was reported in 2009 following their 2008 merger. This resulted in a total of 203 observations, so the final models as estimated had 121 and 148 degrees of freedom. For a list of specific airlines and their beginning, ending, and average variables used see Table 1. Table 1 also lists the standard deviation of each variable over time for each airline and the

[10] Despite this reasoning, models eliminating some or all of the non-significant second order variables were tested, and the measures of RTS, RTD, and E_S were virtually identical to the full model.

Table 1
Description of the industry (dollar amounts adjusted for inflation).

Carrier	Time period	Expenses (millions)	ASM (millions)	Fuel per gallon	Wage per employee (000)	Cost of capital	Average load factor	Average stage length	Number of city pairs
Alaska	Mean	$1203	14,255	$0.82	$52	5.30%	0.66	729	63
	Std. dev.	$436	5507	$0.38	$7	0.94%	0.08	140	8
	1987	$504	5317	$0.61	$48	7.15%	0.55	517	55
	2009	$1700	21,166	$1.12	$68	4.91%	0.79	1,015	77
American	Mean	$9154	106,685	$0.77	$60	5.90%	0.70	952	254
	Std. dev.	$1332	10,343	$0.37	$8	1.46%	0.07	131	51
	1987	$5886	77,724	$0.56	$57	7.15%	0.64	729	311
	2009	$8339	92,929	$1.25	$63	5.60%	0.83	1,049	187
America West	Mean	$1504	21,552	$0.62	$33	7.40%	0.68	755	88
	Std. dev.	$534	5759	$0.29	$6	1.85%	0.07	195	8
	1987	$612	10,318	$0.47	$26	7.18%	0.56	442	70
	2006	$2608	28,562	$1.41	$40	4.14%	0.80	1,014	85
Continental	Mean	$4626	50,442	$0.81	$52	7.18%	0.71	923	162
	Std. dev.	$1042	2979	$0.39	$7	2.09%	0.09	169	34
	1987	$3459	54,626	$0.59	$46	11.12%	0.61	711	251
	2009	$5416	49,393	$1.25	$50	5.46%	0.85	1,185	105
Delta	Mean	$8639	95,770	$0.80	$66	6.78%	0.69	732	312
	Std. dev.	$1493	12,813	$0.42	$11	1.15%	0.09	114	66
	1987	$5255	71,504	$0.58	$54	6.33%	0.55	604	339
	2009	$12,201	107,838	$1.35	$63	4.74%	0.85	907	188
Northwest	Mean	$4804	52,189	$0.80	$59	5.36%	0.70	710	207
	Std. dev.	$908	5915	$0.41	$10	1.22%	0.08	56	15
	1987	$3256	41,499	$0.60	$54	8.40%	0.62	593	219
	2008	$5621	45,004	$2.15	$46	4.15%	0.85	814	161
Southwest	Mean	$3265	51,747	$0.64	$52	4.91%	0.67	478	248
	Std. dev.	$1936	30,837	$0.31	$9	1.16%	0.05	103	131
	1987	$657	11,457	$0.45	$49	6.64%	0.59	368	80
	2009	$6504	98,004	$1.31	$69	4.80%	0.76	639	440
United	Mean	$8424	92,634	$0.80	$56	6.94%	0.72	929	226
	Std. dev.	$1339	11,217	$0.38	$7	1.84%	0.07	116	39
	1987	$6403	86,246	$0.59	$54	11.05%	0.65	759	285
	2009	$6926	69,290	$1.12	$56	6.69%	0.84	1,110	161
US Airways	Mean	$5544	48,834	$0.79	$42	5.80%	0.69	601	265
	Std. dev.	$1401	10,807	$0.37	$8	1.19%	0.08	127	98
	1987	$1807	20,014	$0.58	$37	4.81%	0.65	425	183
	2009	$5342	53,214	$1.08	$47	5.81%	0.83	858	167
All Carriers	Mean	$5298	59,939	$0.76	$53	6.16%	0.69	757	204
	Minimum	$504	5317	$0.38	$24	1.71%	0.53	368	53
	Maximum	$12,201	124,653	$2.15	$87	11.33%	0.85	1,185	449
	Std. dev.	$3078	33,306	$0.37	$12	1.68%	0.08	199	101

minimum, maximum, mean and standard deviation of each variable for the entire sample. This is to show that there is significant variation in all variables in the sample and within each airline over time. As previously stated, all prices were converted to 1987 dollars using the producer price index (PPI), which was obtained from the Bureau of Labor Statistics' website (BLS, 1987–2009).

All data used in the analysis (excluding PPI) was obtained directly from Transtats or calculated using data obtained from Transtats (1987–2009). Most of the data was for scheduled, domestic service only with the exception of input prices and factor shares. Due to data availability, input prices and factor shares are based on system wide values, which seem to be highly correlated with domestic values. One category of output is used: available seat miles (ASM). This is the only measure of output used because it is highly correlated with all other potential measures of output such as revenue seat miles, number of passengers, revenue ton-miles, or available ton-miles; furthermore, ASM is a measure of the total output produced not only the portion sold (as RSM would be), and the carriers included in the study are predominately passenger carriers.

Three categories of input are used: fuel, labor, and capital. Fuel price is calculated by dividing fuel expenses by gallons used, labor price is calculated by dividing payroll expenses by number of employees, and cost of capital is calculated as a weighted average of interest as a percent of debt and depreciation as a percent of capital. Specifically, the measure of cost of capital was calculated as

$$\left[\frac{\text{Assets}}{\text{Assets}+\text{Liabilities}} * \frac{\text{Depreciation Expense}}{\text{Assets}}\right] + \left[\frac{\text{Liabilities}}{\text{Assets}+\text{Liabilities}} + \frac{\text{Interest Expense}}{\text{Liabilities}}\right] \tag{9}$$

Three categories of control variables were employed in Model 1: average load factor (ALF), average stage length (ASL), and city pairs served (CPS). Average load factor is calculated by dividing available seat miles by revenue passenger miles, and average stage length is calculated by dividing total miles by the number of departures. The dependent variable was total operating expenses. Dummy variables were also included for unobserved firm effects and time effects. The parameter estimates of the firm effect variables show if there is any difference between a particular carrier and Alaska Airlines, and the time effect variables show if there is any difference between a particular year and 1987.

6. Estimation and results

The final translog models and a Cobb–Douglas version of each model (second order terms restricted to 0) were estimated using SHAZAM econometric software with the system command. Specifically, the translog cost function and two of the three factor share equations were estimated using seemingly unrelated regression or Zellner estimation. In a Monte Carlo study, Kmenta and Gilbert (1968) showed that a non-iterative version of Zellner estimation is more efficient than an iterative version in the absence of autocorrelation or heteroskedascticity. With the data used in this study, there was evidence of significant autocorrelation in the residuals (Durbin–Watson ranged from 0.69 to 0.88)). As is the case with most statistical packages, SHAZAM does not allow for correction for autocorrelation with the systems of equations commands, so the data was manually transformed by the method described in Whistler et al. (2004) and re-centered. For each model, the system of equations was then estimated using the system command. The Durbin–Watson for these revised models ranged from 1.93 to 2.02, indicating that autocorrelation had been eliminated. Because the autocorrelation was eliminated prior to estimation, the iterative procedure was not used in the final models.

The estimation results for the two translog models as well as those two models with second order terms restricted to zero are shown in Table 2. To determine which model was the best fit to the data, *F*-tests were conducted to test the validity of the various restricted versions of Model 1. With alpha values of 0.000, the hypotheses from both models that all second order coefficients are equal to 0 are rejected. Therefore, both translog models are better than both Cobb–Douglas models. With

Table 2
Results of cost function estimations.[a]

	Model 1 translog	Model 2 translog	Model 1 Cobb–Douglas	Model 2 Cobb–Douglas
System R^2	0.9461	0.9318	0.8562	0.8464
LLF	1431.69	1390.38	1115.80	1097.73
Mean E_S/RTS	1.274**	1.301**	1.297**	1.380**
Mean RTD	1.628**	N/A	2.02**	N/A
Mean E_{ASM}	0.605*	0.769*	0.494*	0.725*
Mean E_{Fuel}	0.303*	0.313*	0.334*	0.337*
Mean E_{Wage}	0.566*	0.558*	0.538*	0.536*
Mean $E_{Capital}$	0.130*	0.129*	0.128*	0.127*
Mean E_{ALF}	−0.262	N/A	−0.508*	N/A
Mean E_{ASL}	−0.155	N/A	−0.134	N/A
Mean E_{CPS}	0.171*	N/A	0.277*	N/A

Where E_{ASM} = total cost elasticity with respect to ASM, E_{Fuel} = total cost elasticity with respect to fuel price, etc.
* Significantly different from 0 at 0.05 level.
** Significantly different from 1 at 0.05 level.
[a] See Appendix for full results.

Table 3
Elasticity of scale at different times.

		1987	1990	1993	1996	1999	2002	2005	2008	Avg.
Alaska	E_S	1.38*	1.32*	1.32*	1.32*	1.31*	1.31*	1.32*	1.31*	1.32*
	RTS	0.95	1.15	1.16	1.23*	1.20*	1.21*	1.25*	1.25*	1.19
	RTD	1.62*	1.61*	1.61*	1.52*	1.56*	1.64*	1.49*	1.47*	1.58*
America West	E_S	1.35*	1.30*	1.31*	1.31*	1.31*	1.30*	1.31*	N/A	1.31*
	RTS	1.02	1.28*	1.21*	1.26*	1.24*	1.25*	1.27*	N/A	1.22*
	RTD	1.51*	1.54*	1.62*	1.64*	1.58*	1.62*	1.49*	N/A	1.58*
American	E_S	1.28*	1.30*	1.29*	1.30*	1.29*	1.28*	1.30*	1.30*	1.29*
	RTS	1.33*	1.32*	1.27*	1.32*	1.31*	1.38*	1.36*	1.31*	1.26*
	RTD	1.86*	1.71*	1.68*	1.61*	1.74*	1.74*	1.53*	1.50*	1.64*
Continental	E_S	1.29*	1.30*	1.30*	1.31*	1.30*	1.30*	1.31*	1.31*	1.30*
	RTS	1.26*	1.25*	1.25*	1.28*	1.29*	1.26*	1.31*	1.28*	1.27*
	RTD	1.81*	1.67*	1.68*	1.63*	1.66*	1.66*	1.49*	1.43*	1.61*
Delta	E_S	1.29*	1.30*	1.29*	1.29*	1.29*	1.29*	1.30*	1.31*	1.29*
	RTS	1.28*	1.27*	1.29*	1.37*	1.32*	1.30*	1.34*	1.32*	1.30*
	RTD	1.94*	1.75*	1.71*	1.64*	1.67*	1.63*	1.48*	1.46*	1.67*
Northwest	E_S	1.30*	1.30*	1.30*	1.30*	1.29*	1.30*	1.31*	1.32*	1.30*
	RTS	1.22*	1.28*	1.27*	1.30*	1.30*	1.28*	1.31*	1.29*	1.27*
	RTD	1.77*	1.65*	1.66*	1.64*	1.63*	1.71*	1.58*	1.45*	1.64*
Southwest	E_S	1.35*	1.31*	1.30*	1.30*	1.30*	1.29*	1.30*	1.31*	1.30*
	RTS	1.03	1.20*	1.26*	1.30*	1.32*	1.28*	1.34*	1.34*	1.33*
	RTD	1.54*	1.55*	1.59*	1.65*	1.74*	1.72*	1.69*	1.71*	1.66*
United	E_S	1.28*	1.30*	1.29*	1.29*	1.29*	1.30*	1.30*	1.30*	1.29*
	RTS	1.35*	1.29*	1.30*	1.32*	1.32*	1.29*	1.33*	1.30*	1.31*
	RTD	1.77*	1.64*	1.62*	1.59*	1.67*	1.71*	1.53*	1.48*	1.63*
US Air	E_S	1.33*	1.29*	1.30*	1.30*	1.30*	1.31*	1.31*	1.31*	1.30*
	RTS	1.14	1.33*	1.27*	1.29*	1.27*	1.22*	1.29*	1.30*	1.31*
	RTD	1.71*	1.67*	1.76*	1.66*	1.75*	1.55*	1.40*	1.50*	1.65*
Average[a]	E_S	1.31*	1.30*	1.30*	1.30*	1.30*	1.30*	1.30*	1.31*	1.30*
	RTS	1.16	1.26*	1.25*	1.30*	1.28*	1.27*	1.31*	1.30*	1.27*
	RTD	1.71*	1.64*	1.66*	1.62*	1.66*	1.66*	1.53*	1.50*	1.63*

* Significantly different than 1 at the 0.05 level; all others are significantly different than 1 at the 0.05 level.
[a] Using the average of all carrier observations for a year to compute.

an alpha value of 0.033, the hypotheses that coefficients related to control variables are all equal to 0 was also rejected at the 0.05 level, however, there was also some evidence of reduced multicollinearity in Model 2 in that Model 1 had a VIF higher than 10 for 80% of the variables, and Model 2 had a VIF higher than 10 for 55% of the variables. Furthermore, all of the high VIFs in Model 2 were on the coefficients of fuel, fuel^2, and the time specific dummy variables. These were all highly significant despite the multicollinearity, indicating that the presence of multicollinearity in Model 2 had little impact on our estimates. Model 1, on the other hand, had high VIFs on many of the parameters of interest, and many of those were non-significant, suggesting that the presence of multicollinearity may have impacted those estimates.

Therefore, the *F*-test favors Model 1, and the VIFs favor Model 2, so the major deciding factor on whether to use Model 1 or Model 2 should be what type of information is sought. Model 1 shows the impact on costs of increasing output while holding load factor, stage length, and/or network size constant or allowing any or all of them to vary proportionally with output, while Model 2 shows the impact on costs of increasing output regardless of changes to load factor, stage length, and network size.

The final translog models had R^2 values of greater than 0.93 for the system and greater than 0.90 for all equations in the systems, indicating an excellent fit to the data. Comparisons of the average total cost elasticity with respect to output and other variables, scale elasticity, returns to scale and returns to density can be seen in Table 2, and a complete list of the estimated coefficients, including the cost share equations, can be seen in the Appendix in Table A1. As evidenced by Table 2, all the models are very similar and show that economies of scale or returns to scale exist in the airline industry. At the mean of the data, all measures of returns to scale are significantly different than one.[11]

The fact that the mean E_S/RTS from both models were very similar and significantly greater than one reveals that the major US airlines enjoy economies of scale under both definitions. Under the definition proposed by Caves et al. (1984) a 1%

[11] A variety of specifications were estimated for each model including without non-significant second order variables, without time specific dummy variables, without second order variables related to output, and without any second order variables (Cobb–Douglas). With all of these specifications, returns to scale existed under both Model 1 and Model 2 definitions and ranged from 1.19 to 1.30 for Model 1 and from 1.20 to 1.38 for Model 2. The consistency of the returns to scale measures is indicative of the robustness of the results, but the full translog specification was selected because it matches the method used by Caves et al. (1984) and will therefore provide a good basis of comparison between the two models.

increase in both ASM and CPS while holding ASL, ALF, and input factor prices constant should result in only a 0.78% increase in expenses, and under the traditional economic definition, a 1% increase in ASM while holding input factor prices constant and allowing CPS, ASL, and ALF to fluctuate should result in only a 0.77% increase in expenses. With regards to RTD in Model 1, a 1% increase in output with no change in network size or configuration would result in a more dense network and a 0.61% increase in expenses.

Elasticities of factor input prices are also very similar between the two models and, as would be expected, reveal that an increase in any of the factor prices will result in increased costs, and a 1% increase in all three factor input prices would result in a 1% increase in total expenses and expenses per ASM and RSM.

Elasticities of control or managerial variables in Model 1 reveal a lack of statistical significance of ALF and ASL, although both are negative values. These results are unsurprising because for ALF to increase while holding ASM constant would require an increase in RSM, and for the data included in this analysis, RSM and ASM were highly correlated. In other words, this is a scenario that may not have occurred in the data analyzed. The lack of significance of ASL seems surprising at first, but it would be unlikely to have an increase in ASL with no change in CPS, ASM, or ALF, so again this specific scenario may have never occurred in the data analyzed, making the lack of significance less surprising. The elasticity of CPS reveals that an increase in CPS while holding all else constant would result in slightly increased costs. This increase in CPS with no change in ASM would require a reallocation of resources and modest increases in ground operations from things such as terminal personnel and lease of terminal space.

Another important aspect of the airline industry that can be explored because this study estimated a translog model are measures of E_S, RTS, and RTD at different points in time and different airlines based on their input factor prices, output, and network configuration. These specific estimates can be calculated by evaluating the first derivative of the cost function with respect to output at various points in the data. Therefore these estimates are impacted by the coefficient of ASM and CPS and the second order coefficients or interaction terms related to ASM and CPS.

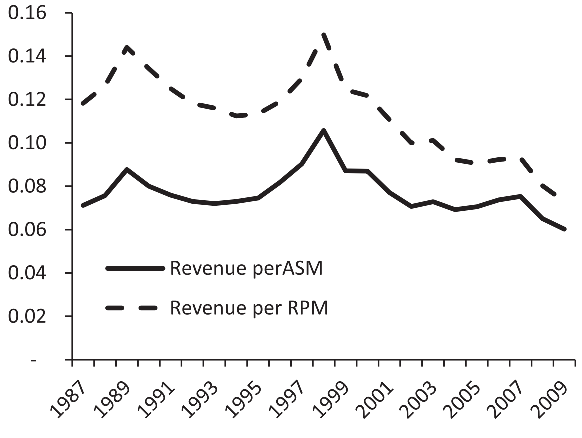

Fig. 1. Revenue per ASM and RPM.

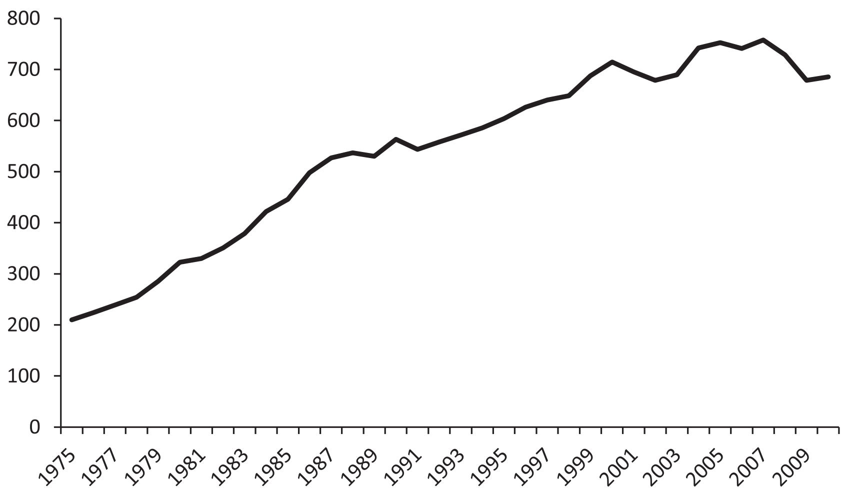

Fig. 2. US ASM 1975–2009.

Table A1
Full estimation results.

Cost equation	Translog model				Cobb–Douglas model	
	Model 1		Model 2		Model 1	Model 2
	$R^2 = 0.949$	$R^2 = 0.944$	$R^2 = 0.936$	$R^2 = 0.936$	$R^2 = 0.935$	$R^2 = 0.924$
First order terms						
Constant	−0.767*	−0.714*	−0.580*	−0.537*	−0.828*	−0.554*
Output	0.614*	0.554*	0.769*	0.740*	0.494*	0.725*
City pairs served	0.171*	0.240*	–	–	0.277*	–
Stage length	−0.155	−0.066	–	–	−0.134	–
Load factor	−0.262	−0.411*	–	–	−0.508*	–
Labor price	0.556*	0.556*	0.558*	0.558*	0.538*	0.536*
Fuel price	0.314*	0.315*	0.313*	0.314*	0.334*	0.337*
Capital price	0.129*	0.130*	0.129*	0.128*	0.128*	0.127*
Second order terms						
$(\text{Output})^2$	0.120	–	0.032	–	–	–
$(\text{City Pairs})^2$	0.187	–	–	–	–	–
Output ∗ City Pairs	−0.245	–	–	–	–	–
$(\text{Labor Price})^2$	0.198*	0.198*	0.203*	0.203*	–	–
$(\text{Fuel Price})^2$	0.218*	0.219*	0.224*	0.224*	–	–
$(\text{Capital Price})^2$	0.057*	0.056*	0.055*	0.054*	–	–
Labor ∗ Fuel	−0.179*	−0.181*	−0.186*	−0.186*	–	–
Labor ∗ Capital	−0.019*	−0.017*	−0.017*	−0.016*	–	–
Fuel ∗ Capital	−0.039*	−0.038*	−0.038*	−0.037*	–	–
$(\text{Stage Length})^2$	0.211	–	–	–	–	–
$(\text{Load Factor})^2$	1.289	–	–	–	–	–
Stage Length ∗ Load Factor	−0.901	–	–	–	–	–
Output ∗ Labor	−0.067*	−0.060*	0.000	–	–	–
Output ∗ Fuel	0.057*	0.060*	−0.009	–	–	–
Output ∗ Capital	0.011	–	0.008	–	–	–
City Pairs ∗ Labor	0.088*	0.081*	–	–	–	–
City Pairs ∗ Fuel	−0.077*	−0.081*	–	–	–	–
City Pairs ∗ Capital	−0.011	–	–	–	–	–
Stage Length ∗ Labor	0.116*	0.107*	–	–	–	–
Stage Length ∗ Fuel	−0.140*	−0.144*	–	–	–	–
Stage Length ∗ Capital	0.024	0.038*	–	–	–	–
Load Factor ∗ Labor	−0.171*	−0.162*	–	–	–	–
Load Factor ∗ Fuel	0.158*	0.162*	–	–	–	–
Load Factor ∗ Capital	0.013	–	–	–	–	–
Output ∗ Stage Length	−0.139	–	–	–	–	–
Output ∗ Load Factor	0.255	–	–	–	–	–
City Pairs ∗ Stage Length	0.113	–	–	–	–	–
City Pairs ∗ Load Factor	−0.548	–	–	–	–	–
Firm dummies						
America West	0.238*	0.213*	0.209*	0.177*	0.253*	0.189*
American	0.608*	0.584*	0.484*	0.477*	0.682*	0.510*
Continental	0.511*	0.479*	0.427*	0.389*	0.533*	0.396*
Delta	0.488*	0.471*	0.475*	0.463*	0.538*	0.488*
Northwest	0.447*	0.403*	0.445*	0.407*	0.435*	0.398*
Southwest	0.009	0.000	0.145**	0.120	−0.022	0.085
United	0.664*	0.632*	0.529*	0.516*	0.712*	0.525*
US Airways	0.631*	0.609*	0.719*	0.682*	0.644*	0.709*
Time dummies						
1988	0.086*	0.069*	0.066*	0.059*	0.087*	0.073*
1989	0.170*	0.151*	0.136*	0.127*	0.177*	0.157*
1990	0.152*	0.122*	0.102*	0.090*	0.138*	0.105*
1991	0.247*	0.210*	0.176*	0.163*	0.231*	0.175*
1992	0.314*	0.275*	0.232*	0.218*	0.302*	0.230*
1993	0.298*	0.268*	0.216*	0.202*	0.300*	0.214*
1994	0.329*	0.288*	0.227*	0.210*	0.336*	0.224*
1995	0.390*	0.346*	0.259*	0.239*	0.396*	0.244*
1996	0.414*	0.369*	0.266*	0.247*	0.406*	0.230*
1997	0.483*	0.435*	0.320*	0.300*	0.482*	0.291*
1998	0.542*	0.493*	0.366*	0.346*	0.583*	0.378*
1999	0.572*	0.523*	0.393*	0.374*	0.610*	0.403*
2000	0.495*	0.444*	0.303*	0.284*	0.501*	0.278*
2001	0.502*	0.456*	0.304*	0.284*	0.520*	0.280*
2002	0.461*	0.413*	0.258*	0.236*	0.491*	0.245*
2003	0.413*	0.365*	0.197*	0.173*	0.437*	0.173*

(*continued on next page*)

Table A1 (*continued*)

Cost equation	Translog model				Cobb–Douglas model	
	Model 1		Model 2		Model 1	Model 2
	R^2 = 0.949	R^2 = 0.944	R^2 = 0.936	R^2 = 0.936	R^2 = 0.935	R^2 = 0.924
2004	0.410*	0.366*	0.182*	0.159*	0.430*	0.147*
2005	0.353*	0.326*	0.128*	0.103*	0.433*	0.118*
2006	0.311*	0.289*	0.082	0.057	0.431*	0.105**
2007	0.295*	0.271*	0.052	0.027	0.416*	0.075
2008	0.198*	0.192*	−0.029	−0.057	0.448*	0.087
2009	0.340*	0.328*	0.098**	0.072	0.441*	0.082
Fuel share equation	R^2 = 0.923	R^2 = 0.923	R^2 = 0.900	R^2 = 0.899	R^2 = 0.136	R^2 = 0.132
Constant	0.314*	0.314*	0.314*	0.314*	0.334*	0.337*
Fuel Price	0.218*	0.219*	0.224*	0.224*	–	–
Labor Price	−0.179*	−0.181*	−0.186*	−0.186*	–	–
Capital Price	−0.039*	−0.038*	−0.038*	−0.037*	–	–
Output	0.057*	0.060*	−0.009	–	–	–
City Pairs Served	−0.077*	−0.081*	–	–	–	–
Stage Length	−0.140*	−0.144*	–	–	–	–
Load Factor	0.158*	0.162*	–	–	–	–
Labor share equation	R^2 = 0.930	R^2 = 0.930	R^2 = 0.910	R^2 = 0.910	R^2 = 0.494	R^2 = 0.493
Constant	0.556*	0.556*	0.558*	0.558*	0.538*	0.536*
Labor Price	0.198*	0.198*	0.203*	0.203*	–	–
Fuel Price	−0.179*	−0.181*	−0.186*	−0.186*	–	–
Capital Price	−0.019*	−0.017*	−0.017*	−0.016*	–	–
Output	0.067*	−0.060*	0.000	–	–	–
City Pairs Served	0.088*	0.087*	–	–	–	–
Stage Length	0.116*	0.107*	–	–	–	–
Load Factor	−0.171*	−0.162*	–	–	–	–
Capital share equation	R^2 = 0.519	R^2 = 0.517	R^2 = 0.505	R^2 = 0.505	R^2 = 0.195	R^2 = 0.194
Constant	0.130*	0.130*	0.129*	0.128*	0.129*	0.128*
Labor Price	−0.017*	−0.016*	−0.016*	−0.015*	–	–
Fuel Price	−0.039*	−0.039*	−0.038*	−0.038*	–	–
Capital Price	0.056*	0.055*	0.054*	0.053*	–	–
Output	0.010	–	0.008	–	–	–
City Pairs Served	−0.010	–	–	–	–	–
Stage Length	0.024	0.038*	–	–	–	–
Load Factor	0.014	–	–	–	–	–

* Significant at the 0.05 level.
** Significant at the 0.10 level.

E_S, RTS, and RTD measured every 3 years on all the airlines in the sample are shown in Table 3. This table illustrates that at most points in time E_S and RTS are very similar, but early in the sample period, there were significant differences between the two measures for the smaller, faster growing carriers in the sample (Alaska, America West, Southwest, and US Air). Rather than this being an issue with scale economies, it is more likely due to the fact that a rapidly growing airline would need to adjust CPS, ALF, and ASL disproportionately to increases in output in order to maintain efficiency and customer service while growing. Tables 2 and 3 also reveal that at the mean of the data (where the estimated parameters are most accurate) RTS or the transportation definition of scale economies are slightly lower than E_S or the traditional economic definition of scale economies, but under either definition, economies of scale do exist for the largest US airlines.

7. Conclusion

The results of this study show that on average the largest major US airlines have enjoyed increasing RTS for the past 22 years. This alone might lead to the conclusion that airlines are able to charge any price they like at the expense of consumers, however, this does not seem to be the case. After adjusting for inflation, the average revenue per ASM for the airlines in our sample actually declined from \$0.07 in 1987 to \$0.06 in 2009, and average revenue per RSM declined from \$0.12 in 1987 to \$0.07 in 2009 (Transtats, 1987–2009). These yearly averages can be seen in Fig. 1.

A bigger issue than potential extra expenses for consumers is that the presence of economies of scale in the domestic passenger airline market indicates inefficiencies in operations in the sense that expenses per unit of output could be reduced by increasing the scale of operations. As long as per unit expenses can be reduced by increasing output, it would be more efficient to increase output, which would lead to greater profitability for airlines or greater value for consumers from lower priced tickets. Of course, increasing output is only a viable option if that additional output can be sold to consumers, which may not be the case in a competitive marketplace. This existence of economies of scale is one possible explanation of the recent increase in large scale mergers in the US airline industry. Of course, with the slack demand following the terrorist attacks of September 11, 2001 and the bankruptcies of the past decade, the existence of constant returns to scale would be enough to justify such a merger.

The results of this study certainly do not invalidate the results of any previous studies, but they do call into question whether the inclusion of operational characteristic variables in cost functions has confused the issue of returns to scale. For example, if Caves et al. (1984) had calculated elasticity of scale from a traditional economic viewpoint, would they still have found no evidence for returns to sale? Fig. 2 shows that the domestic airline industry was experiencing much more rapid growth until 1987 than after 1987 (ATA, 1976–2010), and Table 3 shows that the biggest difference between the two measures is during periods of rapid growth. Therefore it is possible that Caves et al. (1984) would have seen some evidence for economies of scale if they had used the traditional economic definition. Airlines would have had much less flexibility in managerial decisions such as stage length, load factor, and city pairs served under the regulatory environment that existed during a large part of the sample period of the Caves et al. study (1984). Because of these restrictions, airlines may have found it necessary to make less than optimal decisions in regards to network size, stage length, and load factor when expanding output.

From a policy standpoint, there is no reason to think that the conclusion from the studies leading up to deregulation were inaccurate either. It certainly seems likely that, given the set of constraints they faced, US airlines were operating at an efficient level prior to deregulation. If that were the case, they would not have seen any benefit from increasing output without some change taking place in the input price or regulatory environment. While that is exactly what took place with deregulation in 1979, it was probably several more years before the industry as a whole changed enough that an airline could reduce per unit costs by increasing output and making the necessary adjustments to network size, stage length, and load factor.

This study and its results contribute further to the current body of literature in that it re-examines economies of scale in the airline industry using a newer and more extensive set of data entirely from the post-deregulation era. This data begins after the industry had time to adjust its operation to an unregulated environment and extends through a period of significant growth and concentration in the industry. Furthermore, this study details why the addition of control variables in an estimated model actually restricts the interpretation of results to a variety of very specific scenarios in which zero, one, or more of the control variables are allowed to vary proportionally with output while the others remain constant. These are scenarios which may be of interest bur are unlikely to occur in the real world. Model 2, which was estimated in this study, explains what will happen to costs as output is increased and other managerial factors are allowed to vary in the most efficient way possible whether this includes changing network size, altering route structure, adjusting average load factor, or changing the average length of haul, all of which may be necessary in order to meet the economic assumption of operational efficiency. This study also provides one possible explanation for the recent large scale mergers in the US airline industry. These mergers may be more feasible because of slack demand and bankruptcies but would not make as much financial sense without the possibility of reduced expenses following the merger. In conclusion, this study contributes to the current body of literature on economies of scale in the airline industry by analyzing a current and extensive data set and estimating a model that can be applied to any real world situation where output increases and other factors are adjusted in the most efficient way possible rather than remaining constant or increasing at the same rate as output.

Appendix A. Complete estimation results

See Table A1.

References

Air Transport Association (ATA), 1976–2010. The Annual Report of the US Scheduled Airline Industry. <http://www.airlines.org> (accessed 11.01.10).

Antoniou, Andreas, 1991. Economies of scale in the airline industry: the evidence revisited. Logistics and Transportation Review 27 (2), 159–183.

Baltagi, Badi H., Griffin, James M., Rich, Daniel P., 1995. Airline deregulation: the cost pieces of the puzzle. International Economic Review 36 (1), 245–258.

Basso, Leonardo J., Jara-Diaz, Sergio R., 2005. Calculation of economies of spatial scope from transport cost functions with aggregate output (with an application to the airline industry). Journal of Transport Economics and Policy 39 (1), 25–52.

Basso, Leonardo J., Jara-Diaz, Sergio R., 2006a. Are returns to scale with variable network size adequate for transport industry structure analysis. Transportation Science 40 (3), 259–268.

Basso, Leonardo J., Jara-Diaz, Sergio R., 2006b. Distinguishing multiproduct economies of scale from economies of density on a fixed-size transport network. Network and Spatial Economics 6 (2), 149–162.

Bauer, Paul W., 1990. Decomposing TFP growth in the presence of cost inefficiency, nonconstant returns to scale, and technological progress. Journal of Productivity Analysis 1 (4), 287–299.

Borenstein, Severin, 1992. The evolution of US airline competition. Journal of Economic Perspectives 6 (2), 45–73.

Brueckner, Jan K., Spiller, Pablo T., 1994. Economies of traffic density in the deregulated airline industry. Journal of Law and Economics 37 (2), 379–415.

U.S. Bureau of Labor Statistics (BLS), 1987–2009. Economists and Economic Analysts. <http://www.bls.gov> (accessed 06.01.10).

Bureau of Transportation Statistics (Transtats), 1987–2009. Data Library: Aviation. <http://www.transtats.bts.gov> (accessed 06.01.01).

Civil Aeronautics Board (CAB), Office of Plans, 1972. Unit Costs of the Airlines and Their Determinants. Civil Aeronautics Board, Office of Plans, Washington, DC (March).

Civil Aeronautics Board (CAB), Office of Plans, 1972. Economies of Scale in the Domestic Air Transport Industry. Civil Aeronautics Board, Office of Plans, Washington, DC (November).

Caves, Richard E., 1962. Air Transport and Its Regulators: An Industry Study. Harvard University Press, Cambridge, MA.

Caves, Douglas W., Christensen, Laurits R., Tretheway, Michael W., 1984. Economies of density versus economies of scale: why trunk and local service airline costs differ. Rand Journal of Economics 15 (4), 471–489.

Cherington, Paul W., 1958. Airline Price Policy: A Study of Domestic Airline Passenger Fares, Graduate School of Business Administration. Harvard University, Boston, MA.
Crane, John B., 1944. The economics of air transportation. Harvard Business Review 22 (4), 495–509.
Creel, Michael, Farell, Montserrat, 2001. Economies of scale in the U.S. airline industry after deregulation: a Fourier series approximation. Transportation Research Part E: Logistics and Transportation Review 37 (5), 321–336.
Dempsey, P.S., 1993. Airlines, Aviation, and Public Policy: The Need for Regulatory Reform, Address before the National Commission to Ensure a Strong Competitive Airline Industry, (6/1).
Douglas, George W., Miller, James C., 1974. Economic Regulation of Domestic Air Transport: Theory and Practice. The Brookings Institution, Washington, DC.
Eads, George, Nerlove, Marc, Raduchel, William, 1969. A long-run cost function for the local service airline industry: an experiment in non-linear estimation. Review of Economics and Statistics 51 (3), 258–270.
Farrar, Donald E., Glauber, Robert J., 1967. Multicollinearity in regression analysis: the problem revisited. Review of Economics and Statistics 49 (1), 92–107.
Filipini, Massimo, Maggi, Rico, 1992. The cost structure of the Swiss private railways. International Journal of Transport Economics 19 (3), 307–327.
Formby, John P., Thistle, Paul D., Keeler, James P., 1990. Costs under regulation and deregulation: the case of U.S. passenger airlines. The Economic Record 66 (195), 308–321.
Gillen, David W., Oum, Tae Hoon, Tretheway, Michael W., 1985. Airline Costs and Performance: Implications for Public and Industry Policies. University of British Columbia, Center for Transportation Studies, Vancouver, BC.
Gillen, David W., Oum, Tae Hoon, Tretheway, Michael W., 1990. Airline cost structure and policy implications: a multi-product approach for Canadian airlines. Journal of Transport Economics and Policy 24 (1), 9–34.
Goetz, Andrew R., Sutton, Christopher J., 1997. The geography of deregulation in the US airline industry. Annals of the Association of American Geographers 87 (2), 238–263.
Gordon, Robert J., 1965. Airline costs and managerial efficiency. In: Gordon, Robert J. (Ed.), Transportation Economics: A Conference of the Universities National Bureau Committee for Economic Research. Columbia University Press, New York, NY, pp. 61–91.
Harbeson, Robert W., 1970. Economic status of the local service airlines. Journal of Transport Economics and Policy 4 (3), 309–325.
Harper, Jean H., 1982. So You Want To be an Airline Pilot, Airline Pilot Career Booklet, Aurora, CO.
Hurdle, Gloria J., Johnson, Richard L., Joskow, Andrew S., Werden, Gregory J., Williams, Michael A., 1989. Concentration, potential entry, and performance in the airline industry. The Journal of Industrial Economics 35 (2), 119–139.
Jara-Diaz, Sergio.R., Basso, Leonardo.J., 2003. Transport cost functions, network expansion and economies of scope. Transportation Research Part E: Logistics and Transportation Review 39 (4), 271–288.
Jara-Diaz, Sergio R., Cortes, Cristian E., 1996. On the calculation of scale economies from transport cost functions. Journal of Transport Economics and Policy 30 (2), 157–170.
Johnston, Ahren., Ozment, John., 2011. Concentration in the airline industry: evidence of economies of scale? Journal of Transportation Management 22 (2), 59–74.
Jordan, William A., 1970. Airline Regulation in America: Effects and Imperfections. The Johns-Hopkins Press, Baltimore, MD.
Kahn, Alfred, 1988. Surprises of airline deregulation. American Economic Review 78 (2), 316–322.
Keeler, Theodore E., 1972. Airline regulation and market performance. Bell Journal of Economics 3 (2), 399–424.
Keeler, James P., Formby, John P., 1994. Cost economies and consolidation in the US airline industry. International Journal of Transport Economics 21 (1), 21–45.
Kirby, Michael G., 1986. Airline economics of 'scale' and Australian domestic air transportation policy. Journal of Transport Economics and Policy 20 (3), 339–352.
Kmenta, Jan, Gilbert, Roy F., 1968. Small sample properties of alternative estimators of seemingly unrelated regression. Journal of the American Statistical Association 63 (324), 1180–1200.
Koontz, Harold D., 1951. Economic and managerial factors underlying subsidy needs of domestic trunk line carriers. The Journal of Air Law and Commerce 8 (2), 127–156.
Koontz, Harold D., 1952. Domestic airline self-sufficiency: a problem of route structure. The American Economic Review 42 (1), 103–125.
Kraus, Marvin., 2008. Economies of scale in networks. Journal of Urban Economics 64 (1), 171–177.
Kumbhakar, Subal C., 1990. A re-examination of returns to scale, density and technical progress in U.S. airlines. Southern Economic Journal 57 (2), 428–442.
Kumbhakar, Subal C., 1992. Allocative distortions, technical progress, and input demand in US airlines. International Economic Review 33 (3), 723–737.
Kyle III, Reuben, Phillips, Laurence T., 1985. Airline deregulation: did economists promise too much or too little? Logistics and Transportation Review 21 (1), 13–25.
Murphy, Neil B., 1969. Sources of productivity increases in the US passenger airline industry. Transportation Science 3 (3), 233–238.
Neter, John, Kutner, Michael, Wasserman, William, Nachtsheim, Christopher., 2000. Applied Linear Statistical Models, fourth ed. Irwin, Chicago, IL.
Oum, Tae Hoon, Zhang, Yimin, 1997. A note on scale economies in transportation. Journal of Transport Economics and Policy 31 (3), 309–315.
Panzar, John C., 1989. Technological determinants of firm and industry structure. In: Schmalensee, Richard, Willig, Robert (Eds.), Handbook of Industrial Organization. North Holland, New York, NY, pp. 3–56.
Proctor, Jesse W., Duncan, Julius S., 1954. A regression analysis of airline costs. The Journal of Air Law and Commerce 22 (2), 282–292.
Rakowski, James P., Bejou, David, 1992. Birth, marriage, life, and death: a life-cycle approach for examining the deregulated US airline industry. Transportation Journal 32 (1), 15–29.
Reid, Samuel R., Mohrfeld, James W., 1973. Airline size, profitability, mergers and regulation. Journal of Air Law and Commerce 39 (1), 97–112.
Roy, Roger, Cofsky, Diane, 1985. An Empirical Investigation of Production Technology of Canadian Air Services, Canadian Transport Commission, Research Branch, Economic and Social Research Directorate, Report No. 1985/03E, Ottawa, April.
Sarndal, Carl E., Statton, W. Brent., 1975. Factors influencing operating cost in the airline industry. Journal of Transport Economics and Policy 9 (1), 67–88.
Sarndal, Carl E., Oum, Tae H., Statton, W. Brent, 1978. Further evidence on factors influencing operating costs of the U.S. commercial airlines. Journal of Transport Economics and Policy 12 (1), 47–55.
Shephard, Ronald W., 1953. Cost and Production Functions. Princeton University Press, Princeton, NJ.
Sickles, Robin C., 1985. A nonlinear multivariate error components analysis of technology and specific factor productivity growth with an application to the U.S. airlines. Journal of Econometrics 27 (1), 61–78.
Sickles, Robin C., Good, David, Johnson, Richard L., 1986. Allocative distortions and the regulatory transition of the U.S. airline industry. Journal of Econometrics 33 (1/2), 143–163.
Strazheim, Mahlon R., 1969. The International Airline Industry. The Brookings Institution, Washington, DC.
US Congress, Senate, Committee on the Judiciary, 1975. Civil Aeronautics Board Practices and Procedures, Report of the Subcommittee on Administrative Practice and Procedures, pp. 59–63.
Varian, Hal.R., 1996. Intermediate Microeconomics: A Modern Approach, fourth ed. W.W. Norton and Company, Inc., New York, NY.
Viton, Philip A., 1986. Air deregulation revisited: choice of aircraft, load factors, and marginal-cost fares for domestic air travel. Transportation Research 20A (5), 361–371.
Wheatcroft, Stephen, 1956. The Economics of European Air Transport. The Manchester University Press, Manchester, UK.
Whistler, Diana, White, Kenneth J., Wong, S. Donna, Bates, David, 2004. Shazam Econometrics Software User's Reference Manual Version 10. Northwest Econometrics, Ltd., Vancouver, BC.
White, Robert C., 1979. State sales and use taxes – variations, exemptions, and the airline industry. Journal of Air, Law, and Commerce 45(3)(509), 515–516.
Ying, J., 1992. On calculating cost elasticities. Logistics and Transportation Review 28 (3), 231–235.