Title

Estimating Coke and Pepsi's Price and Advertising Strategies

Authors

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

Publication Date

1998-07-01

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

Working Paper No. 789

ESTIMATING COKE AND PEPSI'S PRICE ADVERTISING STRATEGIES

by

Amos Golan

Larry S. Karp

and

Jeffrey M. Perloff

REVISED

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
July, 1998

Estimating Coke and Pepsi's Price and Advertising Strategies

Amos Golan*
Larry S. Karp**
Jeffrey M. Perloff**

July 1998

* American University
** University of California, Berkeley, and Giannini Foundation

We benefitted greatly from George Judge's comments about econometrics and Leo Simon's comments on game theory. We are very grateful to Jean Jacques Laffont, Quang Vuong, and especially Farid Gasmi for generously providing us with the data used in this study and for advice. We received useful suggestions from participants at the "Industrial Organization and Food-Processing Industry" conference at the Institute D'Economie Industrielle at the Université des Sciences Sociales de Toulouse. We thank Gary Casterline and Dana Keil for substantial help with computer programs.

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

Table of Contents

Abstract

A semi-parametric, information-based estimator is used to estimate strategies in prices and advertising for Coca-Cola and Pepsi-Cola. Separate strategies for each firm are estimated with and without restrictions from game theory. These information/entropy estimators are consistent, are efficient, and do not require distributional assumptions. These estimates are used to test theories about the strategies of firms and to see how changes in incomes or factor prices affect these strategies.

KEYWORDS: strategies, noncooperative games, oligopoly, generalized maximum entropy, beverages

JEL: C13, C35, C72, L13, L66

1. INTRODUCTION

This paper presents two methods for estimating oligopoly strategies. The first method allows strategies to depend on variables that affect demand and cost. The second method adds restrictions based on a game-theoretic model. We use these methods to estimate the pricing and advertising strategies of Coca-Cola and Pepsi-Cola.

Unlike most previous empirical studies of oligopoly behavior, 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] Both our approaches recognize that firms may use either pure or mixed (perhaps more accurately, distributional) strategies.

In our application to Coca-Cola and Pepsi-Cola, we assume that the firms' decision variables are prices and advertising. We divide each firm's continuous price-advertising action space into a grid over prices and advertising. Then we estimate the vector of probabilities — the mixed or pure strategies — that a firm chooses an action within a rectangle in the price-advertising grid. We use our estimates to calculate the Lerner index of market structure and examine how changes in exogenous variables affect strategies.

The main advantages of using our method are that we can flexibly estimate firms' strategies subject to restrictions implied by game theory and test hypotheses based on these estimated strategies. The restrictions we impose are consistent with a variety of assumptions regarding the information that firms have when making their decisions and with either pure or mixed strategies.

---

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

For example, suppose that a firm's marginal cost in a period is a random variable observed by the firm but not by the econometrician. Given the realization of marginal cost, the firm chooses either a pure or a mixed strategy, which results in an action: a price-advertising pair. The econometrician observes only the firm's action and not the marginal cost. As a consequence, the econometrician cannot distinguish between pure or mixed strategies. If both firms in a market use pure strategies and each observes its rival's marginal cost, each firm can anticipate its rival's action in each period. Alteratively, firms might use pure strategies and know the distribution but not the realization of their rival's cost. Due to the randomness of the marginal cost, it appears to both the rival and the econometrician that a firm is using a mixed strategy. The equilibrium depends on whether firms' private information is correlated.

All of these possibilities — firms have only public information, firms observe each other's private information but the econometrician does not, or a firm only knows that its private information is correlated or uncorrelated with its rival's — lead to restrictions of the same form. For expositional simplicity, we concentrate on the situation where firms have private, uncorrelated information about their own - but not their rival's - marginal costs (or some other payoff-relevant variable) and choose a pure or mixed strategy.

There have been few previous studies that estimated mixed or pure strategies based on a game-theoretic model. These studies (Bjorn and Vuong 1985, Bresnahan and Reiss 1991, and Kooreman 1994) involve discrete action spaces. For example, Bjorn and Vuong and Kooreman estimate mixed strategies in a game involving spouses' joint labor market participation decisions using a maximum likelihood (ML) technique. Our approach differs from these studies in three important ways. First, they assume that there is no exogenous uncertainty. Second, they allow each agent a choice of only two possible actions. Third, in order to use a ML approach, they assume a specific error distribution and likelihood function. Despite the limited number of actions, their ML estimation problems are complex.

Our problem requires that we include a large number of possible actions in order to analyze oligopoly behavior and allow for mixed strategies. Doing so using a ML approach would be difficult if not impossible. Instead, we use a generalized-maximum-entropy (GME) estimator. An important advantage of our GME estimator is its computational simplicity. Using GME, we can estimate a model with a large number of possible actions and impose 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. However, a special case of our GME estimator is identical to the ML multinomial logit estimator (when the ML multinomial logit has a unique solution).

In Golan, Karp, and Perloff (1998 — henceforth GKP), we used the GME method to estimate mixed strategies for an airline duopoly. Our problem here is more general in two respects. The more important of these is that we allow the firms' strategies to be conditioned on the exogenous random variables observed by the econometrician. The second generalization is that we allow firms to make decisions over two variables, price and advertising, rather than just one, price.

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 Coke and Pepsi. 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, advertising, quantities, and variables that affect cost or demand, such as input prices or seasonal dummies. We assume that two firms, *i* and *j*, play a static game in each period of the sample.

The econometrician observes payoff-relevant public information, such as demand and cost shifters, $\underline{z}$, but does not observe private information known only to the firms. Firm *i* (and possibly Firm *j*), but not the econometrician, observes Firm *i*'s marginal cost or some other payoff-relevant random variable $\varepsilon^i(t)$ in period $t = 1, ..., T$. Where possible we suppress the time variable *t* for notational simplicity. The set of *K* possible realizations, $\{\varepsilon_1, \varepsilon_2, ..., \varepsilon_K\}$, is the same every period for both firms. The distributions are constant over time but may differ across firms. The firms, but not the econometrician, know these distributions. To simplify the description of the problem, we assume that ε^i and ε^j are private, uncorrelated information.

2.1 Strategies

The set of *n* possible actions (price-advertising pairs) for Firm *i* is $\{x^i_1, x^i_2, ..., x^i_n\}$. We now describe the problem where the random state of nature is private information and uncorrelated across firms.

The profit of Firm *i* in a particular time period is $\pi^i_{rsk}(\underline{z}) = \pi^i(x^i_r, x^j_s, \varepsilon^i_k, \underline{z})$, where *r* is the action chosen by Firm *i* and *s* is the action chosen by Firm *j*. In state *k*, Firm *i*'s strategy is $\underline{\alpha}^i_k(\underline{z}) = (\alpha^i_{k1}(\underline{z}), \alpha^i_{k2}(\underline{z}), ..., \alpha^i_{kn}(\underline{z}))$, where $\alpha^i_{kr}(\underline{z})$ is the probability that Firm *i* chooses action x_r given private information ε^i_k and public information $\underline{z}$. If Firm *i* uses a pure strategy, $\alpha^i_{kr}(\underline{z})$ is one for a particular *r* and zero otherwise.

Firm *j* does not observe Firm *i*'s private information, so it does not know the conditional probability $\alpha^i_{kr}(\underline{z})$. Firm *j* knows, however, the distribution of Firm *i*'s private information. The Nash assumption is that Firm *j* knows the unconditional probability of Firm *i* using action *r*. This probability is the expectation over Firm *i*'s private information: $\alpha^i_r(\underline{z}) = E_k\, \alpha^i_{kr}(\underline{z})$, where E_k is the expectations operator. Similarly Firm *i* knows the unconditional probability $\alpha^j_s(\underline{z})$ of Firm *j*.

In state *k*, Firm *i* chooses $\underline{\alpha}_k(\underline{z})$ to maximize expected profits, $\Sigma_s\, \alpha^j_s(\underline{z})\pi_{rsk}(\underline{z})$, where the expectation is taken over its rival's actions. If $Y^i_k(\underline{z})$ is Firm *i*'s maximum expected profits given ε^i_k and $\underline{z}$, then Firm *i*'s expected loss from using action x_r is

(2.1)
$$L^i_{rk}(\underline{z}) \equiv \sum_s \alpha^j_s(\underline{z})\, \pi^i_{rsk}(\underline{z}) - Y^i_k(\underline{z}) \leq 0,$$

which is non-positive. If it is optimal for Firm *i* to use action *r* with positive probability, the expected loss of using that action must be 0. Hence, optimality requires that

(2.2)
$$L^i_{rk}(\underline{z})\, \alpha^i_{rk}(\underline{z}) = 0.$$

The equilibrium to this game may not be unique. Our estimation method selects the pure or mixed strategy equilibrium that is most consistent with 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 private information ε^i_k. By taking expectations, we eliminate these unobserved variables and obtain usable restrictions.

We define $Y^i(\underline{z}) \equiv E_k\, Y^i_k(\underline{z})$ and $\pi^i_{rs}(\underline{z}) \equiv E_k\, \pi^i_{rsk}(\underline{z})$. Taking expectations with respect to k of Equations 2.1 and 2.2 and using the previous definitions, we obtain

$$\text{(2.3)} \qquad \sum_s \alpha^j_s(\underline{z})\, \pi^i_{rs}(\underline{z}) - Y^i(\underline{z}) \le 0\,,$$

$$\text{(2.4)} \qquad \left(\sum_s \alpha^j_s(\underline{z})\, \pi^i_{rs}(\underline{z}) - Y^i(\underline{z})\right)\alpha^i_r(\underline{z}) + \delta^i_r(\underline{z}) = 0\,,$$

where $\delta^i_r \equiv \mathrm{cov}(L^i_{rk}, \alpha^i_{rk}) \ge 0$. For each Firm $i = 1, 2$, we can estimate the unobservable strategies $\underline{\alpha}^i(\underline{z})$ subject to the conditions implied by Firm i's optimization problem, Equations 2.3 and 2.4.[2]

---

[2] If ε^i and ε^j are correlated or observed by both firms, the restrictions are slightly more complicated. If information is correlated, it would be reasonable to suppose that Firm i's beliefs about j's actions depend on the realization of ε^i, so that α^j_s is replaced by α^j_{ks}. If information is observed by both firms, Firm i's beliefs would also be conditioned on the realization of ε^j. In both cases, we can take expectations with respect to the private information and obtain equations analogous to 2.3 and 2.4. However, with either generalization, we would have an additional additive term in 2.3, say θ, and the definition of δ would be changed. The signs of both θ and δ would be indeterminate. This issue is discussed in GKP. In our empirical application to the cola market, all the estimated δ are positive, which is consistent with the model in the text where ε^i and ε^j are uncorrelated.

Firms may use approximately optimal decisions due to bounded rationality, or there may be measurement error. Therefore, we treat Equation 2.4 as a stochastic restriction and include additive errors in estimation. Equation 2.4, however, already has an additive function, $\delta(\underline{z})$, which we cannot distinguish from the additive error in 2.4. Thus, $\delta(\underline{z})$ is the only "error term" we include in this equation.

If we tried to estimate this model (Equations 2.3 - 2.4) 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, 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 may be 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. Our GME method is closely related to the GME multinomial choice approach in Golan, Judge, and Perloff (1996 — henceforth GJP). 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 sample-moment information and adding up constraints on the probabilities.

The frequency that maximizes entropy is an intuitively reasonable estimate of the true distribution when we lack any other information. If we have information from the experiment, such as the sample moments, or non-sample information about the random variable, such as restrictions from economic theory, we want to alter our "intuitively reasonable" estimate. The method of Maximum Entropy proceeds by choosing the distribution that maximizes entropy, subject to the sample and non-sample information.

In our game, the firms' price-advertising decisions are the random variables that correspond to $\underline{x}$ in the previous example. We want to estimate the firms' strategies, which are their probability distributions over their actions. The next two subsections explain how we incorporate sample and non-sample (theory) information. In our application, the sample information for cola manufacturers consists of time series of price-advertising pairs for each firm, quantities sold, and time series of exogenous variables that affect demand (a seasonal dummy and income) and cost (an interest rate, a wage rate, and the price of sugar). The game-theoretic restrictions, Equations 2.3 - 2.4, contain all the non-sample information.

3.2 Incorporating Sample Information

We incorporate the sample information into our GME estimator of the strategies, $\underline{\alpha}^i$, by maximizing the entropy of $\underline{\alpha}^i$ subject to the moment or consistency conditions that contain sample information. We can use either of two approaches, as shown in GJP. If we require that the moment restrictions hold exactly, we derive a ME estimator, which is identical to the ML multinomial logit estimator (when the ML estimate is unique).[3] If we view the moment conditions as stochastic restrictions, we obtain a GME estimator, which is a generalization of the multinomial logit. With either the ME or GME approaches, we obtain estimates of the probabilities $\underline{\alpha}^i$ as a function of the public information, $\underline{z}$.

---

[3] The number of parameters to be estimated using the ML multinomial logit is $(n - 1)L$, (where L = the dimension of $\underline{z}$, the number of covariates) because of the normalization used where the parameters for one category are set equal to zero. If $(n - 1)L < T$ and all categories are observed in the sample, the ML estimator may provide a unique solution. Otherwise, the ML problem is ill-posed. Even where the ML problem is ill-posed, the ME and GME estimators provide unique estimates, although the ME estimator is no longer equivalent to the (nonexistent) multinomial logit ML estimator.

In our problem, there are n actions, which are price-advertising pairs. The variable y^i_{tr} equals one in period t if action r is observed and zero otherwise. The variable y^i_{tr} is a function of the public information:

(3.2) $$y^i_{tr} = G(\underline{z}'_t \underline{\zeta}^i_r) + e^i_{tr} = \alpha^i_{tr} + e^i_{tr},$$

for i = 1, 2, where y^i_{tr} and $\underline{z}_t$ are observed and α^i_{tr}, e^i_{tr}, and ζ^i_r are unknown parameters to be estimated.

By eliminating $\underline{e}^i$ from Equation 3.2 and assuming that $G(\cdot)$ is a known cdf such as the logistic or the normal, we can estimate this model using maximum likelihood multinomial logit or probit. To avoid having to assume a specific cdf, we follow GJP and relate the set of covariates $\underline{z}_t$ to the data y^i_{tr} and the unknown α^i_{tr} and e^i_{tr}. We multiply Equation (3.2) by each covariate variable, z_{tl}, and sum over observations to obtain the stochastic sample-moment (data consistency) restrictions:

(3.3) $$\sum_t y^i_{tr} z_{tl} = \sum_t \alpha^i_{tr} z_{tl} + \sum_t e^i_{tr} z_{tl},$$

for l = 1, ..., L, which is the number of covariates in $\underline{z}$, and r = 1, ..., n. We obtain the basic GME estimator by maximizing the sum of the entropy corresponding to the strategy probabilities, $\underline{\alpha}^i$, and the entropy from the noise, $\underline{e}^i$, subject to that data consistency condition (3.3).

The GME objective is a dual-criterion function that depends on the weighted sum of the entropy measures from both the unknown and unobservable $\underline{\alpha}^i$ and $\underline{e}^i$. By increasing the weight on the $\underline{e}^i$ component of entropy, we improve the accuracy of estimation (decrease the mean square errors of the estimates of $\underline{\alpha}^i$). By increasing the weight on the $\underline{\alpha}^i$ component of entropy, we improve prediction (correct assignment of observations to price-advertising categories). The ME estimator is a special case of the GME, in which no weight is placed on the noise component and is similar to maximizing the logistic likelihood function. As a practical matter, our GME objective weights the $\underline{\alpha}^i$ and $\underline{e}^i$ entropies equally because we lack any theory that suggests other weights.

As we discussed in Section 3.1, the arguments of the Shannon's entropy measures must be probabilities. The elements of $\underline{\alpha}^i$ are probabilities, but the elements of $\underline{e}^i$ range over the interval [-1, 1]. To determine the entropy of $\underline{e}^i$, we reparameterize its elements using probabilities. We start by choosing a set of discrete points, called the support space, $\underline{v}^i$ = $[v_1^i, v_2^i, ..., v_M^i]'$ of dimension $M \geq 2$, that are at uniform intervals, symmetric around zero, and span the interval $[-1/\sqrt{T}, 1/\sqrt{T}]$, where T is the number of observations in the sample. Each error term e_r^i has corresponding unknown weights $\underline{w}_r^i = [w_{r1}^i, w_{r2}^i, ..., w_{rM}^i]'$ that have the properties of probabilities: $0 \leq w_{rm}^i \leq 1$ and $\Sigma_m w_{rm}^i = 1$.

We rewrite each error element as $e_r^i \equiv \Sigma_m v_m^i w_{rm}^i$. For example, if $M = 3$, then $\underline{v}^i$ = $(-1/\sqrt{T}, 0, 1/\sqrt{T})'$, and there exists w_1^i, w_2^i, and w_3^i such that each noise component can be written as $e_r^i = -w_{r1}^i/\sqrt{T} + w_{r3}^i/\sqrt{T}$. Using this parameterization, we represent the GME consistency conditions, Equation 3.3, as

$$\begin{aligned} \sum_t y_{tr}^i z_{tl} &= \sum_t \alpha_{tr}^i z_{tl} + \sum_t e_{tr}^i z_{tl} \\ &= \sum_t \alpha_{tr}^i z_{tl} + \sum_t \sum_m w_{trm}^i v_m^i z_{tl}^i . \end{aligned} \tag{3.4}$$

Other than M, 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 (GJP; Golan, Judge, Miller, 1996) show a substantial decrease in the mean square error (MSE) of estimates when M increases from 2 to 3. Further increases in M provide smaller incremental improvements. The estimates hardly change if M is increased beyond 7.

For notational simplicity, we now drop the firm superscript. If we assume that the actions, $\underline{x}$, and the errors, $\underline{e}$, are independent and define $\underline{w}$ as the vector which contains the elements w_{trm}, 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

$$\underline{1}'\underline{\alpha}_t = 1, \tag{3.6}$$

$$\underline{1}'\underline{w}_{ts} = 1. \tag{3.7}$$

for $s = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$.

The Lagrangean to the GME problem is

$$\begin{aligned} \mathfrak{L}\left(\underline{\lambda}, \underline{\rho}^{\alpha}, \underline{\rho}^{w}\right) = & -\sum_t \sum_r \alpha_{tr} \ln \alpha_{tr} - \sum_t \sum_r \sum_m w_{trm} \ln w_{trm} \\ & + \sum_l \sum_r \lambda_{lr} \left(\sum_t y_{tr} z_{tl} - \sum_t \alpha_{tr} z_{tl} - \sum_t \sum_m v_m w_{trm} z_{tl} \right) \\ & + \sum_t \rho_t^{\alpha} (1 - \underline{1}'\underline{\alpha}_t) + \sum_t \sum_r \rho_{tr}^{w} (1 - \underline{1}'\underline{w}_{tr}). \end{aligned} \tag{3.8}$$

where $\underline{\lambda}$, $\underline{\rho}^{\alpha}$, and $\underline{\rho}^{w}$ are vectors that stack the Lagrange multipliers. The vector $\underline{\lambda}_r$ is the L-dimensional sub-vector of $\underline{\lambda}$ containing the Lagrangean multipliers λ_{lr}, l = 1, 2, ..., L. We can write the GME estimates as functions of $\underline{\breve{\lambda}}_r$, the multipliers that minimize 3.8:

$$\text{(3.9)} \qquad \breve{\alpha}_{tr} = \frac{\exp\left(-\underline{z}_t' \underline{\breve{\lambda}}_r\right)}{\sum_r \exp\left(-\underline{z}_t' \underline{\breve{\lambda}}_r\right)} \equiv \frac{\exp\left(-\underline{z}_r' \underline{\breve{\lambda}}_r\right)}{\Omega_t},$$

$$\text{(3.10)} \qquad \breve{w}_{trm} = \frac{\exp\left(-\underline{z}_t' \underline{\breve{\lambda}}_r v_m\right)}{\sum_m \exp\left(-\underline{z}_t' \underline{\breve{\lambda}}_r v_m\right)} \equiv \frac{\exp\left(-\underline{z}_r' \underline{\breve{\lambda}}_r v_m\right)}{\Psi_{tr}}.$$

The Hessian is negative definite (the first n elements on the diagonal are $-1/\alpha_r$, the rest are $-1/w_{rm}$, and the off-diagonal elements are 0) so the solution is globally unique.[4]

We can reformulate the GME problem as a generalized logit likelihood function, which includes the traditional logit as a special case. Substituting Equations 3.9 and 3.10 into 3.8 to eliminate ln $\underline{\alpha}$ and ln $\underline{w}$, we can rewrite the first three elements of the Lagrangean as

$$\text{(3.11)} \qquad \begin{aligned} \mathfrak{L}(\underline{\lambda}) &= -\sum_t\sum_r \alpha_{tr}\left[-\sum_l z_{tl}\lambda_{lr} - \ln\theta_t\right] - \sum_t\sum_r\sum_m w_{trm}\left[-\sum_l z_{tl}\lambda_{lr}v_m - \ln\Psi_{tr}\right] \\ &\quad + \sum_r\sum_l \lambda_{lr}\left[\sum_t y_{tr}z_{tl} - \sum_t \alpha_{tr}z_{tl} - \sum_t\sum_m v_m w_{trm} z_{tl}\right] \\ &= \sum_t\sum_r\sum_l y_{tr}z_{tl}\lambda_{lr} + \sum_t \ln\left[\sum_r \exp\left(-\sum_l z_{tl}\lambda_{lr}\right)\right] + \sum_t\sum_r \ln\left[\sum_m \exp\left(-\sum_l z_{tl}\lambda_{lr}v_m\right)\right] \\ &\equiv \sum_t\sum_r\sum_l y_{tr}z_{tl}\lambda_{lr} + \sum_t \ln\theta_t + \sum_t\sum_r \ln\Psi_{tr}. \end{aligned}$$

---

[4] Equation 3.9 has the same form as the logistic CDF. GJP show that if $e_{tr} = 0$ in Equation 3.3 then $\underline{\hat{\zeta}}_r = -\underline{\breve{\lambda}}_r$, where $G(\cdot)$ in Equation 3.2 is logistic. This result shows that the ME estimator is equivalent to the ML logistic estimator when the latter is unique.

The second equality is obtained by noting that probabilities sum to one and simplifying. Minimizing the last equality in Equation 3.11 with respect to $\underline{\lambda}$ yields the same estimates as those obtained from maximizing the original formulation represented by Equations 3.4, 3.5, 3.6, and 3.7. The formulation in Equation 3.11 is computationally more efficient than the original formulation.

Henceforth we refer to the GME that uses only sample information as "the GME" estimator. When using the GME estimator, we may estimate $\underline{\alpha}^i$ and $\underline{\alpha}^j$ separately.

3.3 Incorporating the Non-Sample (Game-Theoretic) Information

The "GME-Nash" estimator is obtained by adding the game-theoretic restrictions, Equations 2.3 and 2.4, to the GME estimator. Initially, we suppose that we know the parameters of the profit function, π^i_{rs}. To use the GME-Nash estimator, we need to estimate $\underline{\alpha}^i$ and $\underline{\alpha}^j$ jointly because both strategy vectors appear in Equation 2.4. Further, we also have to estimate $\underline{\delta}^i(\underline{z})$, $i = 1, 2$, from Equation 2.4. As we discussed in Section 2.2, $\underline{\delta}^i(\underline{z})$ is nonzero if the econometrician does not observe firms' private information or if firms make mistakes in optimization. Our first step is to reparameterize $\underline{\delta}^i(\underline{z})$ using probabilities. Let $\underline{v}^a$ be a vector of dimension $J^a \geq 2$ with corresponding unknown weights $\underline{\omega}^a_r$ such that

(3.12) $$\sum_j \omega^a_{rj} = 1,$$

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

for $\underline{a} = \underline{\delta}^1, \underline{\delta}^2$. The support spaces $\underline{v}^a$ are defined to be symmetric around zero for all a_r.[5]

Let $\underline{y} = (\underline{y}^{i\prime}, \underline{y}^{j\prime})'$, $\underline{\alpha} = (\underline{\alpha}^{i\prime}, \underline{\alpha}^{j\prime})'$, $\underline{w} = (\underline{w}^{i\prime}, \underline{w}^{j\prime})'$, and $\underline{\omega} = (\underline{\omega}^{\delta^{i\prime}}, \underline{\omega}^{\delta^{j\prime}})'$. As above, we assume independence between the actions and the errors. The GME-Nash problem is

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

subject to the data consistency conditions 3.4 for each firm, the necessary economic conditions 2.3 and 2.4 for each firm, and the adding-up conditions for $\underline{\alpha}$, $\underline{w}$, and $\underline{\omega}$. The terms δ^i_r in Equation 2.4 are defined by Equations 3.12 and 3.13. Solving the problem (3.14) yields the estimates $\underline{\tilde{\alpha}}$, $\underline{\tilde{w}}$, and $\underline{\tilde{\omega}}$. As with the GME estimator, Equation 3.14 is a dual-loss objective where we maximize the sum of the entropies of the strategies $\underline{\alpha}$ and the error components $\{\underline{w}, \underline{\omega}\}$.

In order to write the Lagrangean for this problem, we need to determine how to impose the game-theoretic restrictions. Ideally, we would require that the game-theoretic restrictions hold for all possible values of $\underline{z}$. We cannot impose these restrictions for all values because we cannot write $\underline{\alpha}(\underline{z})$ in closed form independent of the unknown Lagrangean multipliers. Instead, we impose the weaker condition that the game-theoretic restrictions, Equations 2.3 and 2.4, hold at some or all of the values of $\underline{z}$ in our sample.

Because of the large number of restrictions for each value of $\underline{z}$ in our application, we impose the restrictions for only a subset of the observations. These restrictions, however,

---

[5] We do not have natural boundaries for $\underline{\delta}^i$, 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 larger of the empirical standard deviation of the discrete action space of prices or advertising.

affect the $\underline{\alpha}$ for all observations through the Lagrangean multipliers corresponding to the sample information (the moment conditions Equation 3.2).

If we do not know the parameters of the profit functions, we simultaneously estimate the strategies and the parameters of the profit function, which depend on the demand and cost functions. We do not have observations on cost, but we observe Firm *i*'s output, q^i, and some factor-cost (wages, price of sugar, and interest rate) variables, $\underline{z}_c$. We assume that Firm *i*'s cost is given by a cost function $C^i(q^i, \underline{z}_c; \eta^i)$, where η^i are parameters. Similarly, we assume that the demand for Firm *i*'s product is $q^i(x^i, x^j, \underline{z}_d; \phi^i)$, where $\underline{z}_d$ are demand shifters (income and a seasonal dummy) and ϕ^i are parameters. We substitute these functions in the Constraints 2.3 and 2.4 and estimate η^i and ϕ^i jointly with the other parameters. Because we observe demand (but not cost) we have an additional set of data consistency (sample) restrictions in the form of demand equations for each Firm *i*,

$$q^i = q^i(x^i, x^j, \underline{z}_d; \phi^i) + u^i,$$

where u^i is an error term. We estimate the parameters ϕ^i and η^i using the same method described in the previous subsection for estimating parameters that are not probabilities (see Appendix 1). That is, we choose a support for each such parameter and estimate the probability distribution over that support. We perform this estimation by maximizing the sum of all the entropy measure in equation 3.14 plus the entropy associated with the unknown demand and cost parameters.

3.4 Properties of the Estimators and Normalized Entropy

Both the GME and GME-Nash estimators are consistent, but they differ in efficiency and information content. GJP shows that the GME estimator is consistent given an appropriate choice of the bounds of the error term in the data consistency constraint 3.4. Under the assumption that a solution to the GME-Nash estimation problem exists for all samples, Appendix 2 uses a minor modification of the argument used in GKP to show that the GME-Nash estimator is consistent.

GJP show that the GME estimates of $\underline{\alpha}$ have smaller variances than the ME-ML multinomial logit estimates. The possible solution space for the GME-Nash estimate of $\underline{\alpha}$ 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. GKP reports sampling experiments that support this conjecture.

We can quantify the added information contained in the game-theoretic restrictions by comparing the normalized entropy of $\underline{\alpha}$ with and without the restrictions. The normalized entropy measure is $S(\underline{\alpha}) = -(\Sigma_r \alpha_r \ln \alpha_r)/(\ln n)$. The normalized entropy measure is $S(\underline{\alpha}) = 1$ if all outcomes are equally likely, and is $S(\underline{\alpha}) = 0$ if we know which action will be taken with certainty. The magnitude of the change in normalized entropy from imposing the game-theoretic restrictions provides a measure of the information they contain. See Appendix 2 for a derivation of the properties and inferences results for this estimator.

4. COLAS

Using quarterly data for 1968-1986, we estimate the price and advertising strategies for Coca-Cola and Pepsi-Cola using the GME and GME-Nash approaches. The Coca Cola Company and Pepsico, Inc. dominate the cola and soft-drink markets.[6] We use quarterly data

---

[6] In 1981 for example, Coca-Cola's share of colas was 44.4% and its share of the national carbonated soft-drink market was 27.8% (according to the *Beverage Industry Annual*

for 1968-86, which were obtained from a variety of secondary sources and are described in Gasmi (1988), Gasmi and Vuong (1991), and Gasmi, Laffont, and Vuong (1992).[7]

We assume that firms set prices and advertising and use the demand specification from these earlier studies:

$$q_t^{\,i} = \phi_0^i + \phi_1^i p_t^{\,i} + \phi_2^i p_t^{\,j} + \phi_3^i (A_t^{\,i})^{½} + \phi_4^i (A_t^{\,j})^{½} + \phi_5^i d + \phi_6^i I + u^{\,i}, \tag{4.1}$$

where $i = 1, 2$, $i \neq j$; q^i is the quantity sold, p^i is the real price charged, and A^i is the real advertising by Firm i; d is a seasonal dummy; I is income; u^i is an error term; ϕ_1^i is negative; ϕ_2^i and ϕ_3^i are positive.[8] In Appendix 1, we show how to reparameterize 4.1 so that it can be estimated along with the other parameters in the GME-Nash model. We assume that the marginal and average cost of Firm i is $c^i = \eta_0^i + \eta_1^i \times$ real price of sugar + $\eta_2^i \times$ real unit cost of labor in the nondurable manufacturing sector + $\eta_3^i \times$ real yield on a Moody's AAA corporate bond, where $\eta_0, \eta_1, \eta_2, \eta_3 \geq 0$.[9]

---

for 1986). The corresponding shares for Pepsi were 34.6% and 21.6%.

[7] The data were generously provided by these authors. We especially thank Farid Gasmi for patiently describing the data and making suggestions about the specification of our model.

[8] Some of the previous studies included output lagged one quarter as a proxy for the effects of advertising on market demand. We leave out that term because such intertemporal relationships are inconsistent with our assumption that the firms play repeated static games.

[9] The earlier studies did not include a constant term. Moreover, some of them used separate interest rates for the two companies. Because the correlation between these two interest rate measures is 0.99, we use only one.

4.1 Cola Estimates

For both the GME and GME-Nash models, firms have 35 possible actions in each period. We divide the range of possible prices into seven intervals and the range of possible advertising levels into five intervals.[10]

To estimate the GME-Nash model, we impose sign restrictions from economic theory on both the cost (all cost coefficients are non-negative) and demand parameters (demand falls with a firm's own price and rises with the other firm's price and its own advertising) and the game-theoretic restrictions in 15 periods (every fifth quarter starting with the third quarter). By only imposing the restrictions in about one-fifth of the periods, we greatly reduce the size of the estimation problem.

The Coca-Cola demand coefficients are ϕ_0 = 4.549, ϕ_1= -1.079, ϕ_2 = 2.137, ϕ_3 = 0.741, ϕ_4 = -0.232, ϕ_5 = 7.730, ϕ_6 = 0.737. The corresponding demand coefficients for Pepsi-Cola are -20.021, -1.596, 0.582, 0.808, -0.211, 5.592, and 2.056. The correlation coefficients between observed quantities and those predicted by the demand equation are 0.93 for Coke and 0.94 for Pepsi.

For the GME-Nash, the estimated cost parameters are η_0 = 13.482, $\eta_1 = \eta_2$ = 0 (due to the theoretical restriction that the coefficient be non-negative), and η_3 = 0.208 for Coca-Cola. The corresponding cost coefficients for Pepsi-Cola are 7.251, 0, 0, and 0.

---

[10] Within the sample, the prices range between $10.886 and $17.790 for Coca-Cola and between $6.646 and $9.521 for Pepsi-Cola. This difference in price levels is apparently due to the greater use of Coke syrup at fountains. Advertising expenditures range between 5.726 and 71.966 for Coca-Cola and from 7.058 to 50.814 for Pepsi-Cola.

The GME and GME-Nash estimating procedures produce estimates of the Lagrangean multipliers, which are the negative of the coefficients that would be estimated by a ML multinomial logit. Table 1 shows the GME estimates for Coca-Cola.[11] From the estimated coefficients, we can calculate the strategy probabilities, $\underline{\alpha}$, for each period. We show the estimates for the first quarter of 1977, near the midpoint of the sample, for Coca-Cola in Figure 1 and Pepsi-Cola in Figure 2. In both figures, panel a shows the GME estimates and panel b shows the GME-Nash estimates.

For both companies, the GME probability estimates are more uniform (reflect greater entropy) than the GME-Nash estimates. These figures illustrate that the game-theoretic conditions contain additional information beyond that in the data alone. If this theoretical information is true, it improves our estimates.

The corresponding marginal distributions for price and advertising strategies for both the GME and GME-Nash models are shown in Figure 3 for Coke and in Figure 4 for Pepsi. The GME-Nash marginal distributions put more weight on the category with the largest probability than do the GME marginal distributions.

This pattern is repeated in virtually all periods. We can compare the different estimators empirically using the normalized entropy (information) measure $S(\underline{\alpha})$. The normalized entropy measures for the GME, 0.66 (Coke) and 0.73 (Pepsi), are closer to one (the upper bound of entropy) than are the corresponding GME-Nash measures, 0.31 and 0.41. These numbers show the extent to which the game theoretic restrictions bind: They measure

[11] To save space, we do not report the coefficients for Pepsi or the two tables for the GME-Nash. These tables are available from the authors.

the amount of additional information contained in the restrictions. The pseudo-R^2, which is the expected value of 1 - $S(\cdot)$ for both firms (see Appendix 2), is 0.31 for the GME model and 0.64 for the GME-Nash model.

The GME-Nash model is flexible enough to allow for both pure and mixed strategies. Out of the 76 periods of the sample, there are three periods for each firm where it uses a pure strategy.

4.2 Tests

We now test whether our theory is consistent with the firms' behavior (data), using tests derived in Appendix 2. The entropy-ratio test statistic is [2H(GME) - 2H(GME-Nash)] = 359.38 < $\chi^2_{1050,\ 0.05}$, where $H(\cdot)$ is the optimal value of the objective function. Thus, we conclude that the economic theory represented by the set of conditions (2.3) and (2.4) is *consistent* with the data.

We now compare the strategies (estimated $\underline{\alpha}$) of the GME and the GME-Nash models using the cross-entropy χ^2 test (Appendix 2).[12] We reject the null hypothesis that the GME and GME-Nash estimated strategies are identical in 34 periods (out of the 76 total periods) for Coke and in 25 periods for Pepsi. Thus, we conclude that the profit-maximizing, Nash restrictions are consistent with the data and contain useful information, so that imposing these restrictions affects our estimates of the strategies.

---

[12] These test results are for a 0.05 significance level. As there are seven support points for the price strategy and five for the advertising strategy, there are (7 - 1) × (5 - 1) = 24 degrees of freedom.

Next, we compare the strategies of the two firms for the GME-NASH model. We reject the null hypothesis that the two sets estimated strategies are identical in 74 out of the 76 periods. That is, the firm use *different* strategies.

For example, by comparing Figures 1b and 2b, we see that Coke and Pepsi had very different strategy distributions in the middle of the sample. Coke had a single-modal strategy distribution with most weight on a moderate price-moderately intense advertising strategy, whereas Pepsi had a bimodal distribution with the most weight on a high price-intensive advertising strategy.

Next, we investigate the significance of the individual covariates, l = income, price of sugar, wage, and bond rate. That is, we test whether the estimated coefficient is zero (H_0: $z_l = 0$) or nonzero (H_1: $z_l \neq 0$). The χ^2 test-statistic values are 54.41, 34.38, 78.07, and 56.51 for income, price of sugar, wage, and bond rate respectively. As a result, we reject the null hypothesis for *all* the covariates at the $\alpha = 0.01$ level. Thus, though the factor prices do not greatly affect the marginal costs of the firms, they do affect the strategies the firms use (see Section 4.4).

These estimators fit the data reasonably well, as Table 2 shows. For example, the GME-Nash estimator correctly predicted which of the seven price categories Coke chose 55% of the periods. Moreover, it missed by more than one category in only 11% of the periods. In this study, the GME predictions are more accurate than those of the GME-Nash.[13]

[13] Based on simulations in GKP, the GME-Nash can predict better than the GME. If the GME-Nash restrictions are correct, we expect the GME-Nash to have lower mean squared errors than the GME.

4.3 Lerner Measures

A standard measure of market power is the Lerner index, which is the percentage by which price is set above marginal cost. Usually, the Lerner index ranges between zero (competition) and one.

As we discuss in Appendix 3, we use our estimates of probabilities to calculate the expected Lerner index, $E[(p^i - c^i)/p^i] = \Sigma_r\ \alpha^i_r[(p^i_r - c^i)/p^i_r]$, where c^i is our estimate of Firm *i*'s marginal cost. We suppress the dependence of all functions on the public information, $\underline{z}$, and hold $\underline{z}$ constant for purposes of this discussion. In our study, the average adjusted Lerner index is 0.24 for Coke and 0.27 for Pepsi.

For comparison, we also calculated the Lerner index for the Bertrand-Nash model using the coefficients from Model 1 of Gasmi, Laffont, and Vuong (1992). Averaged over the sample, the index for Coke is 0.42 and for Pepsi is 0.45. Thus, the GME-Nash estimates indicate that firms have less market power than do ML estimates of a Bertrand-Nash equilibrium.[14]

4.4 Effects of the Exogenous Variables

Using our estimated models, we can calculate the effect of a change in each of the exogenous $\underline{z}$ variables on the strategy probabilities, $\underline{\alpha}$, using the same approach as is used with logit and probit models. Table 3 shows the average strategy elasticities using the GME-Nash estimates (the percentage change in expected action divided by the percentage change in

---

[14] These differences in the Lerner indexes are largely due to differences in the estimates of costs. Our GME-Nash cost estimates are substantially higher than their ML-Bertrand estimates.

a *z* variable).[15] Some of these elasticities are large in absolute value because the corresponding probabilities are close to zero.

By inspection of Table 3, we see that an increase in income, which shifts out demand, increases the probability that Coke, and to a lesser extent Pepsi, charge higher prices. The elasticity of Coke's expected price with respect to income is 0.154 and the corresponding elasticity for Pepsi is 0.0013.

An increase in income (and demand) spreads a unit cost of advertising over a greater volume of sales, so we expect higher income to shift the distribution for advertising to the right. Pepsi's advertising strategy does shift to the right, but for Coke more probability weight is shifted to both tails. The elasticity of expected advertising with respect to income is -0.097 for Coke and 0.03 for Pepsi.

We can calculate similar elasticities with respect to the other exogenous variables. According to our estimates, the corporate bond rate does not directly affect Pepsi's costs, and it has a negligible effect on Pepsi's strategy. Despite the absence of a direct effect (via costs), the bond rate might indirectly affect Pepsi's strategy, possibly because Pepsi thinks that it alters Coke's strategy. Coke's costs increase with the interest rate. For Coke, an increase in the interest rate changes the mix of probabilities of charging a high price. An

[15] First, we calculate the derivatives of probabilities and average them over the 59 periods where we do not impose the game-theoretic constraints. Then, we average the probabilities for these periods. Using these averages, we summed over categories to compute the marginals of the averages, and used the results to calculate the elasticities of the marginals. By using the periods where we do not impose the constraints, we are able to use an explicit formula for probabilities when calculating the derivatives. Had we used the periods where the game theoretic constraints are imposed, we would have had to calculate the derivatives of a system of 70 implicit equations.

increase in the interest rate shifts more weight to the tails of Coke's advertising strategy. The elasticities with respect to the corporate bond rate are 0.004 for Coke's expected price, 0.032 for Coke's expected advertising, 0.004 for Pepsi's expected price, and 0.0005 for Pepsi's expected advertising. Thus, for changes in either income or bond rates, Coke responds more than does Pepsi.

5. CONCLUSIONS

We developed two methods of estimating the strategies of firms, which are the probabilities of taking particular actions. In our application to the cola market, the actions are price-advertising pairs. Both methods are free of parametric assumptions about distributions and ad hoc specifications such as those used in conjectural-variations models. Unlike previous studies of oligopoly behavior that only allowed for pure strategies, we allow for both pure and mixed strategies.

Our simplest approach is to use generalized maximum entropy (GME) to estimate the strategies for each firm using only sample information. This method is more flexible and efficient than the standard maximum likelihood multinomial logit (ML) estimator. Both the traditional ML and the GME estimators ignore restrictions imposed by economic theory and some information about demand and costs.

Our generalized-maximum-entropy-Nash (GME-Nash) approach estimates firms' strategies consistent with the underlying data generation process and the restrictions implied by game theory. The application to the cola market demonstrates that both the GME and GME-Nash models can be used practically.

Tests show that the profit-maximizing, Nash restrictions are consistent with the data but that, because they contain information, alter our estimates of firms' strategies. We our able to use our estimates to show how changes in exogenous variables such as income or factor prices affect the firms' strategies.

Our GME and GME-Nash approaches 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 an appropriate alternative criterion.

REFERENCES

Bjorn, P. A. and Quang 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.

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.

Gasmi, Farid, *Econometrics of Duopolistic Games in Prices and Advertising: The Case of the U.S. Soft Drink Industry*, California Institute of Technology Ph.D., 1988.

Gasmi, Farid, Jean Jacques Laffont, and Quang H. Vuong (1992): "Econometric Analysis of Collusive Behavior in a Soft-Drink Market," *Journal of Economics & Management Strategy*, 1, 277-311.

Gasmi, Farid, and Quang H. Vuong (1991), "An Econometric Analysis of Some Duopolistic Games in Prices and Advertising," George F. Rhodes, Jr., ed, *Advances in Econometrics*, 9, Greenwich, Connecticut: JAI Press Inc.

Golan, Amos, and George Judge, "A Maximum Entropy Approach to Empirical Likelihood: Estimation and Inference," Working Paper, U.C. Berkeley, 1996. (Presented at the 1997 Summer Meetings of the North America Econometric Society).

Golan, Amos, George Judge, and Jeffrey M. Perloff (1996): "Recovering Information from Multinomial Response Data," *Journal of the American Statistical Association*, 91, 841-53.

Golan, Amos, Larry S. Karp, and Jeffrey M. Perloff (1998): "Estimating a Mixed Strategy: United and American Airlines," University of California, Berkeley, Department of Agricultural and Resource Economics Working Paper.

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

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.

Mittelhammer, R. C., and N. S. Cardell, "On the Consistency and Asymptotic Normality of the Data-Constrained GME Estimator in the GLM," Working Paper, Washington State University, 1997.

Owen, A., "Empirical Likelihood Ration Confidence Regions," *The Annals of Statistics*, 19 (1990), 1725-1747.

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.

Qin, J., and J. Lawless, "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22 (1994), 300-325.

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 Academic, 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. We rewrite the demand equation for Firm *i* from Equation 4.1 in matrix form as

(A1.1)
$$\underline{q}^{i} = X^{i}\,\underline{\phi}^{i} + \underline{u}^{i},$$

where $\underline{q}^i$ is the quantity vector, $\underline{u}^i$ is a vector of error terms, X^i is a matrix, and $\underline{\phi}^i$ is a *K*-dimensional vector of parameters. To use an entropy approach, we need to map the unknown parameters $\underline{\phi}^i$ and $\underline{u}^i$ into probability space. For notional simplicity, we suppress the firm index *i*. Following Golan, Judge, and Miller (1996), we model these unknown parameters as discrete random variables with finite supports. Let $\underline{\phi}$ be in the interior of an open, bounded hyperrectangle, $A \subset \Re^K$, and, for each ϕ_k, let there be a discrete random variable $\underline{a}_k$, with $M \geq 2$ possible realizations $\underline{a}_{k1},\ldots,\underline{a}_{kM}$ and corresponding probabilities $b_{k1},\ldots,b_{kM}$ such that

(A1.2)
$$\phi_k = \sum_{m=1}^{M} b_{km}\,a_{km}.$$

Letting *A* be the *M*-dimensional support for $\underline{\phi}_k$, any $\underline{\phi} \in A$ may be expressed as

(A1.3)
$$\underline{\phi} = A\,\underline{b} = \begin{bmatrix} \underline{a}_1 & \underline{0} & . & \underline{0} \\ . & \underline{a}_2 & . & \underline{0} \\ . & . & . & . \\ \underline{0} & \underline{0} & . & \underline{a}_K \end{bmatrix} \begin{bmatrix} \underline{b}_1 \\ \underline{b}_2 \\ . \\ \underline{b}_k \end{bmatrix},$$

where A is a ($K \times KM$) matrix and $\underline{b}$ is a KM-vector of weights such that $\underline{b}_K >> 0$ and $\underline{b}'_K\ \underline{1}_M = 1$ for each demand parameter for $k = 1, 2, ..., K$. The upper and lower bounds of $\underline{a}_k$, a_{k1} and a_{kM}, are far apart and known to contain ϕ_k. Further, we use our knowledge of the signs of the unknown parameters from economic theory when specifying the support space A.

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

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

for each $t = 1, ..., T$. 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 J.

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

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

subject to the consistency conditions 3.4, the necessary economic conditions 2.3 and 2.4, the two demand equations for Firms *i* and *j*, Equations A1.1, and the normalizations for $\underline{\alpha}$, $\underline{w}$, $\underline{b}$, and $\underline{\omega}$, where $\underline{\omega} = (\underline{\omega}^{\delta^i}, \underline{\omega}^{\delta^j}, \underline{\omega}^{u^i}, \underline{\omega}^{u^j})'$. The bounds of the error supports for the demand equations are $\pm 3\sigma_q$.

Estimating the unknown cost parameters is handled similarly by reparameterizing the cost parameters in the profit function. The associated probabilities enter directly into the objective function.

Appendix 2: The GME-Nash Estimator

A2.1 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 points of the error support spaces $\underline{v}$ (for each firm) be normed by $\sqrt{T}$, plim($\underline{\tilde{\alpha}}$) = plim($\underline{\breve{\alpha}}$) = $\underline{\alpha}$.

This result holds even when the profit parameters are unknown.

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

Proof:

i) The consistency of the GME estimator is proved in GJP. 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 $\underline{\breve{\alpha}}_T = \underline{\alpha}$.

ii) The GME-Nash with known profit parameters is consistent: By Assumption 1, after we have added the restrictions 2.3 and 2.4, we still have a solution. The argument in (i) together with Assumption 2 implies that plim $\underline{\tilde{\alpha}}_T = \underline{\alpha}$.

iii) The GME-Nash with unknown profit parameters is consistent. Given Assumption 3, the GME is a consistent estimator of $\underline{\phi}$ in Equation A1.1 (Mittelhammer and Cardell, 1996): plim $\underline{\tilde{\phi}}_T = \underline{\phi}$. By the argument in (ii), plim $\underline{\tilde{\alpha}}_T = \underline{\alpha}$. These asymptotic properties can also be established via the empirical likelihood approach (Owen, 1990; Qin and Lawless, 1994; Golan and Judge, 1996).

A2.2 Hypothesis testing

On the basis of the consistency of the estimators, we can define an "entropy ratio statistic" which has a limiting χ^2 distribution. We use this statistic to test hypotheses. In general, let $\underline{\lambda}$* be the vector of Lagrange multipliers for *all* the model's constraints. Let $H_M(\underline{\lambda}_0^*)$ be the entropy value of the problem where $\underline{\lambda}^* = \underline{0}$, or equivalently all the parameters (strategies as well as demand coefficients) are set equal to zero (or at the *center* of their supports). Thus, $H_M(\underline{\lambda}_0^*)$ is the maximum value of the joint entropies (objective function). It can be obtained by maximizing Equation (3.11) subject to no constraints (except for the requirement that all distributions are proper). Doing so yields the total entropy value of the three sets of discrete, uniform distributions $\underline{\alpha}$, $\underline{w}$, and $\underline{\omega}$. Now, let $H_u(\underline{\tilde{\lambda}}^*)$ be the objective (total entropy) value for the full GME-Nash model — the optimal value of Equation (3.14) — where $\underline{\lambda}^*$ is the set of estimated values (that is, they are not forced to equal zero).

The *entropy-ratio statistic* for testing the null hypothesis H_0 that all parameters are zero is

$$\mathcal{E}(\text{parameters} = \underline{0}) = 2H_M(\text{parameters} = \underline{0}) - 2H_u(\underline{\tilde{\alpha}}, \underline{\tilde{w}}, \underline{\tilde{\omega}}).$$

Under the mild assumptions we made above (or the assumptions of Owen, 1990 and Qin and Lawless, 1994), $\mathcal{E}(\text{parameters} = \underline{0}) \rightarrow \chi^2_K$ as $T \rightarrow \infty$ when H_0 is true and K is the number of restrictions. The approximate α-level confidence intervals for the estimates are obtained by setting $\mathcal{E}(\cdot) \leq C_\alpha$, where C_α is chosen so that $\Pr(\chi^2_K < C_\alpha) = \alpha$. Similarly, we can test any other hypothesis of the form H_0: $\underline{\alpha} = \underline{\alpha}_0$ for all, or any subset, of the parameters. We use these entropy-ratio statistics to test whether the economic and Nash restrictions are consistent with the data.

We use the same line of reasoning as above (each constraint, or data point, represents additional potential information that may lower the value of the objective function but can never increase it) to derive a "goodness of fit" measure for our estimator:

$$R^* = 1 - \frac{H_u(\underline{\tilde{\lambda}}^*)}{H_M(\underline{\tilde{\lambda}}^* = \underline{0})},$$

where $R^* = 0$ implies no informational value of the data set, and $R^* = 1$ implies perfect certainty or perfect in-sample prediction.

The small-sample approximated variances can be computed in a number of ways. We discuss two simplest approaches here. First, for each equation (say the two sets of demand equations), we calculate

$$\hat{\sigma}_i^2 = \frac{1}{T}\sum_t \tilde{u}_{it}^2,$$

where $\tilde{u}_{it} \equiv \sum_j \tilde{\omega}_{tj}^u v_j^{\,u}$ and $\hat{\mathrm{var}}(\phi_k^i) \cong \tilde{\sigma}_i^2 (X'X)^{-1}$ for each parameter ϕ_k^i. Similarly, for each set of equations, the relevant σ^2 is estimated.

Because our model is a system of a large number of equations, the elements of the asymptotic variance-covariance matrix, Ω, for the error terms of the entire system are estimated in the traditional way, taking into account all the data and all the restrictions (Equations 2.3, 2.5, 2.6, and A.6).

Finally, we note the relationship between the entropy objective and the χ^2 statistic. This relationship is used for comparison of various estimated strategies and various estimated distribution. The cross-entropy measures is defined as

(A2.1) $$I(\underline{\alpha}, \underline{\alpha}^o) = \sum_k \alpha_k \ln(\alpha_k/\alpha_k^o),$$

where $\underline{\alpha}^o$ is a proper prior probability distribution. Now, let $\{\alpha_k\}$ be a set of K observed frequencies (strategies) over a set of K observed prices. Let the null hypothesis be H_0: $\underline{\alpha} = \underline{\alpha}^o$, then

$$\chi^2_{(k-1)} = \sum_k \frac{1}{\alpha_k^o}(\alpha_k - \alpha_k^o)^2.$$

A second-order approximation of (A2.1) is

$$I(\underline{\alpha}, \underline{\alpha}^o) \cong \frac{1}{2}\sum_k \frac{1}{\alpha_k^o}(\alpha_k - \alpha_k^o)^2,$$

which is the entropy-ratio statistic (for evaluating $\underline{\tilde{\alpha}}$ versus $\underline{\tilde{\alpha}}^o$) that we previous discussed. We conclude by noting that two times the entropy-ratio statistic corresponds (at the limit) to $\chi^2_{(k-1)}$.

Appendix 3: Expected Lerner Index

In the text, we report an expected Lerner index, $E[(p^i - c^i)/p^i] = \Sigma_r\ \alpha^i_r[(p^i_r - c^i)/p^i_r]$, where c^i is our estimate of Firm i's marginal cost. We suppress the dependence of all functions on the public information, $\underline{z}$, and hold $\underline{z}$ constant for purposes of this discussion. In the absence of private information (and with no fixed cost such as advertising), the estimate of a firm's expected profit is positive if and only if the estimate of the expected Lerner index is positive. If there is private information, however, our estimate of the expected Lerner index can be negative even though our estimate of expected profits is positive.

We first note that Firm i's expected profit, conditional upon using its price, p^i_r, and advertising A^i_r, is $\Sigma_s\ \alpha^j_s\pi^i_{rs} = \Sigma_s\alpha^j_s[p^i_r - c^i]q(p^i_r, p^j_s, A^i_r, A^j_s) - A^i_r$, where $q(\cdot)$ is Firm i's demand function. Taking expectations with respect to Firm j's actions, we can rewrite Firm i's expected profit as $(p^i_r - c^i)Q(p^i_r, A^i_r) - A^i_r$, where $Q(p^i_r, A^i_r) \equiv \Sigma_s\ \alpha^j_s q(p^i_r, p^j_s, A^i_r, A^j_s)$ is the expected quantity demanded from Firm i conditional upon taking action r: choosing p^i_r and A^i_r. Using these expressions in Equation 2.4 and rearranging terms, we find that

(A3.1) $$\left[(p_r^{\,i} - c^{\,i})\,Q(p_r^{\,i},\, A_r^{\,i}) - A_r^{\,i} - Y^{\,i}\right]\alpha^i_r + \delta^i_r = 0.$$

By dividing both sides of Equation 3A.1 by a probability α^i_r that is strictly positive, we find that the expected Lerner index, Λ^i_r, for Firm i given it takes action r is

(A3.2) $$\Lambda^i_r \equiv \frac{p_r^{\,i} - c^{\,i}}{p_r^{\,i}} = \frac{Y^{\,i} + A_r^{\,i} - \delta^i_r/\alpha^i_r}{p_r^{\,i}\,Q(p_r^{\,i},\, A_r^{\,i})}.$$

As discussed in Section 2.2, with private uncorrelated information, $\delta^i_r \geq 0$. If it is optimal to use action *r* with positive probability for all realizations of private information ε^i_k, the loss $L^i_{rk} = 0$ for all *k*, and $\delta^i_r \equiv \text{cov}(L^i_{rk}, \alpha^i_{rk}) = 0$. If, however, there are some ε^i_k when it is not optimal to use action *r*, then $\delta^i_r > 0$. Thus, Equation 3A.2 shows that expected profits, Y^i, and the expected Lerner index can have opposite signs.

Suppose, for example, that there is no advertising, $A^i_r = 0$, or other fixed costs and that the market is very competitive so that expected profits, Y^i, are zero. It may still be the case that for some states ε^i_k it is optimal to use action *r* (that is, $\alpha^i_r > 0$) and for some ε^i_k it is not optimal to use action *r*, so $\delta^i_r > 0$. Given $A^i_r = Y^i = 0$, the Lerner index is never positive for any action, so the expected Lerner index is strictly negative. If we perturb the game so that equilibrium expected profits become slightly positive, then — assuming that the mixed strategies vary continuously — we could observe positive expected profits and a negative expected Lerner index. Indeed in our estimates, we observe negative Lerner indexes for some periods.

Thus, this Lerner index may be a misleading measure of market power. It is contaminated by the econometrician's lack of knowledge about private information. The reason that we estimate negative values in some periods is that our estimate of δ^i_r, the term which incorporates private information of firms, is positive. From the viewpoint of Firm *i*, $\delta^i_r = 0$, so if that firm calculated its own Lerner index, it would obtain a positive value where we estimate a negative one.

We can avoid this problem by using an adjusted Lerner index, Λ^{i*} that "purges" the original Lerner index of this private-information effect. By taking expectations over the actions *r*, we obtain:

$$\Lambda^{i*} \equiv \sum_r \left[\alpha_r^i \, \Lambda_r^i \; + \; \frac{\delta_r^i}{p_r^{\,i} \, Q(p_r^{\,i}, A_r^{\,i})} \right].$$

This adjusted Lerner index is unit free and has the same sign as expected profit if there are no fixed costs such as advertising.[16]

[16] We can only calculate this adjusted measure reported in the text for the periods in which we impose the game-theoretic restrictions because we only have estimates of $\underline{\delta}^i$ for those periods. The corresponding unadjusted Lerner indexes are -0.09 for Coke and 0.10 for Pepsi (these averages are the same for the entire sample and for just those quarters where the restrictions were imposed).

Table 1
GME Estimates of Coefficients for Coca-Cola

Price	Advertising	Constant	Seasonal Dummy	Income	Price of Sugar	Wage	Bond Rate
1	2	-46.738	1.851	0.832	1.207	49.393	-0.077
1	3	-43.588	-1.188	0.822	0.295	49.534	-0.115
1	4	1.441	0.017	-0.013	-0.098	-0.955	0.002
1	5	1.428	0.017	-0.013	-0.098	-0.940	0.002
2	1	77.525	0.834	-1.384	-0.041	-90.107	-0.052
2	2	10.040	0.896	-0.274	0.823	-11.263	-0.275
2	3	-20.309	-1.695	0.586	1.785	11.465	-0.396
2	4	51.699	-0.287	-0.860	-0.441	-60.512	-0.277
2	5	33.840	-1.141	-0.769	-0.266	-30.543	-0.024
3	1	1.428	0.017	-0.013	-0.098	-0.940	0.002
3	2	9.887	1.542	-0.196	-0.841	-12.846	-0.021
3	3	-13.102	-0.128	0.203	1.646	7.332	0.109
3	4	-61.368	-1.401	1.070	2.526	65.273	0.420
3	5	-0.719	-0.671	-0.190	1.200	5.090	0.357
4	1	21.993	0.756	-0.262	-2.909	-23.202	-0.168
4	2	-19.817	2.187	0.354	-2.500	25.788	-0.034
4	3	-5.802	0.697	-0.110	-0.903	11.068	0.291
4	4	59.688	-1.339	-1.201	-2.096	-59.023	0.281
4	5	1.428	0.017	-0.013	-0.098	-0.941	0.002
5	1	1.441	0.017	-0.013	-0.098	-0.955	0.002
5	2	1.425	0.017	-0.013	-0.098	-0.938	0.002
5	3	-22.607	-1.331	0.543	-0.379	23.203	-0.110
5	4	1.452	0.017	-0.013	-0.099	-0.966	0.002
5	5	1.462	0.017	-0.014	-0.099	-0.977	0.002
6	1	1.433	0.017	-0.013	-0.098	-0.947	0.002
6	2	1.438	0.017	-0.013	-0.098	-0.952	0.002
6	3	1.418	0.017	-0.013	-0.098	-0.930	0.002
6	4	-23.977	-1.188	0.479	0.901	24.449	0.037
6	5	1.453	0.017	-0.013	-0.099	-0.968	0.002
7	1	1.420	0.017	-0.013	-0.098	-0.932	0.002
7	2	-29.797	1.311	0.546	1.344	29.438	0.026
7	3	1.453	0.017	-0.013	-0.099	-0.968	0.002
7	4	1.418	0.017	-0.013	-0.098	-0.930	0.002
7	5	1.457	0.017	-0.013	-0.099	-0.972	0.002

Notes: The first two columns show the price and advertising categories. The coefficients in the missing first row are normalized to zero.

Table 2: Percent of Categories Correctly Predicted

	GME		GME-Nash	
	Price	Advertising	Price	Advertising
Coke	70	68	55	63
Pepsi	61	33	54	32

Table 3: Strategy Elasticities

	Categories						
	1	2	3	4	5	6	7
Interest Rate							
Coke Price	.000	.000	.000	.001	-2.920	2.998	-.192
Pepsi Price	.000	.000	-.003	.003	.000	.000	.000
Coke Advertising	.262	.000	-.005	-1.266	3.767		
Pepsi Advertising	-.004	.000	.000	.000	.003		
Income							
Coke Price	.000	.000	-.000	.015	-87.737	64.180	15.077
Pepsi Price	.000	.000	-.165	.204	.000	.001	.000
Coke Advertising	10.825	.000	.635	-32.756	75.448		
Pepsi Advertising	-.265	.000	.000	.000	.155		

Table 2: Percent of Categories Correctly Predicted

	GME		GME-Nash	
	Price	Advertising	Price	Advertising
Coke	70	68	55	63
Pepsi	61	33	54	32

Figure 1b: GME-Nash Estimates of Coke's Strategies (First Quarter 1977)

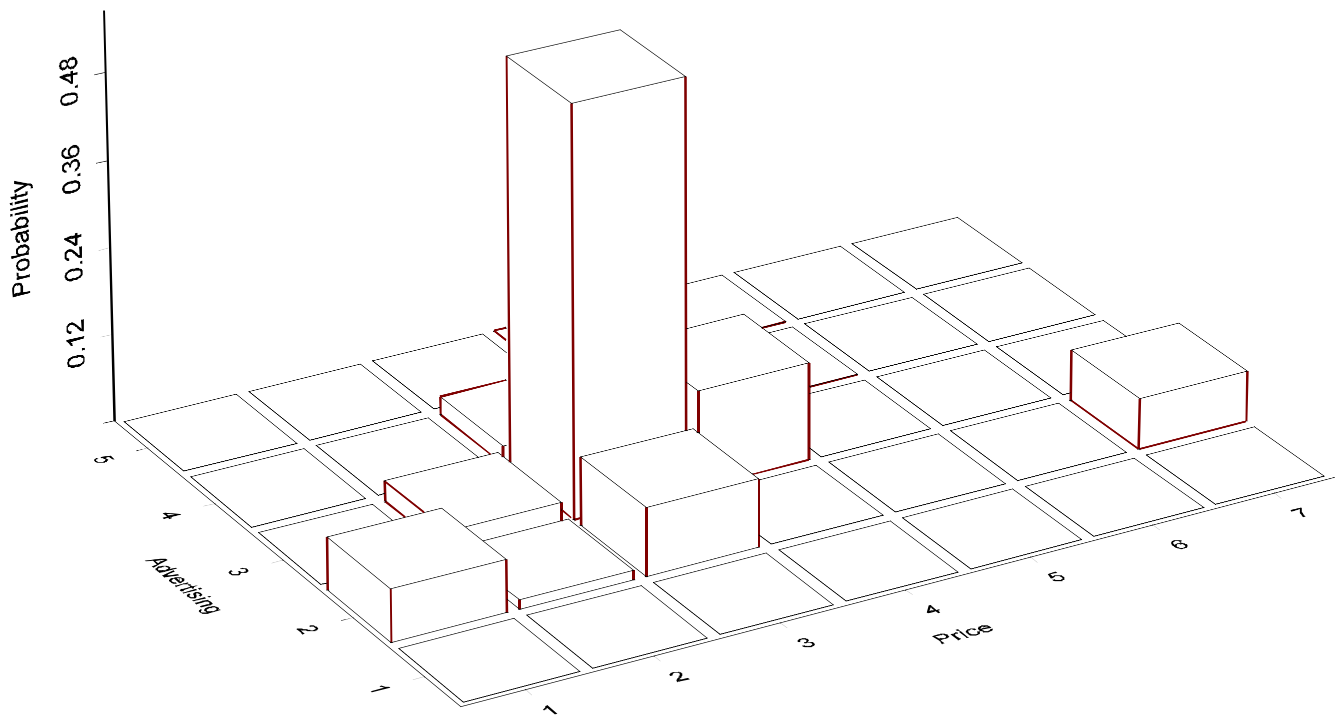


Figure 2a: GME Estimates of Pepsi's Strategies (First Quarter 1977)

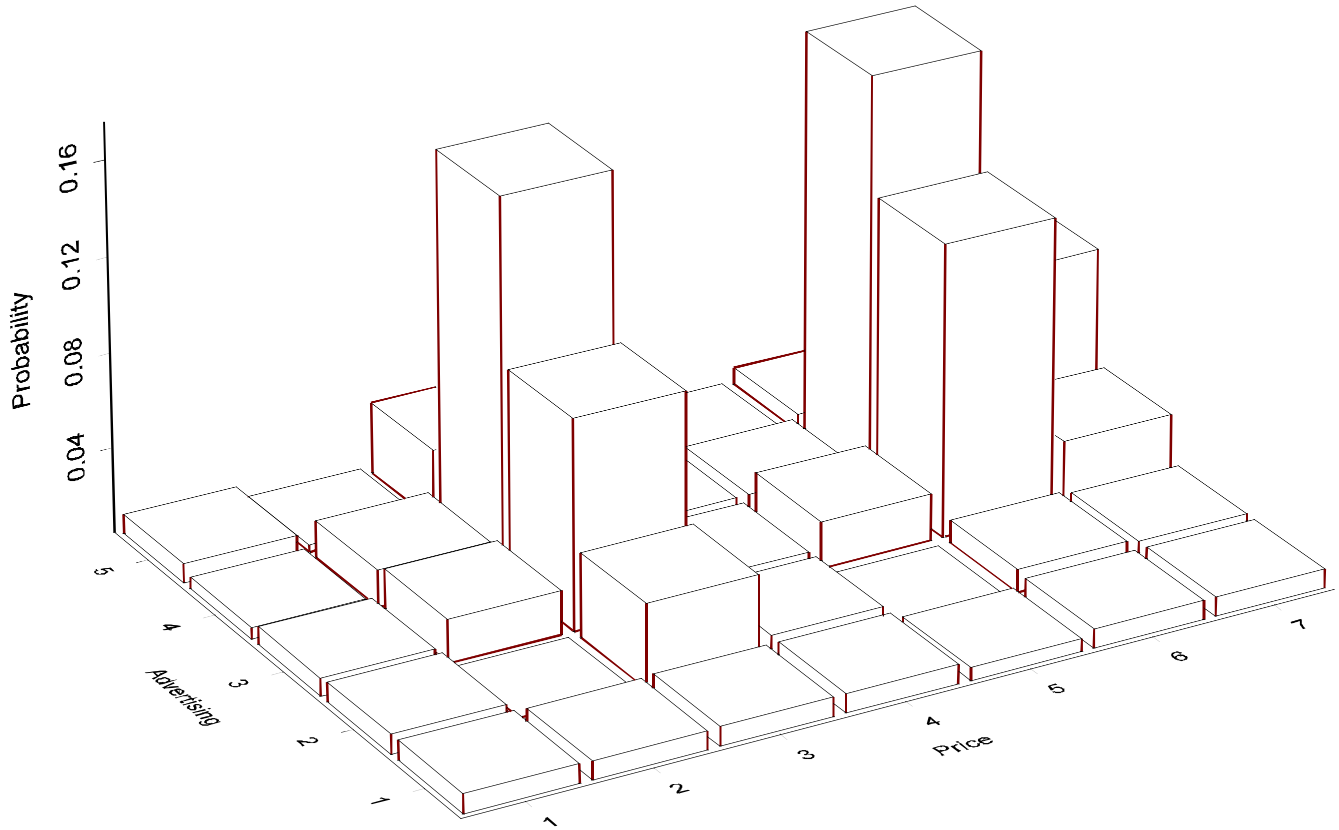


Figure 2b: GME-Nash Estimates of Pepsi's Strategies (First Quarter 1977)

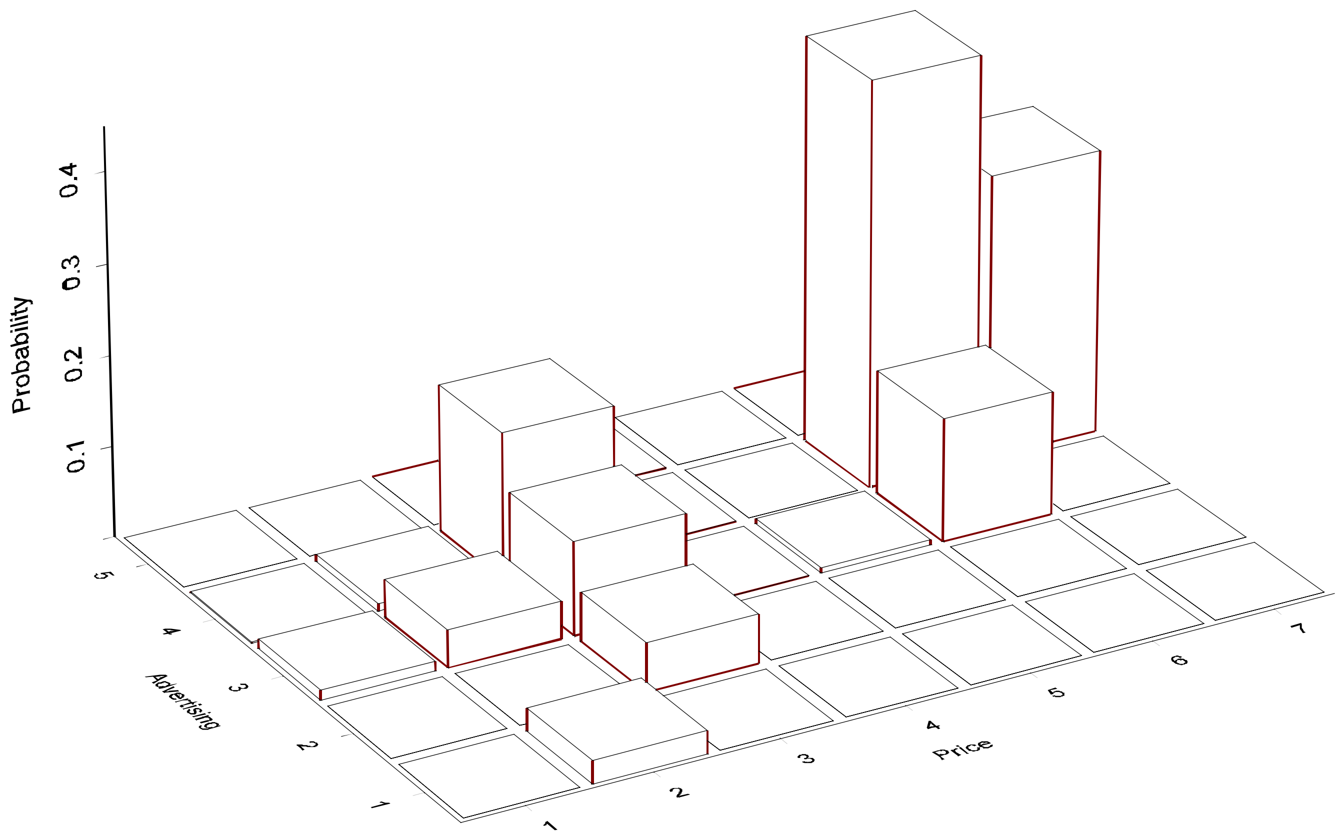
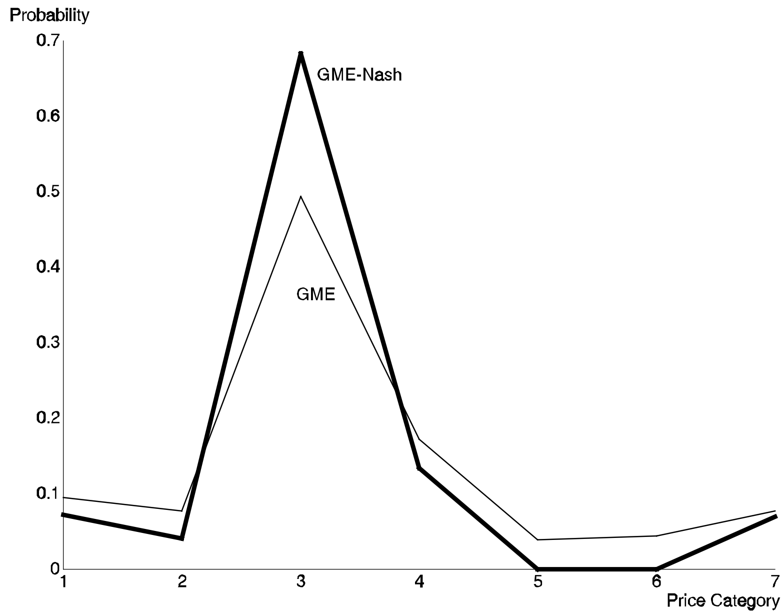


Figure 3: GME and GME-Nash Marginal Strategy Distributions for Coke (First Quarter 1977)

a) Pricing Strategies



b) Advertising Strategies

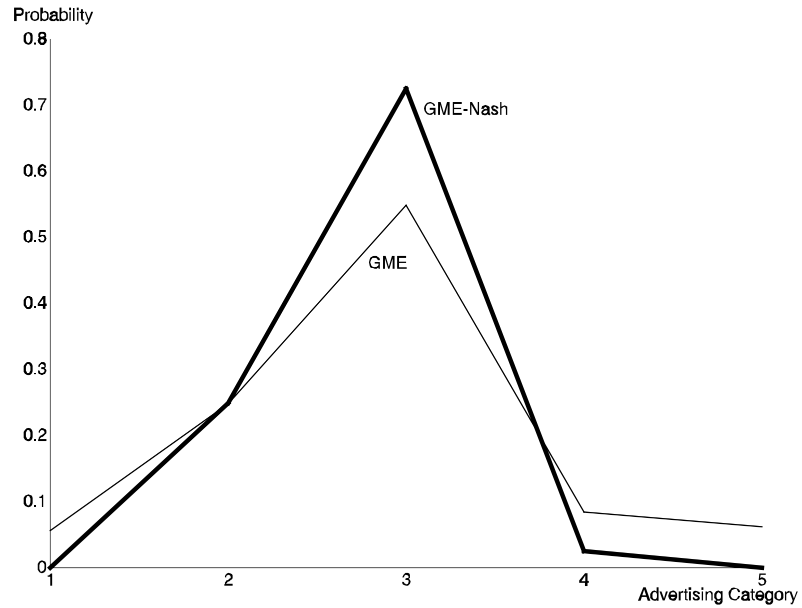
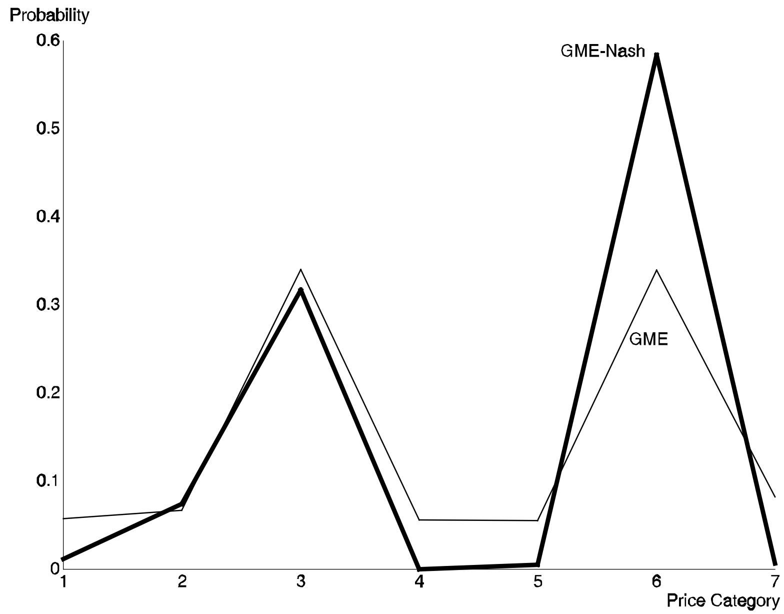


Figure 4: GME and GME-Nash Marginal Strategy Distributions for Pepsi (First Quarter 1977)

a) Pricing Strategies



b) Advertising Strategies