

## DATA APPENDIX AVAILABLE UPON REQUEST.

This study is based on two primary data sources. The first one is a 1982 survey of 616 supermarkets and their characteristics conducted on behalf of the Economic Research Service for a study of supermarket prices. The survey was supplemented by information on the 28 SMSA's where the supermarkets were located. The latter consisted primarily of items such as market growth, which was constructed from various issues of The Progressive Grocer for the same year. A detailed description of this data is available in the study by Kaufman and Handy (1989). The second primary data source was purchased by ERS from Claritas Corporation (Rezide, 1980) and it contains socioeconomic and demographic information on each of the zip code areas where a supermarket is located. This information is derived from the 1980 Census.

A heated debate on the first data source has taken place.<sup>1</sup> One of the essential points of contention, if not the essential one, is whether the price index constructed should try to measure the average price level at a store, firm or SMSA in contrast to the average price level of the same basket of market goods for all stores, firms or SMSA's. There is agreement that the procedure used by Kaufman and Handy (1989, Appendix D) aims to measure the former concept. This is the concept that is suitable for our analysis, because it is the one consistent with the theoretical discussion in the paper. Of course, this does not imply that it is the one most suitable for all other analyses.

$p^*$  is a weighted average of indices of price relatives for each store. For each brand type in a product subcategory an all SMSA average of the unit price for the brand and subcategory was

---

<sup>1</sup>The critique of the data is in Geithman and Marion (1993); the defense of the procedures is in Kaufman and Handy (1993); and both can be found in R. Cotterill (ed.) Competitive Strategy Analysis in the Food System, Boulder, CO: Westview Press, 1993.

constructed<sup>2</sup>. For each item in a brand and product subcategory a price relative index was constructed through division of the unit price by the all SMSA average unit price and multiplying by 100. These indices of price relatives were aggregated into a single store index of price relatives, usually employing the share of sales as weights. This variable was taken directly from the ERS data set.

Data on prices of items were collected for three independent random samples of product subcategories for each supermarket.<sup>3</sup> The surveys were taken about six weeks apart. The index of price relatives for each store ranges from 66 to 135 over the three samples. After cleaning up the data this generates 461 price observations for the first wave, 498 for the second wave and 494 for the third wave. The number of supermarkets for which we have three observations on prices, however, is lower, since the overlap in the three waves is far from perfect. For instance, 430 observations are the same in all three waves; 487 are the same in waves 2 and 3; 434 are the same in waves 1 and 3 and 438 are the same in waves 1 and 2.

D is an index of distribution services. Each store was asked to indicate whether or not it provided each of 20 possible services: a 1 was recorded if it did and a zero otherwise.<sup>4</sup> One of these 20 services was the existence of a front end scanner. We deleted this one from the index on the grounds that it measures an input that made possible other services rather than an output. Through similar logic we used information on store hours to construct an additional service. If a store operates before

---

<sup>2</sup>Four brand type categories were used: advertised brands, private labels, generics and unbranded items.

<sup>3</sup>Product subcategories were selected from a universe of items that accounted for 95% of supermarket sales. In each wave the product subcategories accounted for more than 50% of the universe.

<sup>4</sup>See Appendix B of Kaufman and Handy (1989), which is cited in the references to the paper, for a detailed description of each of these services.

7 a.m. or after 10 p.m. and more than 15 hours per day, we assign it a value of 1 and if not a zero. Thus, the range of possible services goes from 1 to 20. This variable was divided by its average value over all SMSA's and multiplied by 100 to generate an index commensurate with the index of price relatives.

Distribution services can be classified into five broad categories: assortment, information assurance of product delivery, ambiance and accessibility of location. The components of the index described above capture four of these categories. Assortment is captured in the index by four questions on the existence of an area with employee assistance for bakery, deli, meat and fish as well as a question on whether the store accepts utility payments and another one on whether the store offers contests or games. Information is captured in the form of one question on whether product prices are marked and another on the use of unit pricing. Most of the services included can be viewed as providing assurance of product delivery in the desired form or at the desired time. That is, produce wrapping, check cashing, coupon redemption, express checkout, bagging service, carry out service, bottle deposits, trading stamps, the extended hours dummy discussed previously and continuity programs. Finally, ambiance is captured by a question on the use of employee uniforms and another on the availability of music in the store.<sup>5</sup> The only category of services not represented in the index is accessibility of location.

Q is a quantity index of output obtained by dividing annual store sales in hundreds of 1982 \$ by the store index of price relatives. These three variables comprise the set of endogenous variables in

---

<sup>5</sup>In the ERS version of the index the first four assortment services were weighted twice as much as the other sixteen on the grounds that they were more expensive to produce. We constructed two versions of the index: one with equal weights and one in which we weighted the same four assortment services and extended hours twice as much as the other fifteen. The results are very similar; thus, the paper reports only the equal weights version.

our analysis.

We took many of our independent variables from the store survey in the ERS data set. That is, labor compensation, which was measured as the average hourly wage rate per employee plus employee fringe benefits valued as a cost to the employer;  $S_1$ ,  $S_2$ ,  $S_3$  and  $S_4$ , are the store format, scanner and chain store dummies, respectively, which were defined in Section 4 of the paper.;  $X_{41}$  is a shopping center dummy variable which takes on the value of 1 if the store is located with other retail stores in a shopping center complex or mall and zero otherwise; and  $X_{43}$  is the selling area of the store in square feet.

In the original ERS study using this data aggregated to the firm level, the labor compensation variable had a negative sign and was statistically insignificant. Similar results arose at the store level. Part of the problem seems to be that there are substantial variations across SMSA's in this variable, which cloud the results. Consequently, we constructed an index based on this variable that removed all the between SMSA variation. Specifically, to obtain  $v_2$  in Section 4 of the paper we subtracted the mean within the SMSA from the value for the individual store and we added the grand mean overall the observations; finally we divided by the grand mean in order to convert the variable into an index. Hence, all variation in this variable is within SMSA variation.

Four variables were taken directly from the Claritas data set.  $X_1$  is the percentage of households without cars in the zip code area where a supermarket is located.  $X_2$  is an index of socioeconomic level of the zip code area in which a supermarket is located. It is based on four dimensions: household income, educational attainment, occupational status and home value. The index ranges from 1 to 100 and scores between 40 and 60 indicate middle class status. Fifty-five % of the zip codes in the U. S. have scores between 40 and 60. Fifty % of our zip codes have scores between 47 and 61.  $Y$  is the median income of households in the zip code area where a store is located.  $X_{42}$  is

the population in the zip code area where a store is located.

A number of other variables in the ERS data set were constructed on the basis of other published sources. From these we took  $v_1$  or an occupancy cost proxy which is an index of rental rates and utility costs relative to the 28 SMSA sample average;  $M_1$  or the market share of the firm that owns the store in the SMSA where the store is located.  $M_3$  or the average annual real growth of foodstore sales in an SMSA during the five years preceding the price survey and referred to as market growth. We constructed an additional variable, which was not in the ERS data set, from the same sources, i.e.,  $M_2$  or the number of stores per 1000 persons in the SMSA where the store is located.

Finally, the last variable used in the analysis,  $t$ , is an index of the opportunity cost of time in different market areas. This was measured by the index of between group variation in labor compensation across SMSA's. We used this variable to allow the shadow price of distribution services to differ across market areas.