

Appendix E—Data Sources and Construction of Variables

This appendix explains the sources and procedures that were used to construct the data for the empirical analysis in appendix D.

The regressions used quarterly data from the first quarter of 1994 through the third quarter of 2000. The InfoScan market areas that were selected for the sample were those for which 90 percent or more of the population resided in States that belonged to a single WIC contract. In addition, Mississippi was excluded from the regression sample because it does not use the retail distribution system.¹

Dependent Variable

The regressions' dependent variable is a market area's inflation-adjusted supermarket retail price for the specific infant formula products produced by Mead-Johnson, Ross, and Nestle identified in table 6-1. For each manufacturer, the four products selected were distinguished by two product bases (milk-based and soy-based) and by two product forms (liquid concentrate and powder). Chapter 5, "Source of Infant Formula Data" describes the nature and limitations of the InfoScan data on supermarket retail prices for these products. The quarterly CPI-U was used to deflate the nominal supermarket price series.

Independent Variables

The variable *relative size of WIC if contract brand* is an interactive variable, constructed as the product of *relative size of WIC* and a dummy variable, *contract brand status*, for which *contract brand status* equals 1 if the manufacturer's brand is the contract brand for the observation and *contract brand status* equals 0 if it is a noncontract brand.

The variable *relative size of WIC if noncontract brand* is an interactive variable, constructed as the product of *relative size of WIC* and a dummy variable, *noncontract brand status*, for which *noncontract brand status* equals 1 if the manufacturer's brand is a noncontract brand for the observation and *noncontract brand status* equals 0 if it is the contract brand.

Information from USDA's Food and Nutrition Service (FNS) identified which manufacturer was a State's WIC contract brand in each quarter. If the manufacturer of the contract brand changed after the middle of a quarter, the change was treated as taking place at the start of the following quarter.

The variable *relative size of WIC* in a State was computed as the number of formula-fed WIC infants relative to the number of formula-fed non-WIC infants in the State. The number of formula-fed WIC (non-WIC) infants was calculated as the product of the number of WIC (non-WIC) infants in the State and the State's WIC (non-WIC) formula-feeding rate. A State's *relative size of WIC* value was used to proxy the relative size of WIC value in all market areas within the State.

The number of WIC infants, by State, was obtained from USDA's Food and Nutrition Service (FNS). The annual number of live births, by State, was available from the U.S. Department of

¹ Furthermore, certain quarterly observations were dropped for market areas in the sample if either there was more than one contract brand (Open Market System) in effect during a quarter (Phoenix; seven quarters of 1994:1-1995:3) or, in one instance, if there was no contract brand for a quarter (Atlanta; 1994:4).

Health and Human Services' Centers for Disease Control and Prevention (CDC). Quarterly data for each series were obtained by interpolation of annual data. The difference between live births and WIC infants represents the number of non-WIC infants. Formula-feeding rates were obtained from breastfeeding rates using the relationship that a breastfeeding rate and its corresponding formula-feeding rate sum to 1.0. Annual data for the breastfeeding rates of women, by State, were obtained from a breastfeeding survey conducted by Ross Laboratories (Ross Labs Mothers' Survey, various years). Separate data were available for breastfeeding rates for a State's WIC mothers (households) and for a State's non-WIC mothers (households). Annual data were interpolated on a quarterly basis.

The quarterly variable *number of contract brand changes* equals the number of contract brand changes that have occurred in the market area since the inception of the study period (for the product base and product form of each separate price regression). The variable is market-area specific rather than manufacturer-specific: its value is shared within a given market area for all manufacturers. Its values begin at zero in the first quarter of 1994 for each market area, and for some market areas it reaches a value of 2 by the end of the sample period in which case the contract brand changed twice among the manufacturers in the market area.

For each regression—for an infant formula product distinguished by manufacturer, product base, and product form—a quarterly dummy variable was constructed to represent the presence of Wyeth products in a market area. The variable *presence of Wyeth* equaled 1 for a given observation if Wyeth marketed an infant formula product of the relevant product base and product form in the given market area in the given quarter. Although Wyeth phased out production of its infant formulas for the U.S. market in 1996, the quarter in which it ceased marketing in any one market area differed across market areas.

A similar approach was used to construct the dummy variable *presence of private label*. PBM began marketing powdered infant formula in 1997, although the quarter in which it began marketing in any one market area differed across market areas and as of the third quarter of 2000 it was not marketing in every market area in the sample.

Annual data on a market area's number of discount stores and population were obtained from Metro Market Studies; together these data were used to calculate the number of discount stores per 100,000 population, a variable that is labeled *discount stores* in appendix D.² The discount stores that were counted include such retailers as Wal-Mart, K-Mart, and Target who carry infant formula; chains such as T.J. Maxx whose product line does not include infant formula were not counted. Quarterly data were obtained by interpolation.

The variable *supermarket concentration* is the standard Herfindahl-Hirschman Index used to measure concentration within a market area among its supermarket chains. The more closely a market area resembles one that is served by a single, monopoly supermarket chain, the greater is the value of the market area's *supermarket concentration*. Each market area had a single *supermarket concentration* value (based on 2000 data) that was used to represent the entire study period. Market share information was gathered from *MarketScope2000*.³

² The annual publication *Discount Store Distribution Analysis and Guide* is written and published by Metro Market Studies of Tucson, AZ.

³ For each market area, if two or more chains with different names were owned by a single entity, the market shares for those chains were aggregated. The *Supermarket Concentration* utilized all available share information, including independents.

The real (constant purchasing power) median *household income* in a market area was obtained using three series. A State's nominal median income was treated as the nominal median income of its market area(s). State nominal median income figures on an annual basis were obtained from the U.S. Department of Commerce's Census Bureau; quarterly figures were obtained by interpolation. The nominal median income data were adjusted for both intertemporal and cross-sectional differences in the purchasing power of a dollar. The data were adjusted for inflation using the quarterly CPI-U. The data were adjusted for a local market area's cost-of-living relative to the U.S. average cost-of-living using the American Chambers of Commerce Research Association (ACCRA) cost-of-living index.⁴

It is noted that the geographic boundaries of the InfoScan market areas, which are reflected in the determination of the independent variables of supermarket retail prices, do not coincide exactly with the geographic boundaries used for determining the variables *discount stores*, *supermarket concentration*, and *household income*.

The variable *poverty rate* represents the percentage of people in poverty, based on annual data from the U.S. Census Bureau (Department of Commerce). A State's poverty rate was used to proxy for the poverty rate in all market areas within the State. Annual figures were interpolated on a quarterly basis.

For each of a manufacturer's four infant formula products, the variable *wholesale price* was obtained from wholesale price lists published by the manufacturer. Price changes that took effect after the middle of a quarter were treated as taking place at the start of the following quarter. A product's inflation-adjusted wholesale price was obtained by deflating the nominal price series by the CPI-U.

A simple linear time *trend* was created, which changes by one unit each quarter.

⁴ For most market areas, the 1995 value of the ACCRA index was used, although for a few market areas, the value of the index was taken from a different period under the assumption that the market area's cost of living relative to the U.S. average was steady over time.