Title: README file for Variety Pass-Through: An Examination of the Ready-to-Eat Breakfast Cereal Market

Authors: Timothy J. Richards and Stephen F. Hamilton

Introduction

This file is intended to provide readers with all necessary information required to obtain the estimates presented in the paper. As per the Journal data access policy, we acknowledge that the data are proprietary, but describe the necessary procedures to obtain the data, and include the Nlogit code used to estimate the model.

Data

Our data, which describes 33 months (June 2007 - March 2010) of supermarket chain-level retail sales of ready-to-eat breakfast cereal in the Los Angeles retail market, was purchased from IRI InfoScan under a USDA funding agreement. It represents cereal sales for the top six supermarket chains in Los Angeles: Albertsons, Food 4 Less, Ralphs, Safeway, Stater Brothers, Vons and Vons Pavilions. Our data include all branded UPCs, both private labels and national brands, in the breakfast cereal category and we focus our analysis on the 19 top brands (by volume share), subsuming all other brands into the outside option. Similar data may be obtained by contacting IRI at http://www.iriworldwide.com/.

Our wholesale cereal prices are from the PriceTrak data service provided by PromoData, Inc. PriceTrak data reports prices charged by manufacturers prior to the application of slotting allowances, markups of price over unit cost by wholesalers to retailers, the effective date of new case prices, "deal allowances" (off-invoice items offered to retailers by the wholesaler), the type of promotion suggested by the wholesaler to the retailer, and the allowance date. We define the wholesale price to be the price charged to the retailer net of all allowances. Unfortunately, PromoData, Inc. no longer exists, having declared bankruptcy in 2013.

The input-price data are from the Bureau of Labor Statistics database, and is publicly available. These data include average weekly earnings by workers in the grocery retailing industry, an index of healthcare costs paid by retailers, and an index of utility prices. We rely on primary BLS data on wages from the Current Employment Statistics survey, which provides detailed industry data on employment, hours, and earnings of workers, and we acquire utility prices from the market-specific indices provided in the BLS Consumer Price Index program. Finally, we use the Bureau of Census mean and standard deviation for each socio-economic and demographic variable (age and income) for the Los Angeles market to create random draws from the distribution of each variable in the random-coefficient nested logit model. We also rely on Census data for population figures for the Los Angeles metropolitan area, which we match with per-capita breakfast cereal consumption values obtained from the Economic Research Service of the USDA.

Estimation Methods

We estimate our structural model of demand and pass-through in two stages. We first estimate the

demand model using the control function method of Petrin and Train (2010), and then employ the Generalized Method of Moments estimator (Davidson and MacKinnon, 2004) for the second-stage pricing model.

In the demand model, retail prices and variety are likely to be endogenous in our market-level scanner data, and we test for this possibility using the Hausman (1978) specification. In the first stage of the control-function approach, we estimate an instrumental variables (IV) regression in which we regress variety and prices on a set of variables likely to serve as valid instruments. We then use the residuals from this regression as explanatory variables in the mixed-logit demand equation, which is then estimated using simulated maximum likelihood (Train, 2003). By introducing the IV residuals into the demand model, we account for unobservable factors that may be correlated with errors in the demand equation. We employ a variety of instruments to control for these factors. First, because product-specific variation in costs will tend to be correlated with prices for the same product but uncorrelated with unobservable factors in the demand equation, we interact input prices with a set of binary brand-indicators. Second, we include brand-, retailer- and season-specific binary variables to account for idiosyncratic supply factors or quality levels that may be important in determining retail prices. To instrument for variety decisions among retailers, we construct an index of the average number of cereal products stocked each week among all retailers in 23 U.S. markets excluding the Los Angeles market. Finally, to account for the role of retailing costs in determining optimal variety decisions we include a set of retailer-, brand- and season-specific-indicators as additional instruments.

We estimate our price and variety pass-through equations using GMM. This approach requires the use of instruments on the supply-side that reflect variation in demand but are independent of unobservable variables in the pricing equation. We accordingly capture retailer-specific variation in demand by including demographic variables such as income, age, and average household size as well as a set of retailer-specific binary variables and lagged retail-wholesale margins. Second, we include lagged margin values as proxies for current-period margins. Instruments for variety include the demographic variables and retailer-specific indicator variables used as pricing instruments, as well as seasonal indicators, the index of other-market variety described above, an index of total volume in other markets, lagged market-volume and lagged store-share.

Our conceptual model requires wholesale prices to be exogenous to retail pricing decisions. However, wholesale prices may be endogenous if prices set by suppliers in a specific market respond to retail pricing decisions underlying consumer demand. We instrument for wholesale prices using two complementary instrumentation strategies. First, we obtain wholesale prices for RTE breakfast-cereal in the Chicago market, which are highly correlated with wholesale prices in Los Angeles (correlation coefficient of 0.83). Second, we instrument wholesale prices with a vector of manufacturing input prices. The use of input prices as instruments for downstream prices is a standard practice in the literature (Berto Villas-Boas, 2007), as

input prices are correlated with production costs and yet mean-independent of demand for the final product.