Title

Estimating a mixed strategy employing maximum entropy

Authors
Golan, Amos
Karp, Larry
Perloff, Jeffrey M.

Publication Date
1996-02-01

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS
DIVISION OF AGRICULTURE AND NATURAL RESOURCES
UNIVERSITY OF CALIFORNIA AT BERKELEY

WORKING PAPER NO. 780

ESTIMATING A MIXED STRATEGY EMPLOYING MAXIMUM ENTROPY

by

Amos Golan

Larry S. Karp

and

Jeffrey M. Perloff

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
February, 1996

Estimating a Mixed Strategy Employing Maximum Entropy

Amos Golan
Larry S. Karp
Jeffrey M. Perloff

Abstract

Generalized maximum entropy may be used to estimate mixed strategies subject to restrictions from game theory. This method avoids distributional assumptions and is consistent and efficient. We use this method to estimate the mixed strategies of duopolistic airlines.

KEYWORDS: Mixed strategies, noncooperative games, oligopoly, maximum entropy, airlines
JEL: C13, C35, C72, L13, L93

Contact:
Jeffrey M. Perloff (510/642-9574; 510/643-8911 fax)
Department of Agricultural and Resource Economics
207 Giannini Hall
University of California
Berkeley, California 94720
perloff@are.Berkeley.Edu

George Judge was involved in every stage of this paper and was a major contributor, but is too modest to agree to be a coauthor. He should be. We are very grateful to Jim Brander and Anming Zhang for generously providing us with the data used in this study.

Table of Contents

Abstract

Generalized maximum entropy may be used to estimate mixed strategies subject to restrictions from game theory. This method avoids distributional assumptions and is consistent and efficient. We use this method to estimate the mixed strategies of duopolistic airlines.

KEYWORDS: Mixed strategies, noncooperative games, oligopoly, maximum entropy, airlines

JEL: C13, C35, C72, L13, L93

1. INTRODUCTION

We develop a method for estimating oligopoly strategies subject to restrictions implied by a game-theoretic model. Using this method, we estimate the pricing strategies of American and United Airlines.

Unlike most previous empirical applications, we do not assume that firms use a single pure strategy nor do we make the sort of ad hoc assumptions used in conjectural variations models.[1] Our method allows firms to use either pure or mixed strategies consistent with game theory.

First, we approximate a firm's continuous action space (such as price, quantity, or advertising) with a discrete grid. Then, we estimate the vector of probabilities — the mixed or pure strategies — that a firm chooses an action within each possible interval in the grid. We use these estimated strategies to calculate the Lerner index of market structure.

The main advantage of our method is that it can flexibly estimate firms' strategies subject to restrictions implied by game theory. The restrictions we impose are consistent with a variety of assumptions regarding the information that firms have when making their decisions. Firms may use different pure or mixed strategies in each state of nature. Firms may have private or common knowledge about the state of nature, which is unobserved by the econometrician. For example, a firm may observe a random variable that affects its marginal profit and know the distribution (but not the realization) of the random variable that affects its rival's marginal profit. Each firm may choose a pure strategy in every state of nature and regard its rival's action as a

[1] Breshnahan (1989) and Perloff (1992) survey conjectural variations and other structural and reduced-form "new empirical industrial organization" studies.

random variable. Alternatively, there may be no exogenous randomness, but the firm uses a mixed strategy. To the econometrician, who does not observe the firm's information or state of nature, the distribution of actions looks like the outcome of a mixed strategy in either case. The econometrician is not able to determine the true information structure of the game. Nevertheless, the equilibrium conditions for a variety of games have the same form, and by imposing these conditions we can estimate strategies that are consistent with theory.

There have been few previous studies that estimated strategies based on a game-theoretic model. All of the studies of which we are aware (Bjorn and Vuong 1985, Bresnahan and Reiss 1991, and Kooreman 1994) involve discrete games. For example, Kooreman estimates mixed strategies in a game involving spouses' joint labor market participation decisions using a maximum likelihood (ML) technique. Our approach differs from his in three important ways. First, Kooreman assumes that there is no exogenous uncertainty. Second, he allows each agent a choice of only two possible actions. Third, because he uses a ML approach, Kooreman assumes a specific error distribution and likelihood function. Despite the limited number of actions, his ML estimation problem is complex.

Our problem requires that we include a large number of possible actions so as to analyze oligopoly behavior and allow for mixed strategies. To do so using a ML approach would be extremely difficult. Instead, we use a generalized-maximum-entropy (GME) estimator. An important advantage of our GME estimator is its computational simplicity. With it, we can estimate a model with a large number of possible actions while imposing inequality and equality restrictions implied by the equilibrium conditions of the game. In addition to this practical advantage, the GME estimator does not require strong, arbitrary distributional assumptions, unlike

ML estimators. However, a special case of the GME estimator is identical to an ML estimator.

In the next section, we present a game-theoretic model of firms' behavior. In the third section, we describe a GME approach to estimating this game. The fourth section contains estimates of the strategies of United and American Airlines, and sampling experiments that illustrate the small sample properties of our GME estimator. In the final section, we discuss our results and possible extensions.

2. OLIGOPOLY GAME

Our objective is to determine the strategies of oligopolistic firms using time-series data on prices, quantities, and, when available, variables that condition the cost or demand relations. We assume that two firms, *i* and *j*, play a static game in each period of the sample. (The generalization to several firms is straightforward.)

Firm *i* (and possibly Firm *j*), but not the econometrician, observes the random variable $\varepsilon^i(t)$ in period *t*. For notational simplicity, we suppress the time variable *t*. The set of *K* possible realizations, $\{\varepsilon_1, \varepsilon_2, ..., \varepsilon_K\}$, is the same every period and for both firms. This assumption does not lead to a loss of generality because the distribution may be different for the two firms. The firms, but not the econometrician, know the distributions of ε_k. We consider three possible stochastic structures: (1) Firms face no exogenous randomness ($K = 1$); (2) ε_k is private information for Firm *i*; (3) ε_k is common-knowledge for the firms. Because the econometrician does not observe ε_k, even if the firms use a pure strategy in each period, it appears to the econometrician that they are using a mixed strategy whenever their actions vary over time.

2.1 Strategies

The set of n possible actions for either firm is $\{x_1, x_2, ..., x_n\}$. The assumption that the action space is the same for both firms entails no loss of generality because the profit functions can be specified so that certain actions are never chosen. The notation x^i_s means that Firm i chooses action x_s. We now describe the problem where the random state of nature is private information and then discuss alternative assumptions of a single state of nature or common information.

In determining its own strategy, Firm i forms a prior, β^i_{sk}, about the probability that Firm j will pick action x^j_s when i observes ε^i_k. If the firms' private information is correlated, it is reasonable for Firm i to base its beliefs about j's actions on ε^i_k. If the private information is uncorrelated, Firm i form priors that are independent of ε^i_k. We do not, however, assume independence. In state k, Firm i's strategy is $\underline{\alpha}_k = (\alpha^i_{k1}, \alpha^i_{k2}, ..., \alpha^i_{kn})$, where α^i_{ks} is the probability that Firm i chooses action x^i_s. If Firm i uses a pure strategy, α^i_{ks} is one for a particular s and zero otherwise.

The profit of Firm i is $\pi^i_{rsk} = \pi^i(x^j_r, x^i_s, \varepsilon^i_k)$, where r indexes the strategies of Firm j and s indexes the actions of Firm i. In state k, Firm i chooses $\underline{\alpha}_k$ to maximize expected profits, $\Sigma_r \beta^i_{rk}\pi_{rsk}$, where the expectation is taken over the rival's actions. If Y^i_k is Firm i's maximum expected profits when ε^i_k occurs, then $L^i_{sk} \equiv \Sigma_r \beta^i_{rk}\pi_{rsk} - Y^i_k$ is Firm i's expected loss of using action x^i_s in k. Because Y^i_k is the maximum possible expected profit, the expected loss when Firm i uses action s must be nonpositive,

(2.1) $$L^i_{sk} \leq 0.$$

For $\underline{\alpha}_k$ to be optimal, the product of the expected loss and the corresponding probability must equal zero:

(2.2)
$$L^i_{sk}\, \alpha^i_{sk} = 0 .$$

Equation 2.2 says that there is a positive probability that Firm *i* will use action *s* only if the expected profits when action *s* is used are equal to the maximum expected profit.

This problem may have more than one pure or mixed strategy. Our estimation method selects a particular pure or mixed strategy consistent with these restrictions and the data.

2.2 Econometric Implications

Our objective is to estimate the firms' strategies subject to the constraints implied by optimization, Equations 2.1 and 2.2. We cannot use these constraints directly, however, because they involve the unobserved random variables ε^i_k. By taking expectations, we eliminate these unobserved variables and obtain usable restrictions.

Using the expectations operator E_k, we define $\beta^i_r \equiv E_k\, \beta_{rk}$, $Y^i \equiv E_k\, Y^i_k$, $\alpha^i_s \equiv E_k\, \alpha^i_{sk}$, $\pi^i_{rs} \equiv E_k\, \pi^i_{rsk}$, and $E_k\, L^i_{sk} \equiv L^i_s$. If we define $\theta^i_{sk} \equiv L^i_{sk} - (\Sigma_r \beta^i_r \pi^i_{rs} - Y^i)$ and take expectations, then $E_k\, \theta^i_{sk} = \Sigma_r \operatorname{cov}(\beta^i_{rk}, \pi^i_{rsk}) \equiv \theta^i_s$. Thus, $L^i_s \equiv E_k\, L^i_{sk} = \Sigma_r\, \beta^i_r \pi^i_{rs} - Y^i + \theta^i_s$. Taking expectations with respect to *k* of Equation 2.1, we obtain

(2.3)
$$\sum_r \beta^i_r\, \pi^i_{rs} - Y^i + \theta^i_s \le 0 .$$

Taking expectations with respect to *k* of Equation 2.2, we find that

(2.4)
$$\left(\sum_r \beta^i_r\, \pi^i_{rs} - Y^i \right) \alpha^i_s + \delta^i_s = 0 ,$$

where $\delta^i_s \equiv \theta^i_s\alpha^i_s$ + cov(θ^i_{sk}, α^i_{sk}). We can estimate the observable (unconditional) strategy vectors $\underline{\alpha}^i$, i = 1, 2, subject to the conditions implied by Firm i's optimization problem, Equations 2.3 and 2.4.

For the general case of private information, we cannot determine the sign of θ^i_s and δ^i_s. However, if Firm i does not condition its beliefs about Firm j's actions on its own private information (as would be reasonable if the private information is uncorrelated), then β^i_{rk} is constant over k. Here, $\theta^i_s = 0$ and δ^i_s = cov(θ^i_{sk}, α^i_{sk}) = cov(L^i_{sk}, α^i_{sk}) ≥ 0. This last relation holds with strict inequality if and only if the number of states in which it is optimal for Firm i to use action x^i_s, with positive probability, is greater than 1 (so that $\alpha^i_s > 0$) and less than K (so that $L^i_s < 0$). If firms have no exogenous uncertainty but use mixed strategies, then $\theta^i_s = \delta^i_s = 0$. Thus, private, uncorrelated information implies $\theta^i_s = 0$ and $\delta^i_s \geq 0$, whereas the absence of exogenous uncertainty implies $\theta^i_s = 0$ and $\delta^i_s = 0$.

If the information that is unobserved by the econometrician is common knowledge to the firms, Firm i's beliefs and actions may be conditioned on the random variable ε^j_m that Firm j faces. If so, β^i_{rk} is replaced by β^i_{rkm}, and α^i_{sk} is replaced by α^i_{skm}, but restrictions 2.1 and 2.2 are otherwise unchanged. Taking expectations over k and m, we obtain restrictions of the same form as Equations 2.2 and 2.3. Again, in general we cannot sign θ^i_s and δ^i_s.

We have assumed that the econometrician observes the actions that firms choose, but not the information they use to condition these actions, ε^i_k. This assumption simplifies the estimation problem, because it means that the strategies, $\underline{\alpha}^i$, are numbers.[2]

We view 2.3 and 2.4 as stochastic restrictions that hold approximately due to an additive error in each equation. We already have additive parameters (θ^i_s and δ^i_s), so we are able to estimate the sum of those parameters and any additive error, but we cannot identify the two components. Thus, for notational simplicity, we call the sum of the systematic and random components θ^i_s and δ^i_s (rather than add new random variables). We also include an additive error, $\mu^i_s \in [-1, 1]$, associated with α^i_s. That is, we replace 2.4 with

$$\left(\sum_r \beta^i_r \pi^i_{rs} - Y^i\right)\left(\alpha^i_s + \mu^i_s\right) + \delta^i_s = 0 \,. \tag{2.5}$$

We have an analogous set of restrictions for Firm *j*.

The Nash assumption is that agents' beliefs about their rival's actions are correct so that

$$\beta^i_r = \alpha^j_r \,, \tag{2.6}$$

for $i \neq j$. We henceforth maintain the Nash assumption.

If we tried to estimate this model — Equations 2.3, 2.5 and 2.6 — using traditional techniques, we would run into several problems. First, with conventional sampling theory estimation techniques, we would have to specify arbitrarily an error distribution. Second,

---

[2] If the firms' strategies are conditioned on a variable ζ that the econometrician observes, the econometrician may need to estimate functions $\underline{\alpha}^i(\zeta)$ rather than numbers, $\underline{\alpha}^i$. Suppose, however, that Firm *i*'s profits can be written as $f^i(\zeta)\pi^i_{rsk}$, where f^i is a positive function. For example, Firm *i* chooses price p^i, and faces demand $D^i(p^i, p^j)f^i(\zeta)$. Given this multiplicative form, $f^i(\zeta)$ merely rescales the restrictions 2.3 and 2.4 [we can divide each restriction by $f^i(\zeta)$], and the equilibrium strategies, α^i, are independent of ζ. Throughout the rest of this paper, we assume that any variables such as ζ enter the profit functions multiplicatively so that restrictions 2.3 and 2.4 are correct. We will discuss the more general problem in a future paper.

imposing the various equality and inequality restrictions from our game-theoretic model would be very difficult if not impossible with standard techniques. Third, as the problem is ill posed in small samples (there are more parameters than observations), we would have to impose additional assumptions to make the problem well posed. To avoid these and other estimation and inference problems, we propose an alternative approach.

3. GENERALIZED-MAXIMUM-ENTROPY ESTIMATION APPROACH

We use generalized maximum entropy (GME) to estimate the firms' strategies. In this section, we start by briefly describing the traditional maximum entropy (ME) estimation procedure. Then, we present the GME formulation as a method of recovering information from the data consistent with our game. This GME method is closely related to the GME multinomial choice approach in Golan, Judge, and Perloff (1996). Unlike ML estimators, the GME approach does not require explicit distributional assumptions, performs well with small samples, and can incorporate inequality restrictions.

3.1 Background: Classical Maximum Entropy Formulation

The traditional entropy formulation is described in Shannon (1948), Jaynes (1957a; 1957b), Kullback (1959), Levine (1980), Jaynes (1984), Shore and Johnson (1980), Skilling (1989), Csiszár (1991), and Golan, Judge, and Miller (1996). In this approach, Shannon's (1948) entropy is used to measure the uncertainty (state of knowledge) we have about the occurrence of a collection of events. Letting x be a random variable with possible outcomes x_s, $s = 1, 2, \ldots, n$, with probabilities α_s such that $\Sigma_s\ \alpha_s = 1$, Shannon (1948) defined the *entropy* of the distribution $\underline{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_n)'$, as

(3.1)
$$H \equiv -\sum_s \alpha_s \ln \alpha_s ,$$

where $0 \ln 0 \equiv 0$. The function H, which Shannon interprets as a measure of the uncertainty in the mind of someone about to receive a message, reaches a maximum when $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 1/n$. To recover the unknown probabilities $\underline{\alpha}$, Jaynes (1957a; 1957b) proposed maximizing entropy, subject to available data consistency relations, such as moments from the observed data, and adding up constraints.

To use this approach for our game problem, we need to incorporate the data from our sample. Let n_s^i be the number of times x_s^i is observed, out of T total observations. The observed frequency in the sample is $n_s^* \equiv n_s/T$. [We henceforth suppress the firm superscript for notational simplicity whenever possible.] For each firm, the observed frequency equals the true strategy probability, α_s, plus an error term:

(3.2)
$$n_s^* \equiv \frac{n_s}{T} = \alpha_s + e_s ,$$

where the noise term $e_s \in [-1, 1]$.

The traditional ME approach sets e_s in Equation (3.2) equal to zero,

(3.3)
$$n_s^* = \alpha_s$$

and maximizes the Shannon measure (3.1) subject to Equation 3.3. The solution to this problem is trivial in the sense that the constraint 3.3 completely determines the parameter estimate. This ME estimator is identical to the ML estimator, when the x's have a multinomial distribution.

3.2 The Basic Generalized Maximum Entropy Formulation

The GME formulation, which uses restriction 3.2, is a more general version of the ME formulation, which uses restriction 3.3. We obtain the basic GME estimator by maximizing the sum of the entropy corresponding to the strategy probabilities, α, and the entropy from the noise, e, in consistency condition 3.2 subject to that data consistency condition.

In general, the GME objective is a dual-criterion function that depends on a weighted sum of the entropy from both the unknown and unobservable $\underline{\alpha}$ and $\underline{e} = (e_1, e_2, ..., e)'$. By varying the weights, we can put more weight on estimation (accuracy of the α coefficients) or prediction (assignment of observations to a category). The ME estimator is a special case of the GME, in which no weight is placed on the noise component, so that the estimation objective is maximized (thus maximizing the likelihood function). As a practical matter, our GME objective weights the $\underline{\alpha}$ and $\underline{e}$ entropies equally because we lack any theory that suggests other weights.

The arguments of the entropy measures must be probabilities. The elements of $\underline{\alpha}$ are probabilities, but the elements of $\underline{e}$ range over the interval [-1, 1]. To determine the entropy of $\underline{e}$, we reparameterize its elements using probabilities. We start by choosing a set of discrete points, called the support space, $\underline{v} = [v_1, v_2, ..., v_M]'$ of dimension $M \geq 2$, that are at uniform intervals, symmetric around zero, and span the interval [-1, 1]. Each error point e_s has corresponding unknown weights $\underline{w}_s = [w_{s1}, w_{s2}, ..., w_{sM}]'$ that have the properties of probabilities: $0 \leq w_{sm} \leq 1$ and $\Sigma_m w_{sm} = 1$. We reparameterize each error element as

$$e_s = \sum_m v_m w_{sm} .$$

For example, if $M = 3$, then $\underline{v} = (-1, 0, 1)'$, and there exists w_1, w_2, and w_3 such that each noise component can be written as $e_s = w_1(-1) + w_3(1)$. Given this reparameterization, we can rewrite the GME consistency conditions, Equation 3.2, as

$$\underline{n}* = \underline{\alpha} + \underline{e} = \underline{\alpha} + W\underline{v}, \tag{3.4}$$

where row s of the matrix W is the vector of probabilities $\underline{w}_s$, and $\underline{v}$, the support space, is the same for all s.

No subjective information on the distribution of probabilities is assumed. It is sufficient to have two points ($M = 2$) in the support of $\underline{v}$, which converts the errors from [-1, 1] into [0, 1] space. This estimation process recovers M - 1 moments of the distribution of unknown errors, so a larger M permits the estimation of more moments. Monte-Carlo experiments show a substantial decrease in the mean-square-error (MSE) of estimates when M increases from 2 to 3. Further increases in M provides smaller incremental improvement. The estimates hardly change if M is increased beyond 7 (Golan, Judge, Perloff, 1996; Golan, Judge, Miller, 1996).

If we assume that the actions, $\underline{x}$, and the errors, $\underline{e}$, are independent and define $\underline{w} \equiv \text{vec}(W)$, the GME problem for each firm is

$$\max_{\underline{\alpha}, \underline{w}} H\left(\underline{\alpha}, \underline{w}\right) = -\underline{\alpha}' \ln \underline{\alpha} - \underline{w}' \ln \underline{w}, \tag{3.5}$$

subject to the GME consistency conditions, Equation 3.4, and the normalization constraints

$$\underline{1}'\underline{\alpha} = 1 \qquad \underline{1}'\underline{w}_s = 1 \tag{3.6a,b}$$

for $s = 1, 2, \ldots, n$.

The Lagrangean to the GME problem is

$$L(\underline{\lambda}, \rho, \underline{\eta}) = -\sum_s \alpha_s \ln \alpha_s - \sum_s \sum_m w_{sm} \ln w_{sm} + \sum_s \lambda_s \left(n_s^* - \alpha_s - \underline{v}' \underline{w}_s\right) + \rho(1 - \underline{1}'\underline{\alpha}) + \sum_s \eta_s (1 - \underline{1}' \underline{w}_s). \tag{3.7}$$

where $\underline{\lambda}$, ρ, and $\underline{\eta}$ are Lagrange multipliers. Solving this problem, we obtain the GME estimators

$$\breve{\alpha}_s = \frac{\exp\left(-\breve{\lambda}_s\right)}{\sum_j \exp\left(-\breve{\lambda}_j\right)} \equiv \frac{\exp\left(-\breve{\lambda}_s\right)}{\Omega\left(\underline{\breve{\lambda}}\right)}, \tag{3.8}$$

$$\breve{w}_{sm} = \frac{\exp\left(-\breve{\lambda}_s v_m\right)}{\sum_m \exp\left(-\breve{\lambda}_s v_m\right)} \equiv \frac{\exp\left(-\breve{\lambda}_s v_m\right)}{\Psi_s\left(\underline{\breve{\lambda}}\right)}, \tag{3.9}$$

and

$$\underline{\breve{e}} = \breve{W} \underline{v}. \tag{3.10}$$

The Hessian is negative definite (the first n elements on the diagonal are $-1/\alpha_s$, the rest are $-1/w_{sm}$, and the off-diagonal elements are 0) so the solution is globally unique.

Following Agmon et. al (1979), Miller (1994), and Golan et al. (1996), we can reformulate the GME problem as a generalized-likelihood function, which includes the traditional likelihood as a special case:

$$
\begin{aligned}
L(\underline{\lambda}) &= -\sum_s \alpha_s(\underline{\lambda}) \ln \alpha_s(\underline{\lambda}) - \sum_s \sum_m w_{sm}(\underline{\lambda}) \ln w_{sm}(\underline{\lambda}) \\
&\quad + \sum_s \lambda_s \left(n_s^* - \alpha_s - \underline{v}' \underline{w}_s\right) \\
&= -\sum_s \alpha_s(\underline{\lambda}) \left[-\lambda_s - \ln \Omega(\underline{\lambda})\right] - \sum_s \sum_m w_{sm}(\underline{\lambda}) \left[-\lambda_s v_m - \ln \Psi_s(\underline{\lambda})\right] \\
&\quad + \sum_s \lambda_s \left(n_s^* - \alpha_s - \underline{v}' \underline{w}_s\right) \\
&= \sum_s \lambda_s n_s^* + \ln \Omega(\underline{\lambda}) + \sum_s \ln \left[\Psi_s(\underline{\lambda})\right].
\end{aligned}
\tag{3.11}
$$

Minimizing Equation 3.11 with respect to $\underline{\lambda}$ — setting the gradient, $\Delta L(\underline{\lambda}) = \underline{n}^* - \underline{\alpha} - \underline{e}$, equal to zero — yields the same estimates as from the original formulation, Equation 3.8. One advantage of this dual formulation, Equation 3.11, is that it is computationally more efficient.

3.4 Generalized Maximum Entropy Formulation of the Nash Model

We can also use the GME approach to estimate the strategies subject to the game-theoretic restrictions. Here, we require the estimates to satisfy the optimality conditions, Equations 2.3 and 2.5, and the Nash condition, Equation 2.6.[3] Thus, our objective is to recover the strategies, $\underline{\alpha}$, for each firm given the T observations and our knowledge of the economic generating process. We first assume that the econometrician knows the parameters of the functional form of π^i_{rs}, and we later discuss how the problem is changed when some parameters,

[3] Equations 2.3 and 2.4 are not standard econometric restrictions, as each includes an additional unknown parameter: θ^i_s in Equation 2.3 and δ^i_s in Equation 2.4. Therefore, it might appear that the added degree of freedom caused by the new parameter cancels the added information in the restriction. However, when these restrictions are imposed, the parameters θ^i_s and δ^i_s appear in the criterion function. As a result, imposing these restrictions causes the estimates of all the parameters to change and improves the estimates, as we show below.

such as the demand coefficients, must be estimated.

Equation 2.3 includes the noise components $\underline{\theta}^i$, and Equation 2.5 includes the noise components $\underline{\mu}^i$ and $\underline{\delta}^i$, i = 1, 2. Our first step is to reparameterize these six vectors in terms of probabilities. Let $\underline{v}^d$ be a vector of dimension $J^d \geq 2$ with corresponding unknown weights $\underline{\omega}_k^d$ such that

(3.12) $$\sum_j \omega_{sj}^d = 1,$$

(3.13) $$\underline{v}^{d\prime}\underline{\omega}^d = d,$$

for $d = \underline{\mu}^i$, $\underline{\theta}^i$, and $\underline{\delta}^i$, i = 1, 2. The support spaces $\underline{v}^d$ are defined to be symmetric around zero for all d. The natural boundaries for the errors $\underline{\mu}^i$ and $\underline{\mu}^j$ are [-1, 1]. We do not have natural boundaries for $\underline{\theta}^i$ or $\underline{\delta}^i$, so we use the "three-sigma rule" (Pukelsheim, 1994; Miller 1994; Golan, Judge, and Miller 1996) to choose the limits of these support spaces, where sigma is the empirical standard deviation of the discrete action space of prices or quantities.

To simplify the notation, let $\underline{n}^* = (\underline{n}^{*i\prime}, \underline{n}^{*j\prime})'$, $\underline{\alpha} = (\underline{\alpha}^{i\prime}, \underline{\alpha}^{j\prime})'$, $\underline{w} = (\underline{w}^{i\prime}, \underline{w}^{j\prime})'$, and $\underline{\omega} = (\underline{\omega}^{\mu i\prime}, \underline{\omega}^{\mu j\prime}, \underline{\omega}^{\theta i\prime}, \underline{\omega}^{\theta j\prime}, \underline{\omega}^{\delta i\prime}, \underline{\omega}^{\delta j\prime})'$. As above, we assume independence between the actions and the errors. The GME problem is

(3.14) $$\underset{\underline{\alpha}, \underline{w}, \underline{\omega}}{Max}\ H(\underline{\alpha}, \underline{w}, \underline{\omega}) = -\underline{\alpha}' \ln \underline{\alpha} - \underline{w}' \ln \underline{w} - \underline{\omega}' \ln \underline{\omega}$$

subject to the data consistency conditions 3.4, the necessary economic conditions 2.3 and 2.5, the Nash condition 2.6, and the normalizations for $\underline{\alpha}$, $\underline{w}$, and $\underline{\omega}$. The errors θ_s^i, μ_s^i, and δ_s^i in 2.3 and 2.5 are defined by Equations 3.12 and 3.13. Solving this problem yields estimates $\underline{\tilde{\alpha}}$, $\underline{\tilde{w}}$, and $\underline{\tilde{\omega}}$.

If we do not know the parameters of the profit or demand functions, we simultaneously estimate the profit or demand parameters and the strategy parameters using GME estimation procedures. To do so, we need to modify the objective function 3.14 and add the profit (or demand) functions for each firm as additional constraints in the GME-Nash model. This model, for the unknown profit/demand parameters, is described in Appendix 1.

3.5 Properties of the Estimators and Normalized Entropy

All three different estimators, the ME-ML, GME, and GME-Nash, are consistent, but they differ in efficiency and information content. The ML estimator is known to be consistent. Because the ME and ML estimators are identical, as we noted above, the ME estimator is also consistent. Under the assumption that a solution to the GME-Nash estimation problem exists for all samples and given an appropriate choice of the bounds of the error in the data consistency constraint 3.2, we show in Appendix 2 that the GME and GME-Nash estimators are also consistent.

The GME estimator of α has smaller variance than the ME-ML estimator (Golan, Judge, and Perloff 1996). Given that the game-theoretic constraints are correct, the possible solution space for GME-Nash estimate of α is a subset of the solution space of the GME estimate of α. Thus, we conjecture that the GME-Nash estimator has a smaller variance than the GME. In the sampling experiments reported below, this conjecture is always confirmed.

We can compare the different estimators empirically using the normalized entropy (information) measure $S(\underline{\alpha}) = -(\Sigma_s \alpha_s \ln \alpha_s)/(\ln n)$, which measures the extent of uncertainty about the unknown parameters. If there is no uncertainty, $S(\underline{\alpha}) = 0$. If there is full ignorance, in the sense that all actions are equally likely, $S(\underline{\alpha}) = 1$. All else the same, additional information reduces the uncertainty in the data analyzed, resulting in a lower normalized entropy measure. Thus, to the degree that the constraints 2.3, 2.5, and 2.6 bind, the GME-Nash normalized entropy measure is lower than is the GME measure: $S(\underset{\sim}{\tilde{\alpha}}) \leq S(\underset{\sim}{\breve{\alpha}})$.

4. AIRLINES

We estimated the strategic behavior of American and United Airlines using the ME-ML, GME, and GME-Nash approaches. We assume that the airlines set price. We allow for the possibility that American and United provide differentiated services on a given route and assume that the demand curve facing Firm i is

$$q_i = a_i + b_i p_i + d_i p_j + u_i , \tag{4.1}$$

where a_i and d_i are positive, b_i is negative, and u_i is an error term.[4] In Appendix 1, we show how to reparametrize 4.1 so that it can be estimated along with the other parameters in the GME-Nash model.

If firms choose prices, the necessary conditions 2.3 and 2.5 become

$$\sum_r \beta_r^i \left(p_s^{\,i} - c^{\,i}\right) q_{rs}^{\,i} - Y^{\,i} + \theta_s^i \leq 0, \tag{4.2}$$

$$\left[\sum_r \beta_r^i \left(p_s^{\,i} - c^{\,i}\right) q_{rs}^{\,i} - Y^{\,i}\right](\alpha_s^i + \mu_s^i) + \delta_s^i = 0 . \tag{4.3}$$

The Nash condition 2.6 is unchanged.

The data include price and quantity and cost data for 15 quarters (1984:4-1987:4, 1988:2,

[4] We experimented with including various additional right-hand-side variables such as measures of income, population, or economic conditions. None of these variables, however, significantly affected the fit of the equation or the parameters a_i, b_i, and d_i.

1988:4) for various routes between Chicago and other cities.[5] We calculated marginal costs using the formula in Oum, Zhang, and Zhang (1993), and we used the average of these for the parameters c^i. The nominal data are deflated using the Consumer Price Index.

On each of these routes, these two firms had no (or trivial) competition from other firms. We restrict our attention to two city pairs: Chicago-Providence and Chicago-Wichita. We did not estimate our model for the other city pairs in the data set because of two problems. For the Chicago-Las Vegas and Chicago-Sacramento routes, the average marginal cost was higher than the average observed price, hence we were unwilling to make the assumption that the firms were engaged in single-period maximizing behavior. On the basis of economic theory, we require that $b_i < 0$ (demand curves slope down) and $d_i > 0$ (the services are substitutes) for each demand curve. For the remaining routes, the demand curves estimated using ordinary least squares violated these properties.

4.1 The Airline Model Specification

To determine the price space for each city, we first specify the upper and lower bound of the price space. The lower bound is the smallest observed price for both airlines minus 10% and the upper bound is the largest observed price for both airlines plus 10%. We then divide the price space into 20 equal increments.

Because we do not know the true demand curve parameters, we simultaneously estimate linear demand curves for each firm and a price-strategic choice model. As we have a measure

[5] The data were generously provided by James A. Brander and Anming Zhang. They used these data in three excellent papers: Brander and Zhang (1990, 1993) and Oum, Zhang, and Zhang (1993). See these papers for a description of the data.

of the average marginal costs, we did not have to estimate c^i.

4.2 Airline Estimates

In the GME-Nash model, the correlation between the actual and estimated quantities in the demand equations are 0.1 for the American demand equation and 0.2 for the United equation for Providence and 0.5 for both Wichita equations. For Providence, the demand coefficients (a_i, b_i, and d_i) are 1,865.8, -12.1, 4.7 for American Airlines and 1,571.7, -10.2, 4.8 for United Airlines. Given these parameters, the own-demand elasticities at the sample mean are -2.6 for American and -2.1 for United, and both the cross-elasticities are 0.95. In Wichita, the demand coefficient are 694.0, -3.8, and 2.4 for American and 668.7, -4.1, and 2.8 for United. The own-demand elasticities are -1.6 for America and -1.4 for United, and the cross-price elasticities are 0.82 in the American equation and 0.92 in the United equation. Estimating these demand curves using ordinary least squares yields estimates of the same sign and magnitude.

On the Providence route, the estimated strategy parameters, $\underline{\alpha}$, for American Airlines are shown in Figure 1a and for United are shown in Figure 1b. The corresponding distributions for Wichita are shown in Figures 2a and 2b. The ME-ML estimates are the observed frequencies.

The GME distribution is more uniform than that of the ME-ML model because the GME consistency conditions 3.2 allow the estimates to differ from the actual frequency. In attempting to maximize entropy, the GME estimator pushes the probability estimates toward uniformity.

The GME-Nash distribution is smoother than the other two models and has one peak for American and two peaks for United in both cities. The global maximum of the GME-Nash distribution is closer to the average price based on a standard Bertrand model than to a Cournot or Collusive model. [The Cournot — $227 for American and $223 for United — and collusive

— \$286 and \$274 — means are too large to appear in Figure 2 for Wichita.]

Based on Kolmogorov-Smirnov tests, we cannot distinguish between the ME-ML, GME, and GME-Nash distributions. The normalized entropy measure, $S(\underline{\alpha})$, for Providence are 0.66 for American and 0.67 for United for the ME-ML model. The corresponding normalized entropy measures are both 0.90 for the GME model and 0.61 and 0.65 for the GME-Nash model. The normalized entropy measures for Wichita are, respectively, 0.73 and 0.77 for the ME-ML, 0.93 and 0.94 for the GME, and 0.70 and 0.68 GME-Nash. The drop in the entropy measure when we switch from the GME to the GME-Nash shows that the theoretical restrictions contain substantial information.

The estimated expected rents, $\tilde{Y}$, are \$420,000 for American and \$435,000 for United on the Providence route, and \$500,000 for each airline on the Wichita route. These rent calculations are based on the assumption that the average cost equals the marginal cost. These numbers do not include fixed costs. Unless the fixed costs are large, these number suggest that the airlines were making positive profits during this period. The estimated expected rents are consistent with the magnitudes of the prices and quantities observed.

For both airlines for both cities, the average value of $\underline{\tilde{\theta}}$ is practically zero. The average value of $\underline{\tilde{\delta}}$ is positive. For example, in Providence, only 2 out of the 40 values of $\underline{\tilde{\delta}}$ were negative. This sign pattern is consistent with firms having private, uncorrelated information. This pattern is inconsistent with the hypothesis that firms use mixed strategies despite the absence of exogenous randomness.

4.3 Comparing Estimators

How does our approach compare to traditional methods?[6] For the purposes of comparison, we estimated a traditional conjectural variations (CV) model given our heterogenous demand equations. The CV model consists of four equations: the two demand curves and two optimality (first-order) conditions.[7]

Figures 3a and 3b show how the conjectural variations distribution compares to the GME-Nash and ME-ML for Providence. The CV distribution has multiple peaks, with its global maximum slightly higher than the GME-Nash. The CV distribution is significantly different than the ME-ML for United on the Chicago-Providence route based on a Kolmogorov-Smirnov test. Similarly, for United on the Chicago-Wichita route, the CV distribution differs from the GME-Nash strategy distribution.

The estimated market power of these firms differs across the estimators. Table 1 shows how the average Lerner Index of market power (the difference between price and marginal cost

---

[6] We cannot directly compare our results to those in Brander and Zhang (1990, 1993) and Oum, Zhang, and Zhang (1993), because they assume that the services of the two airlines are homogeneous, whereas we estimate demand curves based on differentiated services. Moreover, two of their papers estimate pure strategy models, where we permit mixed or pure strategies. The other paper, Brander and Zhang (1993), estimates a supergame (trigger price) model. If there are punishment periods during the sample, our estimates may show two or more peaks in the distribution of $\underline{\alpha}$. If, however, the firms are using such supergames, we should modify our repeated single-period game model accordingly.

[7] When we tried to estimate the four equations simultaneously, some of the demand parameters took on theoretically incorrect signs. Consequently, we estimated the demand curves and then estimated the optimality conditions treating the estimated demand parameters as exact. Both methods produced similar estimates of the conjectures. Figures 3a and 3b use the second set of estimates where we used the marginal cost in each period to generate a distribution of estimates.

as a percentage of price) varies across the estimators. The ME-ML Lerner Index is identical to the index based on the observed data. The GME indexes are virtually the same or slightly lower than the ME-ML indexes. The average GME-Nash and CV estimates are virtually identical and slightly higher than the sample-based index.

Using the demand parameters from the GME-Nash model, we also calculated the average Bertrand, Cournot, and collusive Lerner Indexes. The average Bertrand index is virtually the same as the average GME-Nash and CV indexes. The Cournot and collusive indexes are much higher.

4.4 Sample Size Sensitivity Experiments

The squared-error loss of each of our three estimators differs as sample size changes. We can demonstrate these properties using sensitivity experiments, where we assume that the estimated demand equations for the Chicago-Wichita route hold with an error term that is distributed N(0, 1). We assume that Firm *i* has information ε^i about its marginal cost and that this information is private and uncorrelated (as is consistent with our estimates), so that Firm *i*'s beliefs, β^i_r, do not depend on ε^i_k. The marginal cost for each firm in each period is drawn from a normal distribution N(60, 5), which closely approximates the distribution of marginal costs for Wichita. We approximate this continuous distribution using a finite grid and use the probabilities associated with the resulting discrete distribution, ρ_k, to determine the Nash restriction that beliefs are correct in equilibrium. This restriction requires that $\Sigma_k \alpha^j_{rk}\rho_k = \beta^i_r$. We then generate Nash equilibrium strategies $\underline{\alpha}$ using this restriction and the necessary conditions 2.1 and 2.2. (We establish by means of sensitivity studies that this equilibrium is unique.) We use the resulting equilibrium probabilities $\underline{\alpha}$ to generate samples of actions by drawing a uniform random number on the unit interval and using that to assign an action for each observation. We generated 200 samples for T (the number of observations in each sample) = 10, 20 and 40, with n (the number of possible actions for each firm) = 20.

According to our analytic results, the GME estimator has lower variance than the ME-ML. We conjecture that the GME-Nash has a lower variance than the GME. These superior finite sample properties of the GME-Nash and GME over the ME-ML are confirmed by our sampling experiments in terms of the empirical mean square error (MSE($\underline{\tilde{\alpha}}^i$) = $\Sigma_{s,t}$ $(\tilde{\alpha}^i_{st} - \alpha^i_s)^2/200$ (where the index t denotes the sample) and the correlation coefficient between the estimated and true α^i_s for each of the models (Table 2). The table shows two sets of results depending on whether the econometrician knows the demand coefficients or has to estimate them. In the latter case, we generate quantities demanded by adding a N(0, 1) term to the demand equation.

The ME-ML and GME perform better (in terms of MSE and correlations) as the number of observations increases. The GME-Nash, however, performs well (relative to the other estimators) for a small number of observations, and the GME-Nash estimates do not improve as the number of observations increases beyond 20. This latter result is very attractive if, as usual, one has relatively few time-series observations. Finally, the GME-Nash estimator yields superior estimates even when the demand coefficients are unknown, without assuming knowledge of the error distributions.

5. CONCLUSIONS

Our generalized-maximum-entropy-Nash (GME-Nash) estimator can estimate firms' strategies consistent with game theory and the underlying data generation process. It is free of parametric assumptions about distributions and ad hoc specifications such as those used in conjectural-variations models.

Our simplest approach to estimating strategies is to use the maximum-likelihood (ML) or maximum-entropy (ME) estimators. These approaches produce the same estimates, which are the observed frequencies in the data. These estimators do not make use of demand or cost information and do not impose restrictions based on theory.

We also estimate two GME models. The basic GME estimator allows greater flexibility than the ML-ME estimator, but does not use demand, cost, or game-theoretic information. We show analytically and through simulations that this GME estimator is more efficient than the ME-ML estimator in terms of mean-square error, correlation, and other measures of variance. The GME-Nash estimator uses all available data, and game-theoretic information. In our sampling experiments, the GME-Nash estimator is more efficient than the basic GME and ME-ML estimators.

In future papers, we plan three generalizations of our approach. First, we will examine whether a price-choice or quantity-choice model is appropriate. Here, we have assumed that the firms chose price. Second, we will estimate more complex games where firms choose price or quantities simultaneously with advertising. Third, we will generalize the model so that strategies may vary with variables the econometrician observes.

We believe that this approach to estimating games can be applied to many problems in addition to oligopoly, such as wars and joint decisions by husbands and wives. To do so only requires replacing profits with some other criterion.

References

Agmon, Noam, Y. Alhassid, and Rafi D. Levine (1979): "An Algorithm for Finding the Distribution of Maximal Entropy," *Journal of Computational Physics*, 30, 250-9.

Bjorn, P. A. and Q. H. Vuong (1985): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force participation," California Institute of Technology working paper 537.

Brander, James A., and Anming Zhang (1990): "Market Conduct in the Airline Industry: An Empirical Investigation," *Rand Journal of Economics*, 21, 567-83.

Brander, James A., and Anming Zhang (1993): "Dynamic Oligopoly in the Airline Industry," *International Journal of Industrial Organization*, 11, 407-35.

Bresnahan, Timothy F. (1989): "Studies of Industries with Market Power," in Richard Schmalensee and Robert Willig, eds., *Handbook of Industrial Organization*, New York: North Holland, 1989.

Bresnahan, Timothy F., and Peter C. Reiss (1991): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48, 57-81.

Csiszár, I. (1991): "Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems," *The Annals of Statistics*, 19, 2032-2066.

Davidson, R., and J. G. MacKinnon (1993): *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Golan, Amos, George Judge, and Jeffrey M. Perloff (1996): "Recovering Information from Multinomial Response Data," *Journal of the American Statistical Association*, forthcoming.

Golan, Amos, George Judge, and Douglas J. Miller (1996): *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, New York: John Wiley & Sons, forthcoming.

Jaynes, E. T. (1957a): "Information Theory and Statistical Mechanics," *Physics Review*, 106, 620-630.

Jaynes, E. T. (1957b): "Information Theory and Statistical Mechanics, II," *Physics Review*, 108, 171-190.

Kooreman, Peter (1994): "Estimation of Econometric Models of Some Discrete Games," *Journal of Applied Econometrics*, 9, 255 - 68.

Kullback, J. (1959): *Information Theory and Statistics*, New York: John Wiley & Sons.

Levine, Rafi D. (1980): "An Information Theoretical Approach to Inversion Problems," *Journal of Physics*, A, 13, 91-108.

Miller, Douglas J. (1994): *Entropy and Information Recovery in Linear Economic Models*, Ph.D. thesis, University of California, Berkeley.

Oum, Tae Hoon, Anming Zhang, Yimin Zhang (1993): "Inter-firm Rivalry and Firm Specific Price Elasticities in Deregulated Airline Markets," *Journal of Transport Economics and Policy*, 27, 171-92.

Perloff, Jeffrey M. (1992): "Econometric Analysis of Imperfect Competition and Implications for Trade Research," in Ian M. Sheldon and Dennis R. Henderson, eds., *Industrial Organization and International Trade: Methodological Foundations for International Food and Agricultural Market Research*, NC-194 Research Monograph Number 1.

Pukelsheim, F. (1994): "The Three Sigma Rule," *American Statistician*, 48, 88-91.

Shannon, C. E. (1948): "A Mathematical Theory of Communication," *Bell System Technical*

Journal, 27, 379-423.

Shore, J. E., and R. W. Johnson (1980): "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," *IEEE Transactions on Information Theory*, IT-26, 26-37.

Skilling, J. (1989): "The Axioms of Maximum Entropy," in J. Skilling, ed., *Maximum Entropy and Bayesian Methods in Science and Engineering*, Dordrecht: Kluwer Acadmic, 173-87.

Appendix 1: GME-Nash with Unknown Demand Coefficients

In order to use the GME-Nash estimator when the parameters of the demand curves are unknown, we estimate the demand curves simultaneously with the rest of the model. For example, let the demand curve facing Firm *i* be

(A1.1)
$$\underline{q}_i = a_i + b_i \underline{p}_i + d_i \underline{p}_j + \underline{u}_i \equiv X_i \underline{\beta}_i + \underline{u}_i ,$$

where $\underline{q}_i$ is the quantity vector, $\underline{p}_i$ is the price vector, a_i and d_i are positive scalars, b_i is a negative scalar, $\underline{u}_i$ is a vector of error terms, X_i is a matrix, and $\underline{\beta}_i$ is a vector of parameters. To use an entropy approach, we need to map the unknown parameters $\underline{\beta}_i$ and $\underline{u}_i$ into probability space. Following Golan, Judge, and Miller (1996), we model these unknown parameters as discrete random variables with finite supports. Let $\underline{\beta}$ be in the interior of an open, bounded hyperrectangle, $Z \subset \Re^K$, and, for each β_k, let there be a discrete random variable $\underline{z}_k$, with $M \geq 2$ possible realizations $z_{k1},..., z_{kM}$ and corresponding probabilities $p_{k1},..., p_{kM}$ such that

(A1.2)
$$\beta_k = \sum_{m=1}^{M} p_{km} z_{km} .$$

Letting Z be the M-dimensional support for $\underline{z}_k$, any $\underline{\beta} \in Z$ may be expressed as

(A1.3)
$$\underline{\beta} = Z \underline{p} = \begin{bmatrix} \underline{z}_1 & \underline{0} & \cdot & \underline{0} \\ \cdot & \underline{z}_2 & \cdot & \underline{0} \\ \cdot & \cdot & \cdot & \cdot \\ \underline{0} & \underline{0} & \cdot & \underline{z}_K \end{bmatrix} \begin{bmatrix} \underline{p}_1 \\ \underline{p}_2 \\ \cdot \\ \underline{p}_k \end{bmatrix},$$

where Z is a $(K \times KM)$ matrix and $\underline{p}$ is a KM-vector of weights such that $\underline{p}_K >> 0$ and $\underline{p}_K' \underline{1}_M =$

1 for each demand parameter for $k = 1, 2, 3$. The upper and lower bounds of $\underline{z}_k$, z_{k1} and z_{kM}, are far apart and known to contain β_k. Further, we use our knowledge of the signs of the unknown parameters from economic theory when specifying the support space Z.

The unknown and unobservable errors, u_{it}, are treated similarly. For each observation, the associated disturbance, u_{it}, is modelled as a discrete random variable with realizations $v_1^u, \ldots, v_J^u \in \underline{v}^u$ with corresponding probabilities $\omega_{t1}^u, \ldots, \omega_{tJ}^u$. That is, each disturbance may be modelled as

(A1.4)
$$u_{it} = \sum_{j=1}^{J} \omega_{tj}^{u} v_j^{u},$$

for each $t = 1, \ldots, T$. The elements of the vector $\underline{v}^u$ form an evenly spaced grid that is symmetric around zero.

Given a sample of data $\underline{q}_i$, a simple way to determine the upper and lower bound of $\underline{v}^u$ is to use the three-sigma rule together with the sample standard deviation σ_q. For example, if $J = 3$, then $\underline{v}^u = (-3\sigma_q, 0, 3\sigma_q)$. Golan, Judge, and Miller (1996) has a detailed discussion of the statistical implications of the choice of bounds and sampling experiments for $\underline{v}^u$ and Z.

Having reparametrized the system of demand equation in this manner, the GME-Nash model with unknown demand parameters is

(A1.5)
$$\max_{\underline{\alpha}, \underline{w}, \underline{p}, \underline{\omega}} H(\underline{\alpha}, \underline{w}, \underline{p}, \underline{\omega}) = -\underline{\alpha}' \ln \underline{\alpha} - \underline{w}' \ln \underline{w} - \underline{p}' \ln \underline{p} - \underline{\omega}' \ln \underline{\omega},$$

subject to the consistency conditions 3.4, the necessary economic conditions 2.3 and 2.5, the Nash conditions 2.6, the two demand equations for Firms i and j, Equations A1.1, and the normalizations for $\underline{\alpha}$, $\underline{w}$, $\underline{p}$, and $\underline{\omega}$, where $\underline{\omega} = (\underline{\omega}^{\mu^i}, \underline{\omega}^{\mu^j}, \underline{\omega}^{\theta^i}, \underline{\omega}^{\theta^j}, \underline{\omega}^{\delta^i}, \underline{\omega}^{\delta^j}, \underline{\omega}^{u^i}, \underline{\omega}^{u^j})'$.

The bounds of the error supports for the demand equations $\pm 3\sigma_q$.

Appendix 2: Consistency

Call the GME-Nash estimates of the strategies $\underline{\tilde{\alpha}}$, the GME estimates $\underline{\breve{\alpha}}$, and the ME-ML estimates $\underline{\hat{\alpha}}$. We make the following assumptions:

Assumption 1: A solution of the GME-Nash estimator ($\underline{\tilde{\alpha}}$, $\underline{\tilde{w}}$, $\underline{\tilde{\omega}}$) exists for any sample size.

Assumption 2: The expected value of each error term is zero, its variance is finite, and the error distribution satisfies the Lindberg condition (Davidson and MacKinnon, 1993, p. 135).

Assumption 3: The true value of each unknown parameter is in the interior of its support.

We want to prove

Proposition: Given assumptions 1-3, and letting *all* the end point of the error support spaces $\underline{v}$ and $\underline{v}^d$ be normed by $\sqrt{T}$, plim($\underline{\tilde{\alpha}}$) = plim($\underline{\breve{\alpha}}$) = plim($\underline{\hat{\alpha}}$) = $\underline{\alpha}$.

When the profit parameters β are unknown, the GME-Nash estimates $\tilde{\beta}$ are consistent.

According to this proposition, the GME-Nash estimates, $\underline{\tilde{\alpha}}$, GME basic estimates, $\underline{\breve{\alpha}}$, and the ME-ML estimates, $\underline{\hat{\alpha}}$, are equal to each other and to the true strategies in the limit as the sample size becomes infinite, $T \rightarrow \infty$. That is, all the estimators are consistent.

Proof:

i) The ME-ML estimates are the observed frequencies: As $T \rightarrow \infty$, the observed frequencies converge to the population frequencies, so the ME-ML estimates are consistent: plim $\underline{\hat{\alpha}}_T = \underline{\alpha}$.

ii) The GME is consistent: Let the end points of the error supports of $\underline{v}$, v_1 and v_m, be $-1/\sqrt{T}$ and $1/\sqrt{T}$ respectively. As $T \rightarrow \infty$, $\psi_s \rightarrow 1$ for all s in the dual-GME, Equation 3.12. Thus, $\Sigma_s \ln \psi_s(\underline{\lambda}) \rightarrow 0$ and plim $\underline{\breve{\alpha}}_T = \underline{\alpha}$.

iii) The GME-Nash with known profit parameters is consistent: By Assumption 1, after we have added the restrictions 2.3 and 2.5, we still have a solution. The argument in (ii) together with Assumption 2 implies that plim $\underline{\tilde{\alpha}}_T = \underline{\alpha}$.

iv) The GME-Nash with unknown profit parameters is consistent: The normed moment version of the linear statistical model, Equation A1.1, is

$$\frac{X'\underline{q}}{T} = \frac{X'X}{T}\underline{\beta} + \frac{X'\underline{u}}{T}. \qquad \text{(A2.1)}$$

Given Assumption 3, the GME is a consistent estimator of $\underline{\beta}$ in Equation A2.1 (Golan, Judge, and Miller, 1996, Ch. 6): plim $\underline{\tilde{\beta}}_T = \beta$. By the argument in (iii), plim $\underline{\tilde{\alpha}}_T = \underline{\alpha}$.

Table 1: Average Lerner Indexes, (p - MC)/p

	Providence		Wichita	
	American	United	American	United
ME-ML: Observed	0.35	0.37	0.62	0.62
GME	0.34	0.35	0.62	0.61
GME-Nash	0.37	0.40	0.65	0.65
Conjectural Variation	0.37	0.40	0.65	0.64
Bertrand	0.37	0.40	0.66	0.65
Cournot	0.40	0.43	0.74	0.73
Collusive	0.45	0.48	0.79	0.78

Table 2: Sample Size Sampling Experiment (n = 20)

	$MSE(\underline{\alpha}^1)$	$MSE(\underline{\alpha}^2)$	Correlation, Firm 1	Correlation, Firm 2
T = 10				
ME-ML	.285	.145	.68	.79
GME	.137	.068	.66	.77
GME-Nash[1]	.086	.037	.79	.89
GME-Nash[2]	.110	.060	.66	.76
T = 20				
ME-ML	.263	.104	.69	.84
GME	.132	.050	.66	.81
GME-Nash[2]	.075	.023	.77	.91
T = 40				
ME-ML	.245	.091	.70	.86
GME	.124	.049	.67	.80
GME-Nash[2]	.075	.026	.78	.90

[1] Known demand coefficients.
[2] Unknown demand coefficients.

Figure 1a: Price Strategy of American Airlines, Chicago-Providence

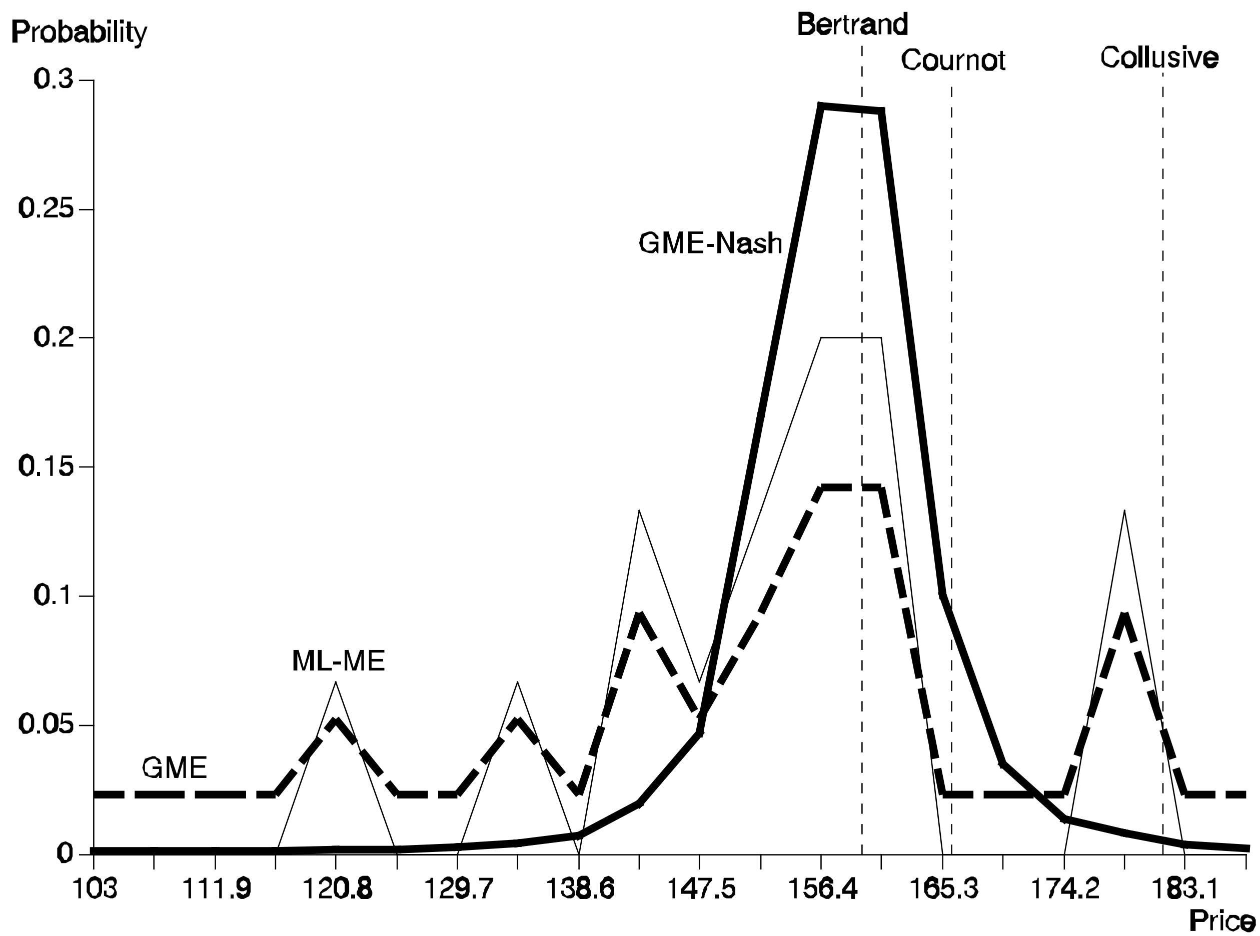

Figure 1b: Price Strategy of United Airlines, Chicago-Providence

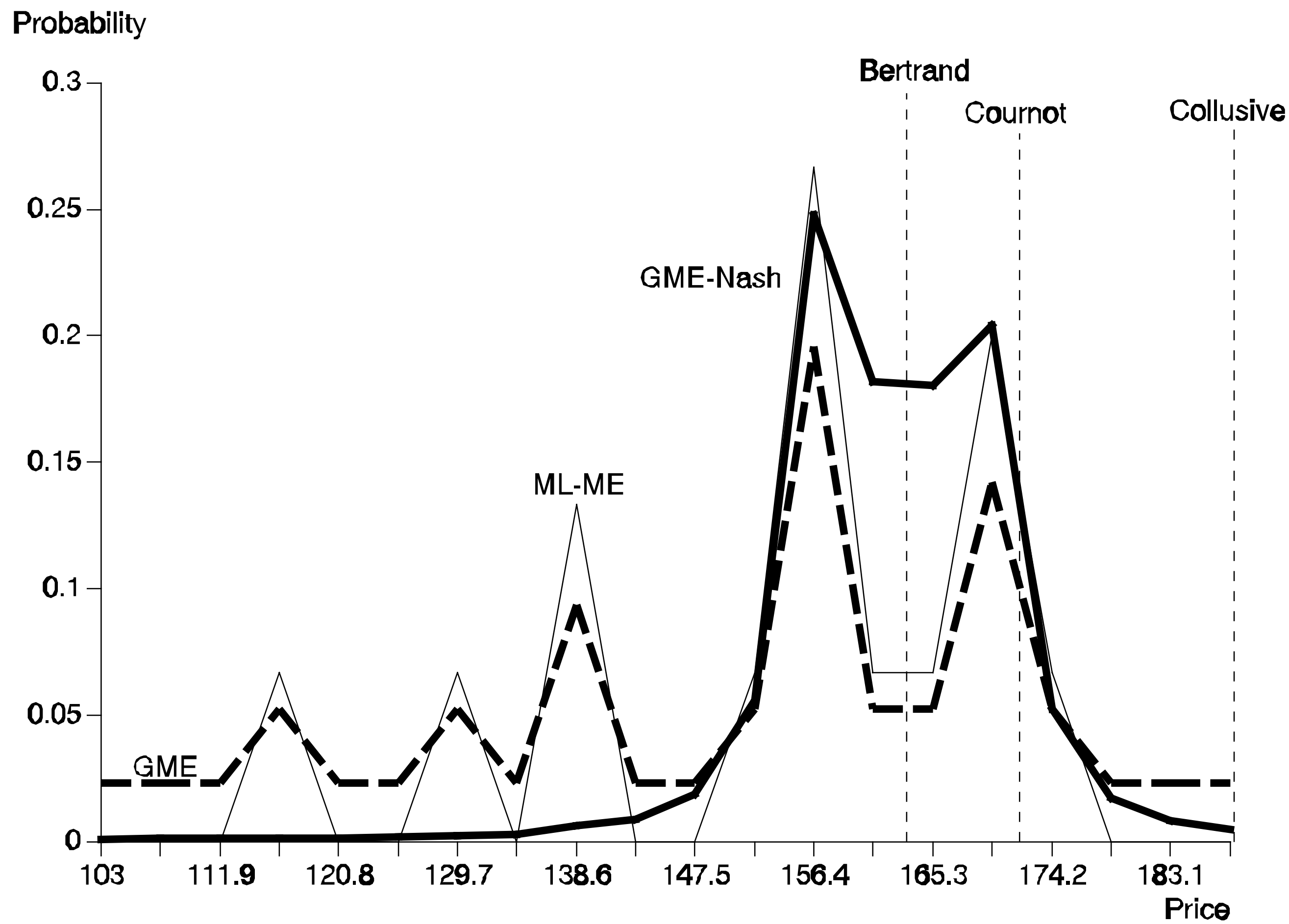

Figure 2a: Price Strategy of American Airlines, Chicago-Wichita

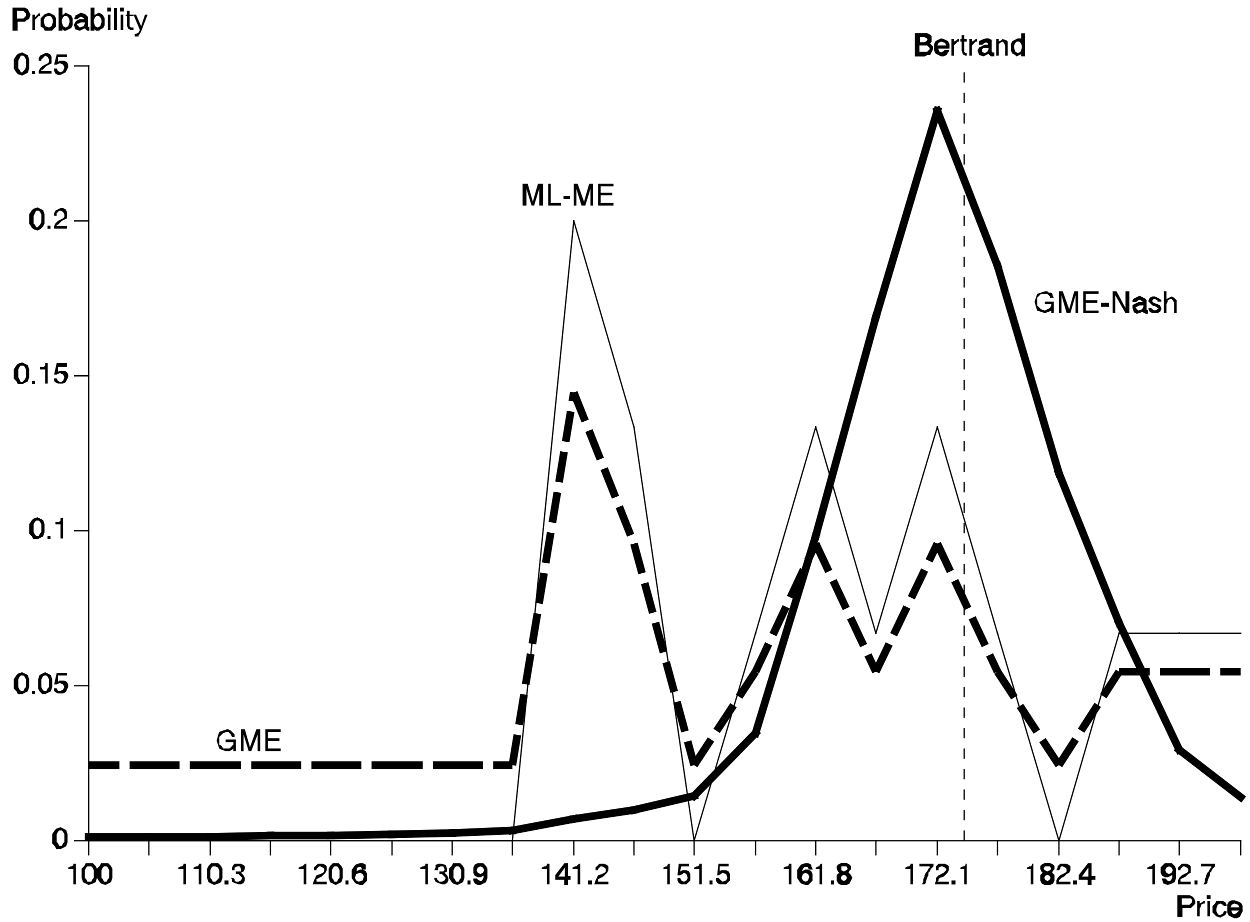

Figure 2b: Price Strategy of United Airlines, Chicago-Wichita

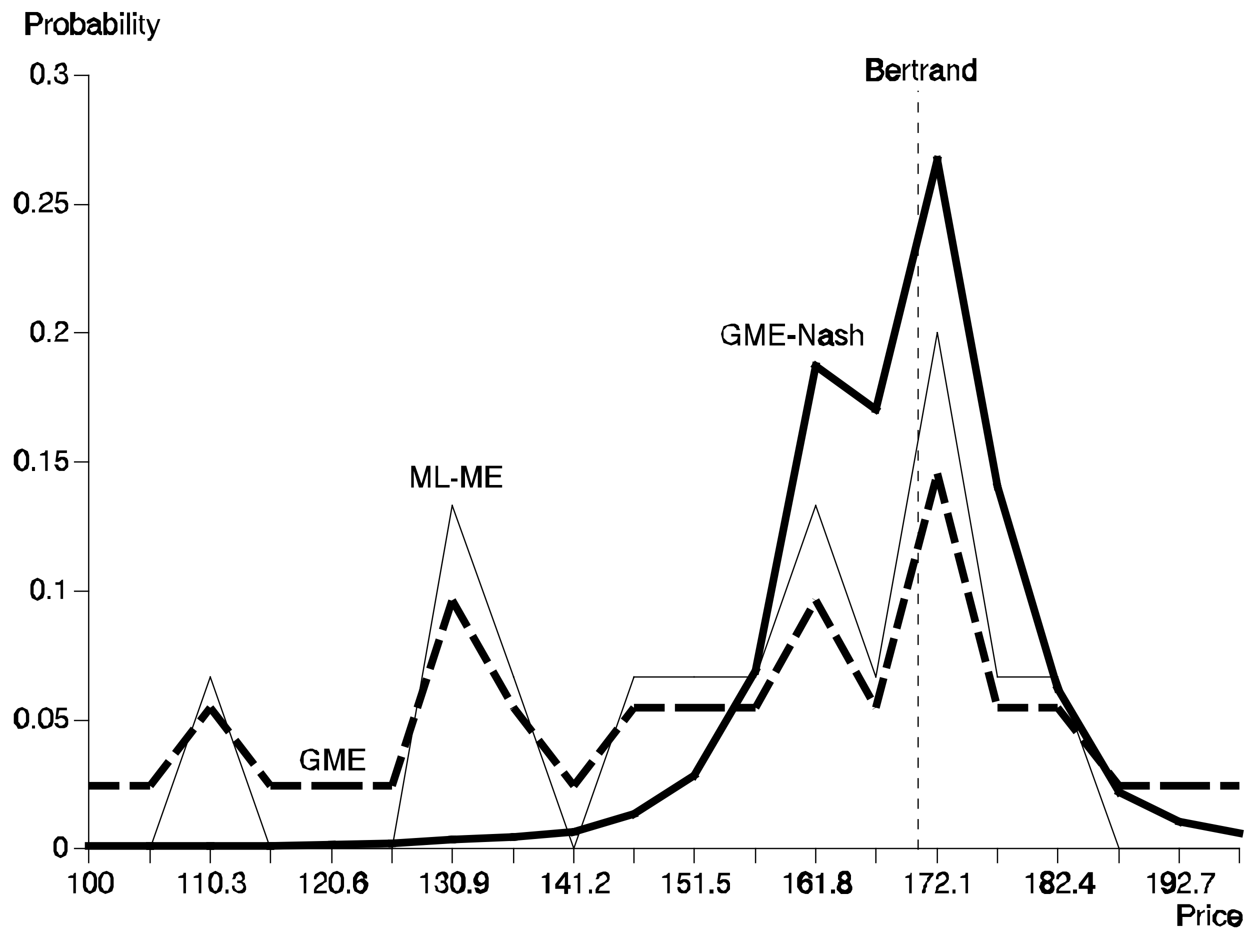

Figure 3a: Strategy and Conjectural Variation Models for American Airlines, Chicago-Providence

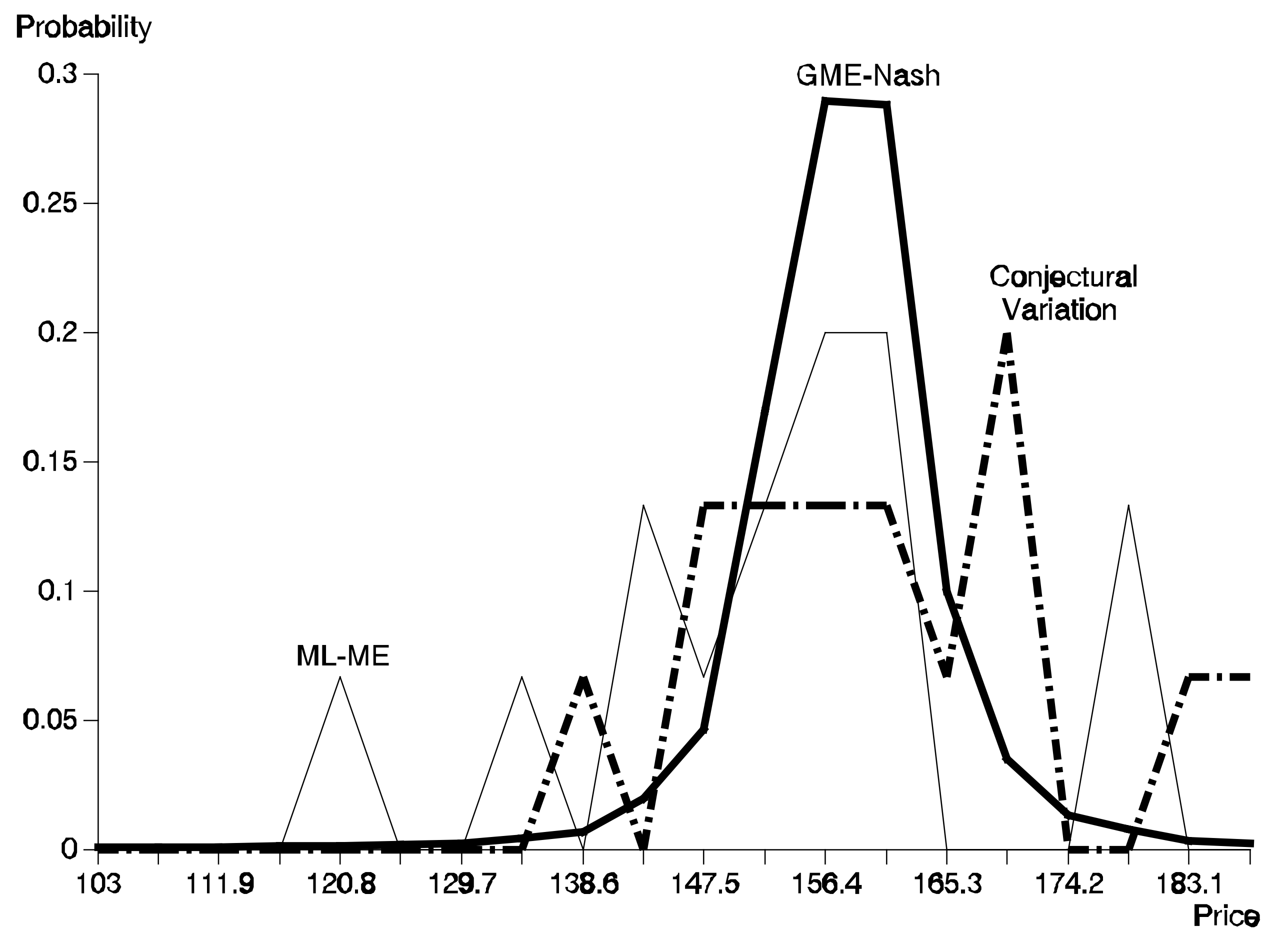

Figure 3b: Strategy and Conjectural Variation Models for United Airlines, Chicago-Providence