The Effect of Market Structure on the Empirical Distribution of Airline Fares

Jeremy A Verlinda[1,2]

October 2005

Abstract

This paper extends the existing literature on the effect of market structure on price dispersion in airline fares by modeling the effect at the disaggregate ticket level. Whereas past studies rely on aggregate measures of price dispersion such as the gini coefficient or the standard deviation of fares, this paper estimates the entire empirical distribution of airline fares and documents how the shape of the distribution is determined by market structure. Specifically, I find that monopoly markets favor a wider distribution of fares with more mass in the tails while duopoly and competitive markets exhibit a tighter fare distribution. These findings indicate that the dispersion of airline fares may result from the efforts of airlines to practice second-degree price discrimination.

[1]Economic Analysis Group, U.S. Department of Justice, 600 E Street, NW - Suite 10000, Washington, D.C. 20530; phone (202) 307-0903, jeremy.verlinda@usdoj.gov.

[2]Jun Ishii, Dale Poirier, Russ Pittman, Chuck Romeo, and Kurt Van Dender provided excellent advice and constructive criticism. This study was partially supported by a fellowship from the University of California Transportation Center at the University of California - Irvine. The views expressed within are mine and are not purported to represent those of the Department of Justice. All mistakes are, of course, my own.

1 Introduction

Airline fares exhibit wide variation across different routes, across different air carriers, and over time. Even on a single flight, travellers throughout the plane will have paid a wide range of fares and few of them will have paid something close to the average fare. Airline fares also tend to be higher in markets that have fewer airlines. The combination of these facts has fascinated the economics profession since at least the early 1990s. Much of the earlier theoretical work attempted to show that this price dispersion could result from the efforts of firms in competitive environments effectively executing price discrimination policies, while much of the more recent work has provided more definitive empirical evidence to support or disprove the claim that the dispersion of airline fares results from price discrimination, usually by demonstrating that as market concentration increases, some measure of price dispersion either increases or decreases. In this paper, I show that market structure affects more than just the standard singular measures of price dispersion. Market structure also influences the entire distribution of airline fares.

The most direct method of exploring a link between price discrimination and market structure would be to collect data on the price-quality schedules offered by various airlines across a wide number of routes. Unlike other price discrimination studies where these price-quality schedules are observed by the econometrician, as in Busse and Rysman (forthcoming), in this study I do not observe the choices available to consumers. This is an explicit feature of the standard airline fare database used in most empirical studies of the airline industry. I do not observe consumer choices, nor do I even observe similar ticket types across markets or time. Whereas most studies handle this complication by collapsing the data into a single statistic at the carrier-route level, in this study I choose to work with the whole distribution of airline fares. I do this with a finite *mixture of normals* model. Usually the mixture of normals model is used to relax the Normality assumption employed in most likelihood-based empirical analyses (see, for example, the discussion in Koop, 2003, or Koop and Tobias, 2004) because the researcher suspects peculiarly fat tails in the distribution or even bimodality in the residuals due to insufficient observable characteristics. But in this study, I exploit the flexibility and tractability of the mixture of normals model in a unique way that allows me to surmount the basic problem that I have an unmatched panel data set.

To anticipate results, I find that the distribution of monopoly fares is both economically and statistically different from the distribution of fares in duopoly and competitive markets. Specifically, the distribution of fares is shifted upwards for monopoly markets, more notably right-skewed, flatter, and fatter tailed. Taken together, these results imply, for example, that the shortest interval of price ranges for any given constant percentage of purchased tickets is greater for monopoly markets than it is for duopoly and competitive markets. The same phenomenon is observed between duopoly and competitive markets, though to a lesser degree.[1] From an economic perspective, the results of the predictive analysis in this paper are consistent with models of second degree price discrimination, i.e., airlines encourage consumers to self-select their preferred tickets based on a trade-off between tickets that more closely match their vertical preferences for quality/scheduling and the price associated with those tickets. This is in contrast to the results of earlier studies that find support for a model of price discrimination where consumers' horizontal preference uncertainty across brands is the primary mechanism driving airlines' pricing decisions.

The paper is organized as follows. Section 2 describes the existing empirical and theoretical literature studying the effect of market structure on price dispersion. In this section I also describe the data used in this study, I introduce a set of definitions to categorize market structure, and I describe a basic breakdown of the data across different market structures. In Section 3 I describe the empirical methodology in more detail, and in Section 4 I explain the main results of the paper. Section 5 concludes and the Appendix provides technical details for the analysis.

2 Market Structure and Price Dispersion

In their seminal paper, Borenstein and Rose (1994; hereafter BR) demonstrated the peculiar phenomenon that the dispersion of airline fares tends to be greater in markets that are less

[1] For example, the smallest interval containing 75% of purchased tickets is predicted to be \$296 for monopoly markets, between \$280 and \$284 for duopoly markets, and \$278 for competitive markets.

concentrated, i.e., in markets with more airlines and for which we would a priori expect greater price competition. This early work has generated a number of additional empirical studies seeking to corroborate and explain the inverse relationship between price dispersion and market concentration, as well as some controversy, since BR argue that their results are consistent with a model of price discrimination where horizontal (interfirm) preference diversity dominates over vertical (intrafirm across ticket quality characteristics). Some of the controversy derives from the discriminatory pricing explanation of price dispersion. Dana (1999) argues that there are possible non-discriminatory sources of price dispersion, in particular that it can arise benignly in a model of costly capacity and demand uncertainty. More importantly, Dana demonstrates an alternative explanation of the BR empirical findings, showing that his model also leads to an inverse relationship between market concentration and price dispersion. The other source of controversy stems from Stole's (1995) theoretical contradiction of the main BR results, claiming instead that the range of prices should increase with market concentration, regardless of whether horizontal or vertical preference uncertainty dominates each firm's pricing decisions.

In this paper, I assume that the source of the controversy is the empirical strategy used to measure the effect of market structure on price dispersion. This is in-line with Stole's suggestion that the BR results are reflective of the averaging technique inherent in collapsing the airline ticket fare data to a single airline-route GINI coefficient that ignores elasticity differences across market segments. Other studies, in particular Liu and Serfes (2005), have attempted more flexible measurement strategies. I argue that there is no reason to collapse the fare data at all. Like other studies, the goal of the analysis is to see how market structure affects the dispersion of airfares. The problem is that different market segments are affected to different extents, but that the researcher cannot actually distinguish each market segment in the data. So, rather than collapse the data to a single statistical measure of dispersion, I work with the entire set of airline fares for a large sample of routes and airlines; my measure of the effect of market structure on prices is to then predict how the entire distribution of fares on a route changes as the market structure changes. In the following subsections, I describe the data that I use for this study, as well as some suggestive results of what the data might be able to tell us by looking at the whole distribution of fares as opposed to single measurement statistics like GINI coefficients or standard deviations.

2.1 Data

The standard data source for most airline pricing studies is the DB1A/B data set provided by the U.S. Department of Transportation (DOT). The data are a random sample of approximately 10% of all airline tickets in the U.S., and include information on the origin and destination airports for each leg of the trip, whether the ticket is restricted or unrestricted, whether it is a roundtrip ticket, the number of passengers flying on that ticket, who the operating airline carrier was, the distance traveled for each leg, as well as various other characteristics. The DOT also provides the T100 database, which contains information at a monthly level on the number of passengers, flights, and seats by aircraft type for each airline and route in the U.S.. I use these two data sets as my primary sources of information in this study. To make the data more tractable, as well as comparable with other studies, I employ only a subset of the data that meet the following criteria:

1. I use the first quarter of 2000 as the time period. The year selection is somewhat arbitrary, although I wanted to avoid any issues associated with post-9/11/2001, in particular the long string of airline bankruptcies and financial problems that resulted afterward. The quarter selection avoids the summer slump that occurs in the 2nd quarter and the holiday jump that occurs in the 4th quarter. Most studies use either the 1st or the 3rd quarters for their data.

2. I only use the top 80% of airline routes based on the number of passengers, and only the top 95% of airlines by passengers, which results in the following airlines: Delta (DL), Southwest (WN), United (UA), American (AA), US Airways (US), Northwest (NW), Continental (CO), TWA (TW), America West (HP), Alaska (AS), AirTran (FL), American Eagle (MQ), Hawaiian (HA), and Aloha (AQ).

3. I exclude tickets with fares less than $30 (which I presume to be purchased with frequent

flier miles), and with fares greater than 5 times the Standard Industry Fare Level (a carryover from airline regulation that associates costs with the distance between routes). Fares in excess of this number are assumed to be data entry errors.

4. I exclude routes with less than 100 ticket observations in the DB1B data. Although the cutoff point is arbitrary, the reason for this exclusion is to provide a reasonably large sample of tickets within a route in order to measure the entire distribution of fares on that route, and not just a single statistic of them.

5. I exclude all tickets that include one or more connecting flights to get to the trip destination, and as such the results are indicative of only direct-flight markets.

Filtering the data in this manner results in 773,811 ticket observations on 1,428 routes and 14 airlines. The number of carriers on each route ranges from 1 to 9 with a median of 2, while the number of routes that each carrier flies on ranges from 10 to 478 with a median of 176. The number of tickets on each route ranges from 100 to 2,688 with a median of 412, and the number of tickets associated with each carrier ranges from 1,732 to 155,123 with a median of 37,597.

It is worth noting the possible adverse effects that could arise from these filters. The most egregious is the elimination of the low-fare tickets which are presumably purchased with frequent flier miles. It is reasonable to assume that travellers prefer to use their frequent flier miles for the most expensive tickets, and to the extent that monopoly routes tend to be the most expensive, the distribution of monopoly fares might be incorrectly shifted upward. However, provided the fraction of tickets purchased with frequent flier miles is only a small percentage of fares, even if the exclusion only affects monopoly markets, it will not be a large impact.[2]

The remaining exclusions will all serve to bias the results away from the hypothesis that monopoly markets exhibit a wider distribution of fares than nonmonopoly markets. Concen-

[2] The percentage of tickets meeting the frequent flier filter is approximately 3%. While I cannot determine if indeed fares would have been higher than average for frequent flier trips, the data do reveal that the the tickets screened out by my frequent flier filter are greater distance trips. Still, I don't expect this to significantly impact my results, as it also the case in these data that monopoly routes actually tend to be shorter distance trips, potentially negating the possible bias.

trating on only the top 95% of airlines and excluding all indirect flights both serve to understate the level of competition in a given market. Thus, some markets will be incorrectly classified as more concentrated than they really are, and a finding that more concentrated markets have a wider distribution of fares would suggest that if I included the remaining air carriers and indirect flights, the difference between the correctly identified distributions would be wider still.

2.2 Empirical Distributions by Market Structure Type

Before discussing the effect of market structure on the distribution of fares, I must first outline how I am going to define market structure. Most empirical studies solely look at the Herfindahl-Hirschman Index (HHI) as their measure of market structure. Although useful, the HHI has only limited ability to differentiate between common market structure types that we might expect to find in the data. BR avoids this issue by defining four categories of markets: Monopoly (one firm with greater than 90% market share), Duopoly (nonmonopoly routes where two firms together have greater than 90% market share), Asymmetric Duopoly (duopoly routes where one carrier has at least 1.5 times greater market share than the other carrier), and Competitive (nonmonopoly, nonduopoly routes); in all cases, BR defines market shares based on the number of flights/week.

While useful, the BR definition cannot capture well certain market structures that most people might accept. For example, if a route was such that one carrier had 50% market share, and ten other firms each had 5%, most people would accept that this is a route that is dominated by primarily one carrier, although they might hesitate to call it a "monopoly" as such. Likewise, if two firms each had 40% of the market, while five other firms each had 4%, we would tend to think of this as a market dominated by two firms, although we might hesitate to call it a duopoly market. I therefore offer an alternative categorization of airline market structure. Let s_1 denote the market share of the largest airline on a route, s_2 denote the market share of the second largest, and so on out to the last carrier on a route, so that $s_1 > s_2 > ... > s_K$. Throughout this paper, I then employ the following four categories, where the titles are merely semantic:

1. Monopoly. A route dominated by one carrier with greater than 50% market share and all other firms are small in comparison. Specifically, $s_1 \geq 0.5$ and $s_1 > 9s_2$.

2. Asymmetric Duopoly. A route dominated by two carriers who together have more than 50% market share while all remaining firms are comparably smaller, but who themselves are sufficiently distinguishable in size. Specifically: $s_1 + s_2 \geq 0.5$, $1.5s_2 < s_1 < 9s_2$, and $s_2 > 9s_3$.

3. Symmetric Duopoly. A route dominated by two carriers who together have more than 50% market share while all remaining firms are comparably smaller, but who themselves are insufficiently distinguishable in size. Specifically: $s_1 + s_2 \geq 0.5$, $s_1 < 1.5s_2$, and $s_2 > 9s_3$.

4. Competitive. A route not dominated by only one or two carriers. Specifically, $s_1 + s_2 < 0.5$ or $s_1 + s_2 \geq 0.5$ but $s_2 < 9s_3$.

Table 1 provides a breakdown of the DB1B fare and T100 passenger data based on these market structure definitions. In this study, I define the market share as the share of passengers enplaned in the quarter for each carrier in each route. As we might expect, average and median fares are unambiguously increasing with market concentration, while the number of passengers appears to be decreasing with market concentration. These results are consistent with textbook economic theory on the market equilibria associated with different market structures. But Table 1 also raises a few questions with regard to the empirical studies listed above that the GINI coefficient is either decreasing or inverse U-shaped with market concentration and that the standard deviation is increasing with market concentration.[3] When we calculate the GINI coefficient and standard deviation for each route and then look at averages across routes (the last section of Table 1), the results are consistent with earlier empirical studies. But when we

[3]The GINI coefficient is essentially a measure of right-skewness. In this paper I employ the following calculation of the GINI coefficient:

$$GINI = \frac{\sum_{i=1}^{n} (2i - n - x_i)}{n^2 \overline{x}},$$

where the x_i are sorted such that $x_1 < x_2 < ... < x_n$. If $GINI = 0$, then all observations are equivalent. As $GINI$ increases towards 1, the distribution of x is said to be more dispersed with most observations at the lower end of the support for x and a few observations near the top end of the support.

calculate the GINI coefficient and standard deviation using all tickets within a given market structure (the first section of Table 1), the results are ambiguous at best. The only summary statistic of the dispersion of prices that is consistent across the two methods is the interquartile range, which we see to be increasing with market concentration. If robust, then based on the theoretical results in BR, this would imply that firms encourage travellers to self-select their airline tickets using price schedules that reflect firms' uncertainties regarding (vertical) consumer preferences for ticket quality.

In this paper, I argue that statistics such as those in Table 1 are insufficient for describing the effect of market structure on fares. Instead, I contend that we should really be examining the distribution of fares to see how it varies with market structure. In Figure 1, I look at how the empirical distribution of fares breaks down across each of the market definitions listed above. For example, the dash-dotted line is a kernel-smoothed plot of the distribution of fares for the collection of all routes in my data that meet the market structure definition above for a monopoly market. We see in Figure 1 that aggregated across all routes in a given market structure type, the range of fares is roughly equivalent and that fares are heavily skewed to the right, as we might expect. We also see that the concentration of fares (that is, the interval of fares containing a fixed percentage of tickets) tends to shift down as the route becomes more competitive, consistent with the notion that more airlines on a route should lead to more competitive pricing on average. It is also worth pointing out that before controlling for any ticket- or route-level characteristics, symmetric duopolies and competitive markets demonstrate nearly indistinguishable fare distributions. Moreover, the one distribution that stands out uniquely from the others is the fare distribution for monopoly markets, which also demonstrates the widest range of fares for any given constant percentage of tickets up to about $900.

Aggregating across all routes smooths out many of the interesting features of the distribution of fares that occurs at the route level. In Figure 2, I contrast the empirical distributions of ticket fares for a subset of routes in my data set. The routes are chosen so as to represent the median route based on the average fare, the standard deviation of fares, and the GINI coefficient within each market structure type. In all cases, the distributions are multimodal and skewed to the right. For the Charlotte/Indianapolis (monopoly) route, the distribution is essentially bimodal, with more of its mass centered on higher fares than any of the other distributions. As with the aggregate results, it is the monopoly route that demonstrates the widest range of fares for any given constant percentage of tickets. In contrast, the La Guardia/Orlando (competitive) route has essentially only 1 mode toward the lower end of the distribution of fares, and is more right-skewed than any of the other distributions. For these route selections, we see that there is little difference in the distribution of fares for the two duopoly route types, which conflicts with the expectation that the symmetric duopoly would have more mass for lower fares and be more right-skewed than the asymmetric duopoly route. There are a couple of possible reasons for this, including the fact that the composition of market shares on the two routes are not markedly different, and in line with Table 1, the mean, standard deviation, and GINI coefficient of the fares on the two routes are very similar.[4] Another possibility is that the LAX/BNA route is mostly dominated by Southwest Airlines, whose aggressive pricing policy might be shifting more of the mass of ticket fares downward than if another carrier had similar market share on this route.

In order to capture these details in a predictive analysis, the remainder of this study attempts to model the effect of market structure on these route-level distributions of fares while controlling for ticket-, route-, and carrier-level characteristics. These include the ticket's restriction type, whether the ticket is for roundtrip travel or not, the number of passengers traveling on a ticket, the distance travelled, the total number of flights on the route, and the relative size of each carrier based on its available seats. This is particularly important if we are concerned that certain market structure types tend to be associated with, say, longer distance routes.[5]

[4] The mean, standard deviation, and GINI coefficient for LAX/BNA are \$361, \$216, and 0.292. For PHX/BWI they are \$374, \$215, and 0.270.

[5] In particular, we might suspect that short-haul routes are best served by a single carrier if there are significant economies in utilizing the same plane on a route for multiple trips on a given day. If, in addition, it is also true that short-haul routes have greater fare dispersion, then we might spuriously conclude that monopolies are associated with a greater spread of fares. The data dispute this. The dispersion of airline fares is greater on longer routes. Moreover, while monopoly routes are on average shorter, the Bayes Factor for the hypothesis that monopoly route distances differ from those of nonmonopolies is only exp(-4.2). More importantly, in a regression of the intequartile range of fares on distance and route types (monopoly, asymmetric duopoly, symmetric duopoly), the Bayes Factor that the coefficients differ for short-haul ($\leq$ 500 miles) versus long-haul ($>$ 500 miles) routes is only exp(-13.6). These Bayes Factors indicate no support for a confounding selection bias with respect to the number of carriers on a route and the distance between cities on a route.

Controlling for these characteristics will allow me to predict how the distribution of airfares might change when market structure changes but the other characteristics do not.[6]

3 Empirically Modelling Unmatched Longitudinal Data

To model the full distribution of prices for the ticket-level data, I consider the following *mixture of normals* specification for ticket i on route j operated by carrier k:

$$y_{ijk} = \delta_k + x_{ijk}\beta + \varepsilon_{ijk}$$

$$p\left(\varepsilon_{ijk}|\pi_j, \alpha_j\right) = \sum_{g=1}^{G} \pi_{gj}\phi\left(\varepsilon_{ijk}; \alpha_{gj}, \eta_{yg}^{-1}\right), \tag{1}$$

where $\pi_j = (\pi_{1j}, ..., \pi_{Gj})'$ and $\alpha_{jk} = (\alpha_{1j}, ..., \alpha_{Gj})'$. In this specification, prices are denoted by y_{ijk} and characteristics that vary at the ticket level are captured in the vector x_{ijk}, including whether the the ticket is a restricted ticket, whether it is a roundtrip ticket, and the number of passengers traveling on the itinerary. The error term is assumed to follow a G-component mixture of normals distribution, with π_{gj} denoting the probability that a ticket on route j falls in component g and $\phi(a; b, c)$ denoting the ordinate of the normal density with mean b and variance c, evaluated at a. Additionally, each component is assumed to have its own mean that is assumed to be unique for each route, as well as its own error precision η_{yg}. Carrier- and route-specific effects are captured in the terms δ_k and α_{gj}. The fact that each route has its own mean within each component is reflective of the data composition. Since the DB1B data do not identify ticket types across routes in any meaningful way, the model as it is currently written describes a pseudo-nonparametric description of ticket-level prices that varies across routes.

If the data did in fact identify ticket types or market segments, then a hierarchical model with individual ticket-type effects could be introduced that would potentially describe the multimodality, skewness and fat tails that I demonstrated across all fares on a given route in Figure 2. It is precisely because of this unmatched longitudinal nature of the data that I introduce

[6] In this study, I solely concentrate on a conditional predictive analysis in an attempt to re-identity the empirical regularities in the data. The issue of estimating a model with endogenous airline entry on each route is assumed to be beyond the scope of the paper.

the mixture of normals distribution for the residual term. The alternative used in the airline literature has been to collapse the data into sample statistics to describe the distribution of fares for a given carrier-route combination, such as the mean, standard deviation, or GINI coefficient. But as I show in Table 1, and as Stole (1995) argues, the inferences we draw from such an approach are sensitive to how the data are aggregated before calculating the statistics. The mixture of normals model allows me to estimate the entire distribution of fares, and additionally to associate the location and shape of the various components of the distribution with observable characteristics such as market structure.

I introduce location effects for each component with the following distributional assumption for the random effects α_{gj}:

$$\alpha_{gj} \overset{ind}{\sim} N\left(a_j \alpha_g^*, \eta_{\alpha g}^{-1}\right) \text{ for } g = 1, ..., G \text{ and } j = 1, ..., J,$$

where a_j is a vector of carrier- and ticket-invariant characteristics that include dummy indicators if the route is a monopoly, symmetric duopoly or asymmetric duopoly, the HHI (Herfindahl index) for routes that are competitive, the number of flights on the route, and the length of the route in miles.[7] Note that this specification also implicitly introduces scale (shape) effects that vary by route analogous to a variance decomposition model. I also assume that the carrier-specific effects δ_k are random, such that

$$\delta_k \overset{ind}{\sim} N\left(b_k \delta^*, \eta_\delta^{-1}\right) \text{ for } k = 2, ...K,$$

where b_k is a vector of characteristics that vary only by carrier, in this case the number of available seats based on the combination of planes flown throughout the quarter for a particular carrier.[8] Analogous to the specification for the route-level effects, this specification for the carrier-level effects introduces a common correlation component over tickets with common

[7] As with my construction of the market structure indicator variables, I use market shares based on passenger enplanements to construct the HHI.

[8] Since I have a complete set of route-level effects for all routes, which are themselves centered with an intercept term, and because I want to include an intercept term in the conditional mean for the carrier-effects, I must exclude one of the carriers. For this analysis, I exclude Southwest Airlines from the list of carrier-effects, which means that each of the estimated carrier-effects represent differences from the average (across routes) Southwest fare after removing the effect of the ticket-level characteristics.

carriers. However, for the sake of parsimony and because the data suggest that the dominant source of asymmetry/skewness in ticket fares is route-specific, the carrier-level effects are assumed to be just a simple normal distribution and not a mixture of normals as with the route-level effects.

In the next section, I describe the posterior results from combining the mixture of normals model with the data. Estimation is executed via a standard application of the Gibbs Sampler with the prior distributions described in Appendix A. Appendix B contains details on the conditional distributions used in the MCMC algorithm and Appendix C summarizes the numerical precision of the Gibbs sampler for this study. An essential part of the analysis is to explore the predictive distributions across different market structures. If we denote by θ the nonlatent parameters in the model and $p(\theta|y)$ their posterior distribution, then the posterior predictive distribution for ticket fares is just[9]

$$p\left(y_{ijk}^f|y, a_j^f, b_k^f, x_{ijk}^f\right) = \int \sum_{g=1}^{G} \pi_{gj} \phi\left(y_{ijk}^f; a_j^f \alpha_g^* + b_k^f \delta^* + x_{ijk}^f \beta, \eta_g^{-1}\right) p(\theta|y)\, d\theta,$$

where the superscript f indicates a forecasted variable and

$$\eta_g^{-1} = \eta_{yg}^{-1} + \eta_{ag}^{-1} + \eta_\delta^{-1}.$$

Given MCMC draws from $p(\theta|y)$, it is a straightforward calculation to derive a Rao-Blackwellized estimate of $p\left(y_{ijk}^f|y, a_j^f, b_k^f, x_{ijk}^f\right)$ and to then plot this distribution over a grid of values for y_{ijk}^f. For this paper, I am primarily interested in comparing the simple hypothesis that market structure has no effect on airfares to the hypothesis that market structure does influence the distribution of airfares. If I let α_l^* denote the coefficients on the market structure variables in a_j, then the no-effect hypothesis is that $\alpha_l^* = 0$. I can therefore calculate the Bayes Factor for this hypothesis versus the hypothesis that there is a market structure effect using the Savage-Dickey density ratio

$$BF = \frac{p(\alpha_l^* = 0|y)}{p(\alpha_l^* = 0)}.$$

[9] It should be noted that this expression only describes the marginal posterior predictive for y_{ijk}^f. The predictive distribution is not independent across j or k, implying that if, for example, we wish to describe the distribution of fares for a route with multiple carriers, there would be nonzero off-diagonal terms in the covariance matrix.

As with the posterior predictive distributions, I use the MCMC draws and the posterior conditional distribution $p(\alpha_l^*|y,\theta)$ described in Appendix B to Rao-Blackwellize the numerator of the Savage-Dickey density ratio; the denominator is known a priori and easily calculated.

4 Market Structure and the Predictive Distribution of Ticket Fares

The main results of this study are contained in Tables 2-3 and in Figure 3.[10] In Table 2, I describe the moments of the marginal posterior distributions for each of the parameters in each component of the mixture distribution. We see that after controlling for route, ticket, and carrier-level characteristics, the market structure variables have an important impact on the distribution of airline fares. To begin with, monopoly fares are uniformly (across all components) greater than the fares for all other markets—the posterior probability that the distribution of monopoly fares exceeds the distribution of fares for all other market structure types is arbitrarily close to one. Asymmetric duopoly fares are also greater than competitive fares (with posterior probabilities of the differences ranging between 0.58 and 1.0 for all components), but not markedly different from symmetric duopoly fares in all components.[11]

Another interesting finding is that the Herfindahl index does not seem to have a large impact on the dispersion of airfares. The posterior means of the HHI coefficients in each component vary only slightly around their common value, whereas for the other market structure parameters there is at least one component that is centered at considerable distance from the others. After controlling for the presence of a competitive market (through the intercept and market structure indicators), it would appear that the concentration of firms only weakly affects the distribution of fares. This is consistent with the finding that the distribution of fares is not strongly distinguishable between symmetric and asymmetric duopolies. Still, the

[10] The results of the analysis are conditioned on a 5-component mixture model. At minimum, given the plots in Figures 1 and 2, I require 3 components to capture the kurtosis and skewness of the observed distributions. Moreover, if the observable characteristics are inadequate to capture all of the features in the empirical distributions described in Figure 3, I potentially require even more than 3 components.

[11] The posterior probabilities of a positive difference between asymmetric and competitive routes are, for each component respectively, 0.58, 0.93, 0.96, 0.84 and 1.0. The differences between asymmetric and symmetric duopoly route effects are 0.60, 0.61, 0.75, 0.61, and 0.92.

various market structure definitions are capturing significant variation in the distribution of airline fares—the log-Bayes Factor for the hypothesis that market structure has an effect on the distribution of fares versus the hypothesis that it has no effect is 87. A log-Bayes Factor of this order is overwhelming evidence that market structure influences the distribution of airline fares.

To derive the posterior predictive distribution of airline fares, I must either condition on specific route and carrier effects, as well as ticket characteristics, or integrate these features out to try and capture what I think is the underlying effect of the market structure variables. In this analysis, I choose to integrate out the route and carrier effects. I also integrate out the conditioning characteristics restricted ticket status and roundtrip status based on their joint incidences in the data. I condition the predictive analysis on the following remaining characteristics: the number of passengers traveling on the ticket is 2, the number of flights and the distance travelled on the route are both fixed at their mean values, the size of the airline(s) on the route is (are) fixed at the mean value for the monopoly market structure, 0.7 and 0.3 times the mean value for the asymmetric duopoly, 0.5 times the mean value for each airline in the symmetric duopoly, and one-third the mean value for the competitive market structure.

In Figure 3 I plot the posterior predictive distributions of airline fares for each of the market structures based on these choices. The main results demonstrated in Figure 3 are mostly consistent with the simple kernel-smoothed histograms described in Section 2. We see that as market concentration increases, the distribution of airline fares (controlling for the other characteristics) shifts to the right, and, at least for the monopoly case, demonstrates a wider dispersion of fares for any given constant percentage of tickets. One way to see this is to compare the coefficient parameters in component 5 across each of the market structure types, where we see fares in this component \$103 greater than asymmetric duopoly routes and \$134 greater than symmetric duopoly routes. All of the distributions exhibit significant non-normality near the mode—they are all heavily skewed to the right, and the tails are much fatter than we might get from even a Student-t distribution.[12] Also, consistent with Table 2,

[12] As we move away from the mode, the tails start to exhibit more Gaussian-like behavior. This is an explicit feature of the model, since the components at the extreme ends of the distribution are each conditionally normal

we see that the fare distributions for both types of duopoly markets lie to the right of the fare distribution for competitive markets, and that the fare distribution for asymmetric duopolies is only slightly to the right of the fare distribution for symmetric duopolies. Again, it seems that including variables for the basic market structure—one, two, or more firms dominating a market—captures most of the observable variation in the distribution of airline fares.

These findings suggest, as I described earlier based solely on the descriptive characteristics of the data, that the predictive distributions are consistent with second degree price discrimination in oligopoly markets. This conclusion is based on the theoretical results described in the Stole (1995) and Liu and Serfes (2005) models which have dispersion increasing with market concentration. Conditional on the theoretical results in BR, the results suggest that the vertical (ticket quality) preference diversity of consumers dominates the nonlinear price scheduling of air carriers, and not horizontal (brand) preferences as documented empirically by BR. The findings here are consistent with the conclusions reached in Liu and Serfes (2005) based on their empirical results regarding the relationship between market concentration and the standard deviation of fares. Indeed, the summary of the posterior predictive densities provided in Table 4 directly supports the Liu and Serfes results. Not only is the distribution of airfares shifting to the right as market concentration increases, it is also becoming flatter. The estimated posterior standard deviations are increasing in market concentration, and more importantly, the length of the highest posterior density intervals (intervals of shortest length for a given percentage of the distribution) are also increasing with market concentration. These results also directly contradict the predictions in Dana's (1999) non-discriminatory model of price dispersion, which predicts that the support of the price distribution decreases with market concentration and that the standard deviation of prices also falls with market concentration.

Most of the control variables have signs that are consistent with the standard economic literature. The number of flights on a route is a potential proxy for a demand shifter (and a potential instrument for HHI if we are concerned with possible endogeneity problems), and indeed we see a strong positive effect between the number of flights and airline fares. The

distributed, which ensures propriety of the predictive distributions and the existence of moments.

distance between cities on a route has the largest economic effect on the distribution of airline fares, shifting each component upward by anywhere from \$40 to \$314, reflecting the higher costs for airlines to service routes that are further apart, and also indicating that the range of fares is itself increasing with the distance between cities on a route. The remaining unobserved variation at the route level is significant, with standard deviations in each component ranging from \$66 to \$255, suggesting the importance of controlling for unobserved route-level effects in the analysis.

In Table 3 I summarize the carrier- and ticket-level covariates. At the carrier level I include only one covariate to preserve parsimony of the model. Since Southwest Airlines is the excluded carrier-effect, the coefficient on available seats should be interpreted as the effect of a larger fleet size on the deviation from the average Southwest residual fare. If we were to suppose that airlines with larger fleets would have lower per-ticket or per-flight costs, then we would expect to see that as available seats increase, airlines that normally charge a higher price than Southwest should charge a differential that is otherwise closer to the Southwest fare. In other words, the coefficient on available seats would then be negative. Instead, it appears that an increase in available seats has a positive effect on this differential, with a posterior probability that the effect is negative of only 0.09. One possible explanation of this is that the airlines with the largest fleets are also the oldest airlines with more entrenched labor costs, and that they tend to have much higher costs than some of the airlines with smaller fleets. At any rate, there is not a great deal of inter-carrier variation in the deviation from Southwest fares that remains unobserved relative to the unobserved variation at the route-level, since the posterior standard deviation is only \$22.

In contrast, all of the ticket-level characteristics are easily explained as standard features of the airline industry. Restricted tickets tend to be cheaper than unrestricted tickets, while roundtrip tickets tend to be more costly for travellers than one-way trips. Tickets that include only one passenger, presumably business-travel tickets, tend to be higher than tickets with multiple passengers, although the discounts associated with extra travellers are not significant unless the number of passengers on the ticket exceeds 10. Lastly, allowing for route-specific and carrier-specific effects seems to have captured a sizable percentage of the unobserved variation in fares overall. The standard deviation of fares in the data is \$333, while the posterior means of the residual standard deviations in each component range from \$58 to \$210. Moreover, a back-of-the-envelope calculation yields an approximate standard deviation for the entire distribution of the residuals of approximately \$207.

5 Conclusion

In this paper I explore the effect of market structure on the entire distribution of airline fares. I find that, after controlling for carrier-, route-, and ticket-level characteristics, monopoly markets exhibit higher fares and a wider range of fares for a constant percentage of tickets than both duopoly and competitive markets, and that duopoly markets have a similar but diminished relationship with respect to competitive markets. These findings imply that as market concentration on a route increases, not only do consumers tend to pay more for their tickets, but the range of fares paid by similar percentages of consumers within each market structure type is wider. This result is consistent with theoretical models of second-degree price discrimination.

Appendix

A Specifying a Prior Distribution for the Unmatched Longitudinal Model

To ease computation within the Gibbs Sampler, I employ independent normal-gamma priors for the model specified in Section 3. Specifically, I condition on

$$\beta \sim N\left(\underline{\beta}, \underline{Q}_{\beta}^{-1}\right), \ \eta_{yg} \overset{ind}{\sim} G\left(\frac{\underline{\omega}_{yg}}{2}, \frac{2}{\underline{\mu}_{yg}}\right),$$

$$\eta_{\alpha g} \overset{ind}{\sim} G\left(\frac{\underline{\omega}_{\alpha g}}{2}, \frac{2}{\underline{\mu}_{\alpha g}}\right),$$

$$\delta^* \sim N\left(\underline{\delta}^*, \underline{Q}_{\delta}^{-1}\right), \ \eta_{\delta} \sim G\left(\frac{\underline{\omega}_{\delta}}{2}, \frac{2}{\underline{\mu}_{\delta}}\right),$$

and

$$\pi_j = [\pi_{1j}, ..., \pi_{Gj}] \sim Dir\left(\underline{\rho}_j\right),$$

where Dir denotes the Dirichlet distribution. Lastly, for identification purposes, I impose a labeling restriction on the first coefficient in each of the α_g^* through the joint distribution of the α_g^*:

$$p\left(\alpha^* = \left[\alpha_1^{*\prime}, ..., \alpha_G^{*\prime}\right]'\right) \propto \prod_{g=1}^{G} \phi\left(\alpha_g^*; \underline{\alpha}_g^*, \underline{Q}_{\alpha g}^{-1}\right) \mathbb{I}\left(\alpha_{11} < \alpha_{21} < \ldots < \alpha_{G1}\right),$$

where $\mathbb{I}(a \in A) = 1$ if $a \in A$ and yields 0 otherwise.

Regarding choice of prior hyperparameters, I follow what has become largely standard in hierarchical linear regression models when a diffuse prior is desired. I first standardize all of the nondiscrete independent variables in the analysis (by subtracting the sample mean and dividing by the sample standard deviation). Then, for all of the coefficient parameters (β, α^*, and δ^*), I set the prior mean equal to zero and the prior variance equal to 1600 times the identity matrix. I set the intercept parameters in α^* for all components at 400, which implies that my prior predictive densities are symmetric with Gaussian tails, centered at \$400. Under this parameterization, my prior expected value for the effect of each covariate is that increasing the level by one standard deviation will have no effect on the distribution of airfares.

However, since I want to avoid shrinking the covariate effects too much toward having no effect, I specify relatively large prior variances on the coefficient parameters which, nevertheless, preserve propriety of my prior distributions.

For each of the components of the residual precisions, I set $\underline{\omega}_{yg} = 3$ and $\underline{\mu}_{yg} = 2500$, which centers the prior for the residual standard deviation at approximately \$50 in each component but is otherwise extremely flat. For each of the components of the route-effect precisions, I set $\underline{\omega}_{\alpha g} = 3$ and $\underline{\mu}_{\alpha g} = 100$, which centers the prior for the route-effect standard deviation at approximately \$10 in each component, but is otherwise extremely flat. For the carrier-level effects, I set $\underline{\omega}_{\delta} = 3$ and $\underline{\mu}_{\delta} = 4$, which centers the prior for the standard deviation of the carrier-level effects at approximately \$2. Note that in each case for the route- and carrier-level effects, I have centered the prior standard deviations at relatively small numbers and very flat distributions. This contrasts with the usual notion that we should make the prior on residual variances large if we wish to center our beliefs on the regression coefficients at zero, and is a direct result of not wishing to inadvertently introduce extra unobserved variation at the route and carrier levels.

Regarding prior sensitivity, I explore the impact of multiplying and dividing the prior variances on the regression coefficients by four. Figure 5 plots the predictive density for airfare based on these choices. These densities are indeed sensitive to the prior variance of the coefficient estimates, and are suggestive of the influence the prior variances can have on the analysis when they are too small. I condition the analysis in this paper on the case where the prior variances are each equal to 1600, which covers the expected support of airline fares reasonably well, and in particular excludes any significant mass for negative airfares.

B Posterior Conditional Distributions for the Gibbs Sampler

In this section I describe the full set of posterior conditional distributions that I use for estimation in the Gibbs Sampling algorithm. A standard procedure in modeling normal mixtures models is to introduce a latent component labeling vector $c_{ij} = [c_{1ij}, ..., c_{Gij}]$, where $c_{gij} = 1$ if ticket i on route j is associated with component g and is equal to 0 otherwise. The model then becomes

$$p\left(\varepsilon_{ijk}|c_{ij}\right)=\prod_{g=1}^{G}\left[\phi\left(\varepsilon_{ijk};\alpha_{gj},\eta_{yg}^{-1}\right)\right]^{c_{gij}}$$

$$p\left(c_{ij}\right)=\prod_{g=1}^{G}\left[\pi_{gj}\right]^{c_{gij}}.$$

We recognize the marginal distribution of c_{ij} as a Multinomial distribution with one trial, i.e., $c_{ij}\sim Mult\left(1,\pi_{j}\right)$, and note additionally that the marginal distribution for ε_{ijk} is still given by equation (1). Throughout, I employ the notational convention that θ is the vector of nonlatent parameters in the model while

$$\Gamma=\left[\{\alpha_{gj}\},\{\delta_{k}\},\{c_{gijk}\},\theta'\right]'$$

is the vector of all latent and nonlatent parameters. Furthermore, I denote by Γ_{-x} the vector of all model parameters except x. Additional notational requirements concern the number of observations associated with a given component, route or carrier. To that end, let $n_{gj}=\sum_{i}c_{gij}$, and $n_{g}=\sum_{j}n_{gj}$ denote the number of observations within component g and broken down by route j. I also define n_{jk} as the number of ticket observations on route j and carrier k, and n_{k} as the number of ticket observations associated with carrier k, so that: $n_{j}=\sum_{k}n_{jk}=\sum_{g}n_{gj}$ and $n=\sum_{k}n_{k}=\sum_{j}n_{j}=\sum_{g}n_{g}$, which is the total number of unique observations in the data set. Lastly, letting the number of airlines operating on route j be K_{j} and the number of routes that airline k operates on be J_{k}, I denote the total number of carrier-route observations as $L=\sum_{j}K_{j}=\sum_{k}J_{k}$, the total number of unique airlines as K, and the total number of unique routes as J.

As is often the case in Bayesian analyses, computation is aided by conditioning on the latent data. I therefore work throughout with the conditional likelihood

$$p\left(y|\Gamma\right)=\prod_{j=1}^{J}\prod_{k=1}^{K_{j}}\prod_{i=1}^{n_{jk}}\prod_{g=1}^{G}\left[\phi\left(y_{ijk};\alpha_{gj}+\delta_{k}+x_{ijk}\beta,\eta_{yg}^{-1}\right)\right]^{c_{gijk}},$$

where $y=[y_{ijk}]$ and all stacking is done first by all ticket observations associated with carrier k on route j, then each of these stacked by carrier on route j, then over j, and finally, over g where appropriate. At times, it will also be convenient to stack the latent parameters $\{c_{g}\}$ in matrix form, so that we have instead the $n \times G$ matrix $\widetilde{c} = [c_1, ...c_G]$. An additional notational convenience will be to let $\sum\limits_{ijk}$ denote the more formal triple summation $\sum\limits_{j=1}^{J}\sum\limits_{k=1}^{K_j}\sum\limits_{i=1}^{n_{jk}}$.

Loosely following Koop (2003), the conditional posterior $\beta|y, \Gamma_{-\beta} \sim N\left(\overline{\beta}, \overline{Q}_\beta^{-1}\right)$ has parameters

$$\overline{Q}_\beta = \underline{Q}_\beta + \sum_{ijk}\sum_{g} c_{gijk}\eta_{yg}x'_{ijk}x_{ijk}$$

and

$$\overline{\beta} = \overline{Q}_\beta^{-1}\left(\underline{Q}_\beta\underline{\beta} + \sum_{ijk}\sum_{g} c_{gijk}\eta_{yg}x'_{ijk}\left(y_{ijk} - \alpha_{gjk} - \delta_k\right)\right).$$

This is the standard conditional posterior in linear models, augmented to allow $\sum_g c_{gijk}$ to pick off the correct component intercept and precision. However, since programs like Matlab are optimized in such a way that they favor vector notation over loops, the conditional posterior can be rewritten to eliminate the summation notation. If we let $\widetilde{\alpha}_{ij} = \sum_g c_{gij}\alpha_{gj}$, $\widetilde{\alpha} = [\widetilde{\alpha}_j]$, and $\eta_y = [\eta_{y1}, ..., \eta_{yG}]'$, then the parameters of the conditional posterior for β are

$$\overline{Q}_\beta = \underline{Q}_\beta + X' diag\left(\widetilde{c}\eta_y\right)X$$

and

$$\overline{\beta} = \overline{Q}_\beta^{-1}\left(\underline{Q}_\beta\underline{\beta} + X' diag\left(\widetilde{c}\eta_y\right)\left(y - \widetilde{\alpha}\right)\right).$$ [13]

The conditional posterior $\eta_{yg}|y, \Gamma_{-\eta_{yg}} \sim G\left(\frac{\overline{\omega}_{yg}}{2}, \frac{2}{\overline{\mu}_{yg}}\right)$ has the following parameters:

$$\overline{\omega}_{yg} = \underline{\omega}_{yg} + n$$

and

$$\begin{aligned}\overline{\mu}_{yg} &= \underline{\mu}_{yg} + \sum_{ijk} c_{gijk}\left(y_{ijk} - \delta_k - \alpha_{gjk} - x_{ijk}\beta\right)^2 \\ &= \underline{\mu}_{yg} + \left(y - \mathcal{K}\delta - \mathcal{J}\alpha_g - X\beta\right)' diag\left(c_g\right)\left(y - \mathcal{K}\delta - \mathcal{J}\alpha_g - X\beta\right),\end{aligned}$$

[13] For large data sets, for which this is one, the available system memory to store the diagonal matrix $diag\left(\widetilde{c}\eta_y\right)$ will be quickly overwhelmed. However, since this matrix will only contain n nonzero values and $n\left(n-1\right)$ zero values, a useful tool here is to employ Matlab's "sparse" matrices tool, which only stores locations and values of the nonzero elements of a matrix. For example, in a simulated exercise I worked with $n = 800{,}000$. The memory required to store $diag\left(\widetilde{c}\eta_y\right)$ with this n is approximately 4.66 Terabytes. Using Matlab's sparce matrix, this is shrunk to approximately 6.87 Megabytes. The sparse tool proves highly useful in storing $\widetilde{c}$ as well.

where $\mathcal{K}$ is a $n \times (K-1)$ matrix of dummy indicators across carriers and $\mathcal{J}$ is a $n \times J$ matrix of dummy indicators across routes. The conditional posterior for the latent parameters $\{\alpha_{gj}\}$ is

$$\alpha_{gj}|y, \Gamma_{-\alpha_{gj}} \sim N\left(D_\alpha^{-1} d_\alpha, D_\alpha^{-1}\right) \text{ for } j = 1, ..., J, \text{ and } g = 1, ...G,$$

where

$$D_\alpha = \eta_{\alpha g} + \eta_{yg} n_{gj},$$

$$\begin{aligned} d &= \eta_{\alpha g} a_j \alpha_g^* + \eta_{yg} \sum_{k=1}^{K_j} \sum_{i=1}^{n_{jk}} c_{gij} \left(y_{ijk} - \delta_k - x_{ijk}\beta\right) \\ &= \eta_{\alpha g} a_j \alpha_g^* + \eta_{yg} n_{gj} \overline{\left[y_j - \mathcal{K}^{(j)}\delta - x_j\beta\right]_g}, \end{aligned}$$

$\mathcal{K}^{(j)}$ is a $n_j \times (K-1)$ matrix of dummy indicators for the carriers on route j, and $\overline{\left[y_j - \mathcal{K}^{(j)}\delta - x_j\beta\right]_g}$ is the sample average (over i, k) of the residuals in component g.

To derive the conditional posteriors for δ_k, I require the notation $\mathcal{K}_k$ to denote the k^{th} column of $\mathcal{K}$. It follows that

$$\delta_k|y, \Gamma_{-\delta_k} \sim N\left(D_\delta^{-1} d_\delta, D_\delta^{-1}\right) \text{ for } k = 2, ..., K,$$

where

$$D_\delta = \eta_\delta + \sum_{j=1}^{J_k} \sum_{i=1}^{n_{jk}} \sum_{g=1}^{G} c_{gij} \eta_{yg},$$

$$d_\delta = \eta_\delta b_k \delta_g^* + \sum_{j=1}^{J_k} \sum_{i=1}^{n_{jk}} \sum_{g=1}^{G} c_{gij} \eta_{yg} \left(y_{ijk} - \alpha_{gj} - x_{ijk}\beta\right),$$

which can be rewritten as

$$D_\delta = \eta_\delta + \mathcal{K}_k' diag\left(\widetilde{c}\eta_y\right) \mathcal{K}_k,$$

and

$$d_\delta = \eta_\delta b_k \delta_g^* + \mathcal{K}_k' diag\left(\widetilde{c}\eta_y\right)\left(y - \widetilde{\alpha} - X\beta\right).$$

The posterior conditional for α_g^* is

$$p\left(\alpha^*|y, \Gamma_{-\alpha_g^*}\right) \propto \left[\prod_{g=1}^{G} \phi\left(\alpha_g^*; \overline{\alpha}_g^*, \overline{Q}_{\alpha g}^{-1}\right)\right] \mathbb{I}\left(\alpha_{11} < \alpha_{12} < \ldots < \alpha_{1G}\right)$$

where

$$\overline{Q}_{\alpha g} = \underline{Q}_{\alpha g} + \eta_{\alpha g} A' A$$

and

$$\overline{\alpha}^*_g = \overline{Q}^{-1}_{\alpha g} \left(\underline{Q}_{\alpha g} \underline{\alpha}^*_g + \eta_{\alpha g} A' \alpha_g \right).$$

We also have $\eta_{\alpha g} | y, \Gamma_{-\eta_{\alpha g}} \sim G\left(\frac{\overline{\omega}_{\alpha g}}{2}, \frac{2}{\overline{\mu}_{\alpha g}} \right)$, where

$$\overline{\omega}_{\alpha g} = \underline{\omega}_{\alpha g} + J$$

and

$$\overline{\mu}_{\alpha g} = \underline{\mu}_{\alpha g} + \left(\alpha_g - A\alpha^*_g \right)' \left(\alpha_g - A\alpha^*_g \right).$$

Conditional on δ, we have a standard linear regression model, and so we get $\delta^* | y, \Gamma_{-\delta^*} \sim N\left(\overline{\delta}^*, \overline{Q}_{\delta}^{-1} \right)$ and $\eta_\delta | y, \Gamma_{-\eta_\delta} \sim G\left(\frac{\overline{\omega}_\delta}{2}, \frac{2}{\overline{\mu}_\delta} \right)$, where

$$\overline{Q}_\delta = \underline{Q}_\delta + \eta_\delta B' B,$$

$$\overline{\delta}^* = \overline{Q}^{-1}_{\varphi g} \left(\underline{Q}_\delta \underline{\delta}^* + \eta_\delta B' \delta \right),$$

$$\overline{\omega}_\delta = \underline{\omega}_\delta + K - 1$$

and

$$\overline{\mu}_\delta = \underline{\mu}_\delta + \left(\delta - B\delta^* \right)' \left(\delta - B\delta^* \right).$$

To obtain the latent $\{c_{gijk}\}$, note that

$$p\left(c_{ij} | y, \Gamma_{-c_{ij}} \right) \propto \prod_{g=1}^{G} \left[\pi_{gj} \right]^{c_{gij}} \left[\phi\left(y_{ijk}; \delta_k + \alpha_{gj} + x_{ijk}\beta, \eta_{yg}^{-1} \right) \right]^{c_{gij}},$$

which implies that $c_{ij} | y, \Gamma_{-c_{ij}} \sim Mult\left(1, \overline{\pi}_{ij} \right)$, where

$$\overline{\pi}_{gij} = \frac{\pi_{gj} \phi\left(y_{ijk}; \delta_k + \alpha_{gj} + x_{ijk}\beta, \eta_{yg}^{-1} \right)}{\sum_{g=1}^{G} \pi_{gj} \phi\left(y_{ijk}; \delta_k + \alpha_{gj} + x_{ijk}\beta, \eta_{yg}^{-1} \right)}.$$

Lastly, the posterior conditional $\pi_{gj} | y, \Gamma_{-\pi_{gj}} \sim Dir\left(\overline{\rho}_j \right)$ has parameters

$$\overline{\rho}_{gj} = \underline{\rho}_{gj} + n_{gj} \quad \text{for} \quad g = 1, ..., G \quad \text{and} \quad j = 1, ..., J.$$

C Numerical Precision of the Gibbs Output

The numerical standard errors in Tables 2-3 and the plots in Figure 4 are provided in order to demonstrate the effectiveness of the Gibbs Sampling algorithm in estimating the model described in Section 3. In all cases, the moments of the marginal posterior distributions for all of the parameters in the model are estimated with relatively high precision based on the numerical standard errors. I calculate the numerical standard errors based on the "batching" method described in Carlin and Louis (2000). As we can see in Figure 4 though, the Gibbs chain mixes fairly well—for most of the parameters, the autocorrelation function drops to below 0.1 after about 20 lags, and in all cases drops below 0.5 after about 20 draws. Figure 4 also demonstrates how quickly the Gibbs sampler converges to the posterior distribution. For this study, I run 2 parallel chains of the sampler each for 1500 draws and discard the first 500 in each as burn-in, which seems very reasonable based on the draws illustrated in Figure 4.

Tables & Figures

Table 1

Summary Statistics by Market Structure of the DB1B/T100 Data

	Monopoly	Asymmetric Duopoly	Symmetric Duopoly	Competitive
Airline Fare (aggregate)				
Min	31	31	31	31
Mean	431	404	391	375
Median	325	297	276	273
Max	2872	3398	3266	3713
IQR	325	308	288	263
SD	330	332	348	329
Gini	0.380	0.401	0.420	0.407
Passengers Enplaned (by route)				
Min	2,307	15,183	20,043	25,721
Mean	52,423	82,911	82,167	136,470
Median	44,306	70,181	65,579	122,460
Max	232,400	362,160	451,010	401,180
IQR	32,132	57,790	51,205	86,466
SD	29,285	51,651	64,864	75,294
Airline Fare (by route)				
Min	60	46	47	38
Mean	395	359	358	324
Median	325	297	294	272
Max	1070	1142	1146	1194
IQR	305	279	282	240
SD	228	211	209	193
Gini	0.284	0.290	0.287	0.293

Table 2

Posterior Distribution of Route Effects Parameters in the Mixture of Normals Model

Route Effects		Component 1 Moment Estimate	Component 1 Numerical Std. Error	Component 2 Moment Estimate	Component 2 Numerical Std. Error	Component 3 Moment Estimate	Component 3 Numerical Std. Error	Component 4 Moment Estimate	Component 4 Numerical Std. Error	Component 5 Moment Estimate	Component 5 Numerical Std. Error
Intercept	Post. Mean	231.402	(0.931)	276.114	(0.996)	289.386	(0.717)	316.724	(0.799)	523.407	(0.953)
	Post. Std. Dev.	13.507	(0.100)	14.657	(0.106)	11.269	(0.147)	12.658	(0.157)	17.032	(0.260)
Monopoly	Post. Mean	-6.537	(1.015)	42.830	(1.424)	33.520	(1.423)	40.941	(0.833)	160.855	(1.063)
	Post. Std. Dev.	14.787	(0.096)	20.819	(0.136)	21.072	(0.159)	13.897	(0.171)	20.102	(0.297)
Symmetric Duopoly	Post. Mean	-15.419	(0.972)	10.950	(1.749)	-7.641	(1.014)	-7.443	(1.074)	26.704	(0.920)
	Post. Std. Dev.	14.682	(0.100)	26.134	(0.174)	17.080	(0.174)	18.339	(0.194)	22.183	(0.306)
Asymmetric Duopoly	Post. Mean	-13.764	(0.920)	13.944	(1.235)	0.981	(1.165)	-2.269	(0.876)	58.217	(0.951)
	Post. Std. Dev.	13.649	(0.097)	18.604	(0.148)	17.952	(0.152)	14.779	(0.159)	19.868	(0.276)
Competitive * HHI	Post. Mean	-46.632	(1.922)	-21.053	(1.432)	-67.665	(1.575)	-47.566	(1.287)	-42.598	(1.066)
	Post. Std. Dev.	30.408	(0.258)	28.881	(0.405)	30.531	(0.383)	29.189	(0.401)	35.091	(0.576)
Flights (in 000's)	Post. Mean	-2.927	(0.093)	13.276	(0.359)	8.645	(0.268)	17.202	(0.148)	29.961	(0.254)
	Post. Std. Dev.	2.411	(0.037)	6.101	(0.072)	5.187	(0.060)	4.671	(0.079)	8.493	(0.146)
Distance (in 000's miles)	Post. Mean	39.066	(0.150)	154.364	(3.119)	137.968	(0.781)	191.095	(0.189)	313.772	(0.536)
	Post. Std. Dev.	2.786	(0.031)	44.097	(0.062)	11.449	(0.055)	4.467	(0.064)	10.048	(0.107)
Standard Deviation	Post. Mean	65.940	(0.213)	111.362	(1.388)	117.181	(0.539)	135.128	(0.230)	255.461	(0.650)
	Post. Std. Dev.	3.293	(0.025)	19.698	(0.047)	7.923	(0.038)	4.104	(0.046)	10.333	(0.090)

* The (1,1) element for each parameter indicates the posterior mean; the (1,2) element the numerical standard error (in parentheses) for the estimated posterior mean; the (2,1) element the posterior standard deviation; and the (2,2) element the numerical standard error (in parentheses) for the estimated posterior standard deviation.

Table 3

Posterior Distribution of Carrier- & Ticket-level Parameters in the Mixture of Normals Model

Carrier Effects		Moment Estimate	Numerical Std. Error	Component Mixing Weights	Moment Estimate	Numerical Std. Error
Intercept	Post. Mean	6.538	(0.148)	Component 1	0.268	N/A
	Post. Std. Dev.	9.054	(0.154)		0.010	N/A
Available Seats (in millions)	Post. Mean	6.201	(0.119)	Component 2	0.142	N/A
	Post. Std. Dev.	6.717	(0.129)		0.017	N/A
Standard Deviation	Post. Mean	22.398	(0.121)	Component 3	0.148	N/A
	Post. Std. Dev.	4.611	(0.092)		0.009	N/A
				Componet 4	0.127	N/A
					0.004	N/A
				Component 5	0.316	N/A
					0.001	N/A
Ticket-Level Effects				**1st-Level Residual Standard Deviation**		
Restricted	Post. Mean	-43.287	(0.0196)	Component 1	65.030	(0.054)
	Post. Std. Dev.	0.389	(0.0046)		0.777	(0.003)
Roundtrip	Post. Mean	42.303	(0.0177)	Component 2	59.221	(0.197)
	Post. Std. Dev.	0.376	(0.0049)		2.791	(0.006)
1 Passenger	Post. Mean	3.124	(0.0054)	Component 3	62.014	(0.096)
	Post. Std. Dev.	0.257	(0.0047)		1.380	(0.006)
Num. Passengers (2-10)	Post. Mean	0.002	(0.0012)	Component 4	58.427	(0.122)
	Post. Std. Dev.	0.057	(0.0010)		1.741	(0.006)
Num. Passengers (>10)	Post. Mean	-0.028	(0.0001)	Component 5	210.391	(0.077)
	Post. Std. Dev.	0.003	(0.0001)		1.126	(0.005)

* The (1,1) element for each parameter indicates the posterior mean; the (1,2) element the numerical standard error (in parentheses) for the estimated posterior mean; the (2,1) element the posterior standard deviation; and the (2,2) element the numerical standard error (in parentheses) for the estimated posterior standard deviation. The reported mixture component weights are average values for the posterior means and standard deviations of the route-specific component weights.

Table 4

Posterior Predictive Distributions of Fares by Market Structure Type

	Posterior Mean	Posterior Median	Posterior Std. Dev	75% HPD Interval		Interval Length
Monopoly	276	249	187	84	468	384
Asymmetric Duopoly	247	232	157	72	428	356
Symmetric Duopoly	243	224	155	68	420	352
Competitive	240	221	152	66	410	344

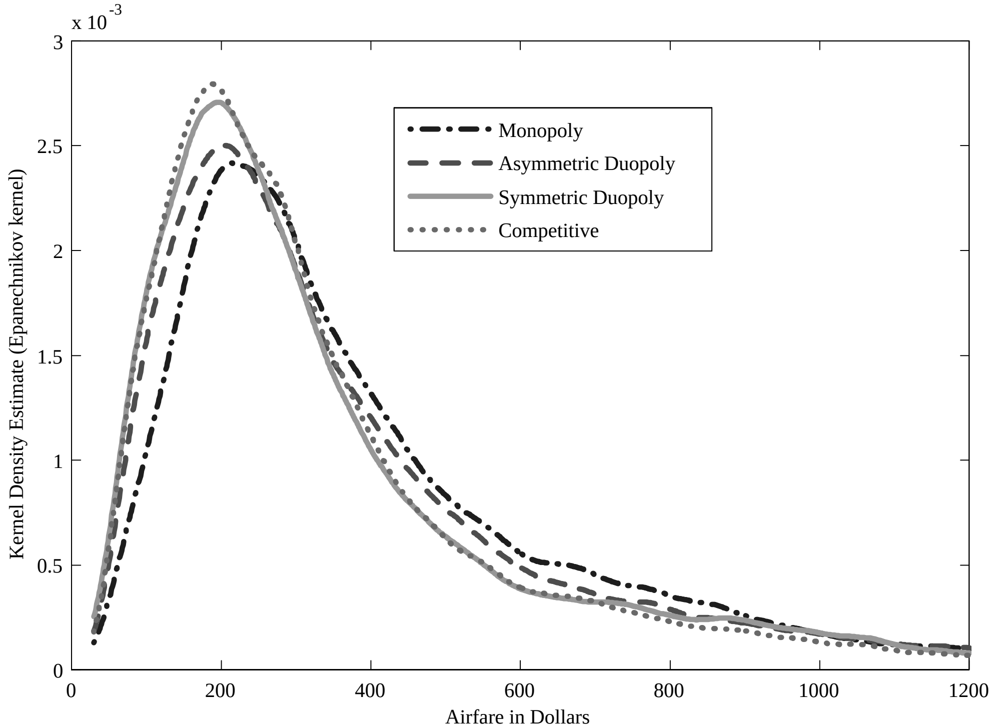

Figure 1: Empirical distribution of airline fares by market structure type. The empirical densities are smoothed with the Epanechnikov kernel based on all ticket observations associated with a given market structure type.

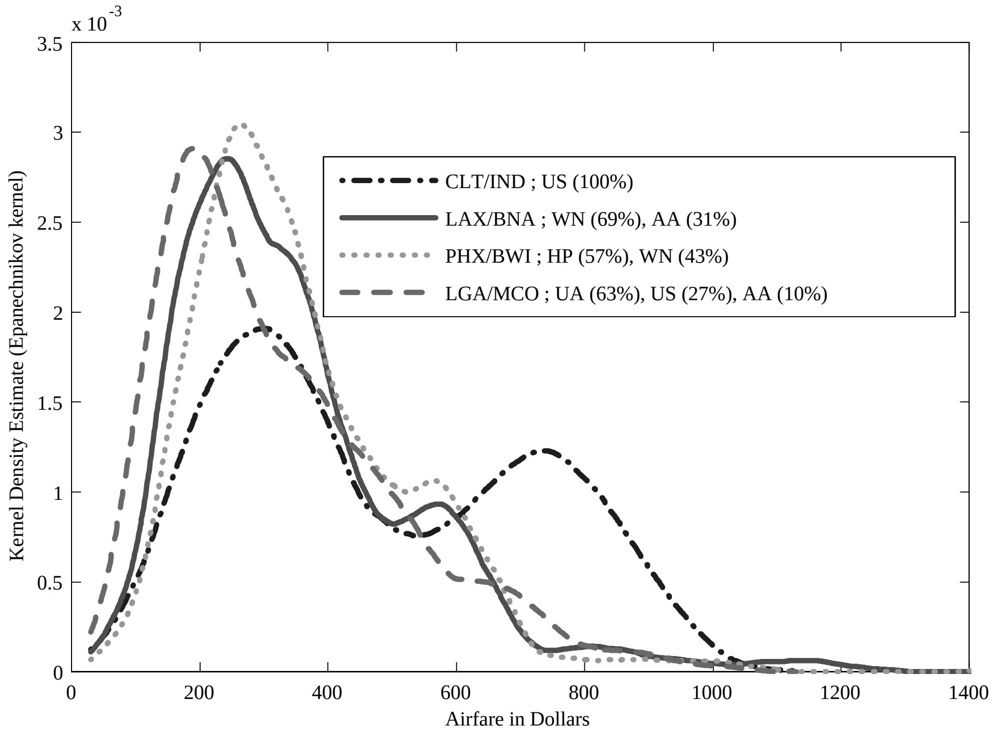

Figure 2: Empirical distribution of fares for a subset of routes. CLT/IND is a monopoly route between Charlotte, NC and Indianapolis, IN dominated by US Airways. LAX/BNA is an asymmetric duopoly route between Los Angeles, CA and Nashville, TN dominated by Southwest and American Airlines. PHX/BWI is a symmetric duopoly route between Phoenix, AZ and Baltimore, MD dominated by Amerian West and Southwest Airlines. LGA/MCO is a competitive route between New York's La Guardia airport and Orlando, FL, operated by United, US, and American Airlines.

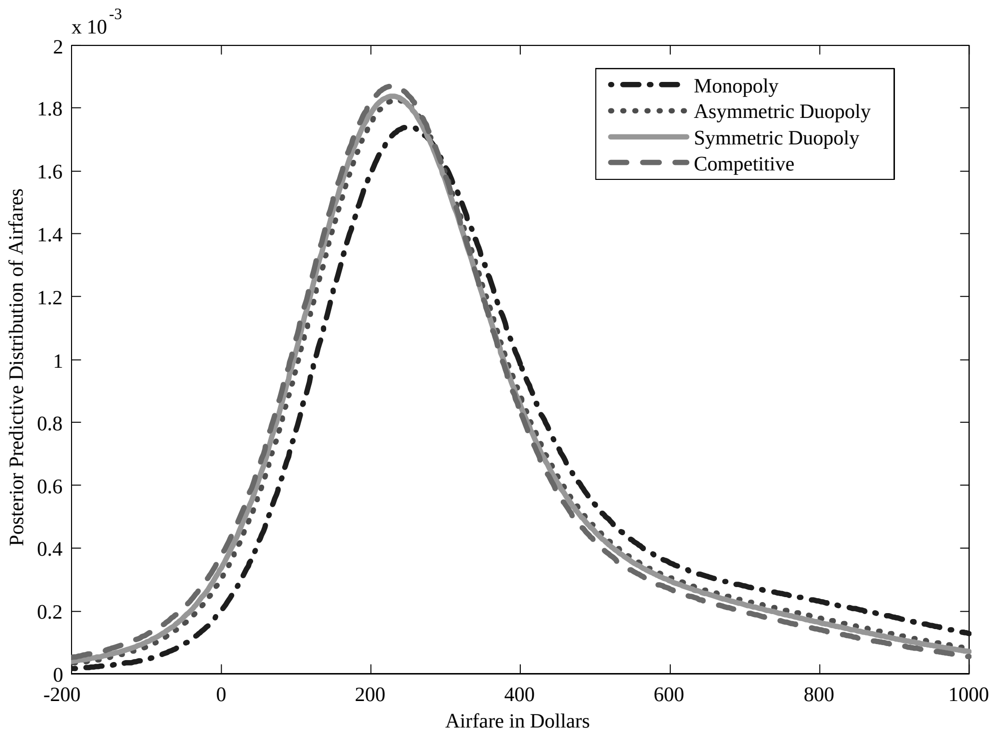

Figure 3: Posterior predictive distribution of airline fares for various market structures. Restricted ticket status and roundtrip status are integrated out based on their joint sample incidences. The number of passengers traveling on a ticket is fixed at 2, and the number of flights and distance travelled on the route are fixed at their median sample values. The number of seats offered is fixed at the median sample value for a monopoly, 0.7 and 0.3 times the median value for asymmetric monopolies, 0.5 times the median value for symmetric duopolies, and one-third the median value for competitive markets.

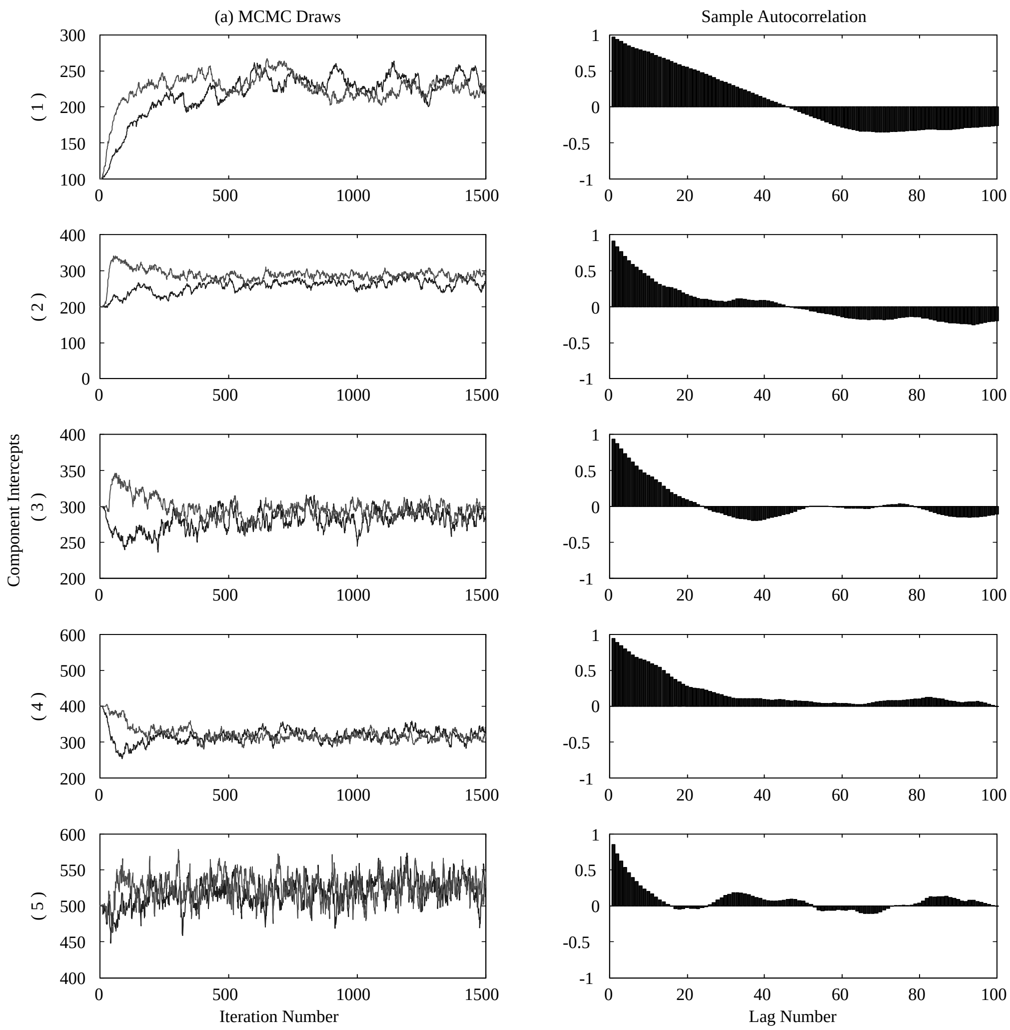

Figure 4: Plots of MCMC draws and autocorrelations for the coefficients on the route effect of Monopoly market structure. Each panel along the vertical axis corresponds to the coefficient in the indicated component of the mixture distribution.

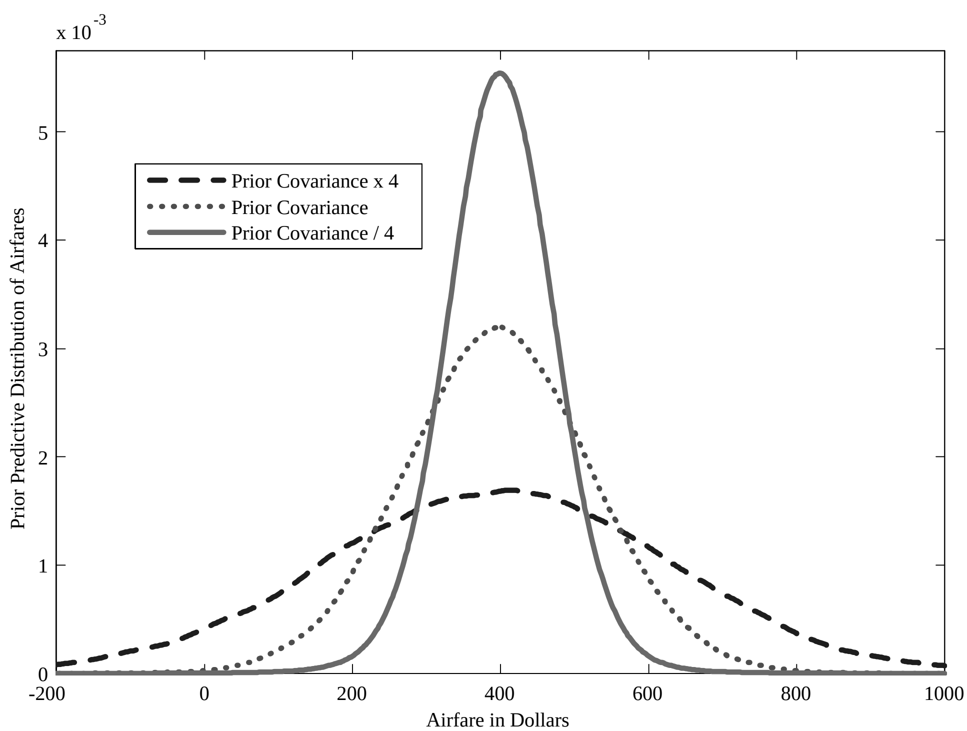

Figure 5: Sensitivity of the prior predictive distribution of airline fares to changes in the prior covariance of the model coefficients.

References

Albert, James H. and Siddhartha Chib, "Bayesian Analysis of Binary and Polychotomous Response Data" Journal of the American Statistical Association 88 (1993), 669-679.

Borenstein, Severin and Nancy L. Rose, "Competition and Price Dispersion in the U.S. Airline Industry" Journal of Political Economy 102 (1994), 653-683.

Busse, Meghan R. and Marc Rysman, "Competition and Price Discrimination in Yellow Page Advertising" RAND Journal of Econometrics (forthcoming).

Carlin, Bradley P. and Thomas A. Louis, "Bayes and Empirical Bayes Methods for Data Analysis" 2nd Ed., Chapman and Hall/CRC (2000).

Chib, Siddhartha, "Marginal Likelihood from the Gibbs Output" Journal of the American Statistical Association 90 (1995), 1313-1321.

Chib, Siddhartha and Bradley P. Carlin, "On MCMC Sampling in Hierarchical Longitudinal Models" Statistics and Computing 9 (1999), 17-26.

Dana, J. Jr., "Equilibrium Price Dispersion Under Demand Uncertainty: The Roles of Costly Capacity and Market Structure" RAND Journal of Econometrics 30 (1999), 632-660.

Hayes, Kathy J. and Leola B. Ross, "Is Airline Price Dispersion the Result of Careful Planning or Competitive Forces?" Review of Industrial Organization 13 (1998), 523-541.

Koop, Gary "Bayesian Econometrics" Wiley Interscience (2003).

Koop, Gary and Justin L. Tobias, "Learning About Heterogeneity in Returns to Schooling" Journal of Applied Econometrics 19 (2004), 827-849.

Laird, N.M. and J.H. Ware, "Random Effects Models for Longitudinal Data," Biometrics 38 (1982), 963-974.

Lindley, D.V. and A.F.M. Smith, "Bayes Estimates for the Linear Model (With Discussion)" Journal of the Royal Statistical Society B 34 (1972), 1-41.

Liu, Qihong and Konstantinos Serfes, "The Effect of Market Structure on Price Dispersion: An Analysis of the U.S. Airline Industry" mimeo, University of Massachusetts, Amherst and Drexel University (2005).

Poirier, Dale J., "Intermediate Statistics and Econometrics: A Comparative Approach" The MIT Press, Cambridge (1995).

Stavins, Joanna, "Price Discrimination in the Airline Market: The Effect of Market Concentration" Review of Economics and Statistics (2001), 200-202.

Stole, Lars A., "Nonlinear Pricing and Oligopoly" Journal of Economics & Management Strategy 4 (1995), 529-562.