INTEGRATED MEASURES OF SALES, MERCHANDISING, AND DISTRIBUTION

John D.C. Little

M.I.T. Sloan School of Management, Cambridge, MA 02142, U.S.A.

Sloan Working Paper # 3997
July 1998

ABSTRACT

Managers track marketing performance with measures of sales, distribution, and merchandising. These can be calculated at different levels of aggregation with respect to geographic areas, time periods, and products. To be most useful, performance measures should have parallel and consistent meanings across the different levels. Starting from the decomposition of sales into base and incremental volume as provided by data suppliers at the underlying level of item-store-week, we define a set of performance measures that fit together into a simple model. It is shown that these definitions permit consistent aggregation into analogous measures and models at higher levels. An example drawing on Ocean Spray Cranberries data illustrates the advantages of the measures for comparing marketing performance across levels of aggregation.

Key words: scanner data, distribution, merchandising, aggregation

Forthcoming in *International Journal of Research in Marketing*

1. Introduction

The most frequently quoted numbers from scanner databases are aggregate statistics. Although bar code readers generate huge amounts of detailed data, the sheer quantity of numbers forces a search for meaningful summaries. These should express, in simple and sensible ways, what is happening to a brand or product line over broad geographic regions and time periods. Such aggregates provide top line information about market status and customer response and so assist marketers in managing their brands. The same measures are used by senior managers and others, inside and outside a company, who need to stay abreast of market trends.

Summary measures answer such questions as: "How is a brand or product line performing overall? Have our products increased in national distribution over the past six months? How effective has our trade support been?" Answers often trigger deeper studies that drill down in the databases for further insight. Because the most commonly used numbers throughout a company are summaries, their construction merits scrutiny for consistency and clarity.

Marketing scientists have exploited scanner data to study many important phenomena, but have done relatively little with aggregation issues. Exceptions are Christen, Gupta, Porter, Staelin, and Wittink (1997), Gupta, Chintagunta, Kaul, and Wittink (1996), Foekens, Leeflang, and Wittink (1994), Link (1995), and Allenby and Rossi (1991). These authors investigate biases introduced by applying econometric techniques to data that has been assembled by aggregating from individuals to stores or stores to markets.

The focus here, however, is quite different and not econometric at all. Most of the aggregate information produced by the suppliers of scanner data is more analogous to accounting reports than to econometric modeling. Relatively simple arithmetic sums and weighted averages serve to accumulate data over products, geographical areas, and time periods. The measures commonly produced have the important advantage that they are relatively transparent in their calculation and interpretation. Such numbers are viewed by thousands of marketers and salespeople daily.

1.1 Scanner Databases.

To give perspective, we briefly sketch the size and scope of the retail store data collected by the two principal suppliers of scanner data to the consumer packaged goods industry in the U.S. and Europe, namely, Information Resources, Inc. (IRI) and the A.C.Nielsen Company (Nielsen). Using the U.S. as a benchmark, the universe of stores being monitored consists of approximately 31,000 supermarkets, 38,000 drug stores, 6,400 mass merchandise outlets, and 140,000 convenience stores. There are eight million Universal Product Codes (UPCs) in IRI's data dictionary, each representing an identifiably different product. Many of them are inactive at any given time, but over 3 million show product movement during a year. The samples of stores in the IRI and Nielsen services are quite large. For example, in mid-1997 IRI's InfoScan contained about 4,515 stores, consisting of approximately 3,050 supermarkets, 600 drug stores, 290 mass merchandisers, and 575 convenience stores. Typically, the raw data arrives from the stores as weekly sales totals in dollars and units by UPC. A store might stock 30,000 UPCs. In addition, stores in the sample are monitored weekly for displays and features. Each store generates from six to ten permanently retained measures per week per UPC carried. These basic measures are analyzed in various ways and combined with information from other sources into over 800 different measures available to clients.

To swell the data further, information is increasingly collected from all stores in a retail chain, not just a sample. Such *census* data is replacing samples for many applications. In particular, for a manufacturer's sales force dealing with retail chains, it helps two-way communication between manufacturer and retailer when both work from the same, complete data. In 1997 IRI collected information from over 24,000 stores and its online database exceeded two terrabytes.

In other countries the number of products, although not quite as large, has a similar order of magnitude. A chief difference is that the size of the country determines the number of stores. Different countries also have different mixes of store types and may have different product codes. In Europe products are identified by European Article Numbers (EAN), in Japan by Japanese Article Numbers (JAN).

Individual store data for, say, a half dozen measures over two years by item and week for a major category runs into billions of records and is prohibitively time consuming to work with except for special studies. Therefore, Nielsen and IRI perform a first step of aggregating data across stores into markets or other useful groupings before delivery. In the case of sample data, such aggregation involves projection.

To bring some order into the large number of UPCs, products are partitioned into categories. Although over 820 categories are in common use and many UPCs are not active, this still leaves many products per category. For example, a large category such as cereals or carbonated soft drinks would contain several tens of thousands of individual products. Further organization is required. Products within a category are therefore grouped hierarchically, for example, by manufacturer, brand, sub-brand, size, package, and flavor, the scope and order being specified by the client requesting the data. Each level in such a hierarchy is a candidate for aggregation.

1.2 Aggregation

Marketers want aggregations over at least three dimensions: geography, time, and product. Projections and aggregations often require weighting store data by the size of the store. In such cases size is usually measured in terms of a store's all commodity volume (ACV), which would be expressed, for example, in millions of dollars per year. This procedure is standard and will be assumed, although other methods could be used and sometimes are necessary for special purposes. Aggregation over time usually means adding (or averaging) over weeks. We shall argue that this should be done in certain cases where it is not done now. Aggregation over products is a less obvious process and will be a major concern of the paper.

Sales. Sales themselves are usually easy to add up, even when they include items of different sizes. Most manufacturers define a physical unit, such as kilograms or liters, or other "equivalent volume." Equivalent units make different sizes comparable. We assume that this step has been taken for all products being aggregated so that sales of items across a product line or category add to meaningful totals. Adding sales across time periods is also straightforward.

More troublesome, however, are distribution and merchandising. Both IRI's Infoscan service and Nielsen's Scantrack produce widely disseminated "non-additive" measures for distribution and merchandising. As an example, Table 1 shows a week's data for a product line, *Total Ocean Spray Cranberry Drinks 64 Oz*, and eight items that make it up. Shown are the distribution and display measures for the individual items and, as would commonly be reported, for the line as a whole. The measures are called non-additive because their values for aggregates are not simple sums or averages of their components.

Distribution. Distribution for an individual item is the percentage of stores with non-zero sales of the item, where each store is weighted by its size as measured by all commodity volume (ACV). Therefore distribution has units of %ACV. In Table 1, for example, 98.2 %ACV distribution for Ocean Spray Cran-Cocktail 64 Oz means that this product was sold in a set of stores that constituted 98.2% of the all commodity volume in the geographic area under consideration.

Merchandising. Merchandising is a generic name for promotional activity conducted by a store to increase sales. Data companies report four mutually exclusive types of merchandising: (1) *display only*, (2) *feature only*, (3) *display and feature together*, and (4) *unsupported price cuts*. Several of these can be subdivided further, if desired. Features are advertisements in newspapers or store flyers. Unsupported price cuts refer to temporary shelf price reductions in the absence of special display or feature.

For illustrative purposes, consider *display-only*. Measures of merchandising activity can be defined analogously to those for distribution. For example, in Table 1, the value of 4.6 %ACV with display-only for Ocean Spray Cran-Cocktail 64 Oz means that this product had a display (but no feature) in a set of stores that represent 4.6% of all commodity volume in the region under consideration.

Questions arise, however, about the rules for aggregating distribution and merchandising. For example, the display of a product line like Total Ocean Spray Cranberry Drink 64 Oz is commonly determined as follows: The line is considered to be *on display* in a store for the week if *any* of its component items are on display. Similarly, the product line is *in distribution* in a store if *any* of its components are in distribution.

Thus, in Table 1, we see that Total Ocean Spray Cranberry Drink 64 Oz has 99.7% distribution even though one of its individual items is in stores representing only 8.8% of ACV. Similarly, the display-only measure for the aggregate is much larger than any of its individual items.

An analogous rule is commonly used for time aggregation. Take, for example, a 12-week period. The 64 ounce bottle of Cran-Grape is said to be on display in a store during the period if it was on display during any of the 12 weeks. (Strict enforcement of this definition requires excessive computation and so data companies use an approximation, but the results are similar.)

PRODUCT	DISTRIBUTION (% ACV)	DISPLAY-ONLY (%ACV)
OCEAN SPRAY - CRAN-COCKTAIL 64 OZ BTL	98.2	4.6
OCEAN SPRAY - CRAN-APPLE 64 OZ BTL	94.8	5.0
OCEAN SPRAY - CRAN-CHERRY 64 OZ BTL	44.8	1.2
OCEAN SPRAY - CRAN-GRAPE 64 OZ BTL	93.6	3.5
OCEAN SPRAY - CRAN-RASPBERRY 64 OZ BTL	91.0	2.4
OCEAN SPRAY - CRAN-STRAWBERRY 64 OZ BTL	80.6	3.7
OCEAN SPRAY - CRAN-KIWI 64 OZ BTL	8.8	1.7
OCEAN SPRAY - CRAN-CURRANT 64 OZ BTL	71.1	1.7
TOTAL OCEAN SPRAY CRANBERRY DRINK 64 OZ.	99.7	14.0

Table 1. With conventional measures a product line like Total Ocean Spray is said to be in distribution if any of its component products are. Similar rules apply to merchandising conditions like display-only. (IRI data for Total US in the week ending 13 July 97.)

Although these measures tell something about what is happening, they quickly saturate to nearly 100% for large aggregates and, a more serious problem, are poor indicators of the depth and strength of distribution and merchandising within the product line and over time.

Lift. A key response indicator for merchandising is *lift*. Lift is a measure of short term merchandising effectiveness, defined as the fractional increase in sales volume attributable to a merchandising activity during the week that it takes place. Typically, an activity such as a display, feature, or temporary price cut produces an immediate jump in sales that continues for the week or two during which the merchandising is running. Lift therefore measures short run effects. It can be calculated for any type of merchandising. Whereas measures of activity tell how much merchandising has taken place, lift tells how effective it has been.

A serious shortcoming of reporting lift to measure merchandising effectiveness and %ACV to measure merchandising activity is that the pair do not combine in any simple way to yield the incremental volume attributed to the merchandising. This problem, along with the difficulties in interpreting aggregates as previously described, motivate our search for measures that fit together more harmoniously.

1.3 Approach to constructing consistent aggregate measures

Are there better ways to construct overall measures? Ideally, aggregate measures of merchandising and distribution should summarize activity and permit comparisons across *markets* (or other groups of stores), *time periods*, and *products* (both individual items and product lines). In addition, such measures should connect controllable actions (merchandising) to market response (sales). Finally, they should have consistent meanings across levels of aggregation.

Nielsen and IRI generate *baselines* by performing time series analyses on data for each store and item (UPC). These partition sales for each week into a *base volume* and an *incremental volume* attributed to merchandising activity. IRI's methods have been described by Abraham and Lodish (1993). Numerical algorithms process data for each store and item to estimate a baseline of sales that would have occurred without the merchandising. The underlying idea is to identify weeks without merchandising and draw a smooth line through their sales over time, adjusting for various overall market effects. This provides the baseline. Then incremental sales equal actual sales minus baseline sales. We take this decomposition as given and construct aggregate measures that employ base and incremental volume at the item-store-week level as "raw" data.

The measures to be developed may be thought of as variables in deterministic models that classify the sales volume of a product consistently into quantities that account for all base and incremental sales. Such an accounting-like approach represents quite a different paradigm from econometric modeling. Each approach has advantages.

Baseline methods pick out the large first-order effects of merchandising. These are easily visible as bumps in plots of sales over time, since the lift due to merchandising is often several times base sales. However, current baseline methods do not, for example, measure the complex influences that the merchandising of one product may have on related products. Such effects can often be estimated by econometric modeling in studies designed for the purpose. Econometric techniques also permit the examination of other issues not addressed in standard IRI and Nielsen measures. Foekens, Leeflang, and Wittink (1994) present a type of multiplicative model that is frequently employed. Such models can often yield valuable information about complex market responses.

Econometric and baseline approaches have different characteristics. An econometric model will ordinarily estimate parameter values from many pieces of data. For example, a single display effectiveness parameter would normally be estimated from data containing many merchandising events. In contrast, Nielsen and IRI baseline methods estimate a separate display effectiveness for each event. Since events differ in their execution, the individual event analysis often provides important information. Furthermore many marketers like to have values of sales and incremental sales that add up to total actual sales in an accounting-like way. Econometric models yield estimates that ordinarily do not do this. On the other hand, econometric models are good at measuring more subtle phenomena than can be discerned by the current baseline approaches. Marketers often want both kinds of information and so use both types of analysis.

The focus here is on improving the consistency and meaningfulness of accounting-like measures of sales, distribution, and merchandising and so takes as given the time-series decomposition into base and incremental volume. If methods are devised that produce better calculations of these inputs, they can be substituted in what follows.

Our plan of attack is as follows:

(1) Adopt an 'atomic' unit of data collection: the item-store-week. An item is defined by its finest grain of identification, normally, its UPC, EAN, JAN, or other code.

(2) Construct a model to represent base and incremental sales for an item at the store level in a given week. The model connects incremental sales to merchandising activity by means of the lift for each type of merchandising.

(3) Sum the item-store-week model in different ways, applying sampling projection factors if needed, to calculate higher levels of aggregation.

(4) In each case, define arithmetic combinations of data elements to create summary measures of sales, merchandising, lift, and distribution for aggregate units, such as markets, multi-week periods, brands, and other useful collections. Aggregate measures are found that are consistent in the sense that they (a) satisfy a model of the same form as the item-store-week model and (b) account for all the observed sales. Because of this consistency and because the variables fit together in the model, we call them an *integrated* set.

(5) Compare the integrated measures with those commonly used for each's ability to interpret sales and marketing performance.

2. Analytic Development

The simple ideas that we use repeatedly are: (1) sales volume = base volume + incremental volume, (2) incremental volume = (base volume) x (merchandising activity) x (lift due to merchandising activity), and (3) base volume = (base volume per unit of distribution) x (distribution).

2.1 Store level data and model

The atomic unit of observed data is the item-store-week. Let

s_{urt} = sales volume of item (UPC) u in retail store r in week t (equivalent units)

$\{k\}$ = a set of mutually exclusive and exhaustive types of merchandising.

m^k_{urt} = 1 if merchandising k is present for item u in store r during week t,

= 0 otherwise.

$R = \{r\}$ = set of stores in sample.

a_r = all commodity volume (ACV) of store r (millions of dollars/year).

$a_R = \Sigma_{r \epsilon R}\, a_r$ = ACV of sample R (millions of dollars/year).

d_{urt} = 1 if item u is in distribution in store r during week t,

= 0 otherwise.

Data suppliers calculate and provide

s_{0urt} = base volume of u in store r in week t, i.e., sales in the absence of merchandising (equivalent units).

The preceding measures are "raw" data for the developments to follow. Two further quantities are immediately calculable from them, incremental volume and lift.

$w_{urt} = s_{urt} - s_{0urt}$ = incremental volume of item u in store r during week t (equivalent units).

$w^k_{urt} = w_{urt}\, m^k_{urt}$ = incremental volume attributable to merchandising type k (equivalent units).

$\ell^k_{urt} = w^k_{urt}/s_{0urt}\, m^k_{urt}$ = lift of merchandising type k on item u in store r during week t (dimensionless).

The last two relationships are valid because the types of merchandising are mutually exclusive and exhaustive and because merchandising activity is a 0-1 indicator variable at the store level.

The preceding definitions lead algebraically to the following model of sales response to merchandising:

$$s_{urt} \;=\; s_{0urt}\,[1 + \Sigma_k\, \ell^k_{urt}\, m^k_{urt}] \tag{1a}$$

Model (1a) is a tautological arithmetic relationship among the measures just defined. In other words, it is a logically consistent way to express sales volume in terms of base volume, lift, and merchandising activity at the level of item-store-week.

Anticipating subsequent developments in the paper, we introduce definitions of *average number of items in distribution* and *average base sales per item in distribution*, even though these measures are rather uninteresting for an item-store-week. First observe that

d_{urt} = average number of items in distribution for item u store r and week t.

Let

$v_{0urt} = s_{0urt}/d_{urt}$ = average base sales per item in distribution for item u, store r, and week t (equivalent units).

Then (1a) can be rewritten explicitly to include distribution

$$s_{urt} \;=\; v_{0urt}\, d_{urt}\,[1 + \Sigma_k\, \ell^k_{urt}\, m^k_{urt}] \tag{1b}$$

Equation (1b) is our basic model. We now seek analogous relations at higher levels of aggregation.

2.2 Store Groups and Markets

As previously noted, IRI and Nielsen deliver data to their clients already aggregated over stores. For concreteness, we describe this as an aggregation into markets, although the methods apply to any group of stores, for example, all stores in a specific grocery chain.

Standard measures include:

s_{ut} = sales volume of item u in the market for week t (equivalent units),

d_{ut} = distribution of u, the fraction of market ACV selling item u in week t (dimensionless).

s_{0ut} = base volume of u in the market during week t (equivalent units).

These measures are calculated by projecting individual store data to the store group as follows:

$$s_{ut} = (a/a_R) \Sigma_{r\epsilon R}\, s_{urt} \qquad (2a)$$

$$d_{ut} = \Sigma_{r\epsilon R}\, (a_r/a_R)\, d_{urt} \qquad (2b)$$

$$s_{0ut} = (a/a_R) \Sigma_{r\epsilon R}\, s_{0urt} \qquad (2c)$$

where a = ACV of the market (millions of dollars/year). For census data, $a = a_R$.

In these formulas distribution is a dimensionless number with a range of 0 to 1. Standard practice is to report such measures on a scale of 0 to 100 percentage points. Thus, $d_{ut} = .73$ is often described as 73 points of ACV distribution. Notice, furthermore, that we can interpret d_{ut} in two important, additional ways:

d_{ut} = probability that item u is in distribution in week t at a randomly selected store in the market, when selection is made proportional to store size.

= expected (average) number of products in distribution in the market for item u in week t.

Finally, analogous to the case of individual stores, define

$v_{0ut} = s_{0ut}/d_{ut}$ = average base volume per product in distribution for item u and week t (equivalent units).

This measure tells how well an item sells in the stores that carry it during a week when there is no merchandising. As such, v_{0ut} reflects an inherent strength of the product with consumers in those stores. Packaged goods marketers often use v_{0ut} to rank items within a product line. A product that scores highly will be a strong candidate for marketing effort to increase the number of stores stocking it, unless the stores that already do so are very different from a marketing point of view (e.g., represent a special ethnic or economic group).

Next, define aggregate measures of incremental volume, merchandising activity and lift. Let

$$w^k{}_{ut} = (a/a_R)\, \Sigma_r\, w^k{}_{urt} \qquad (3a)$$

$$m^k{}_{ut} = \Sigma_r\, s_{0urt}\, m^k{}_{urt} \,/\, \Sigma_r\, s_{0urt} \qquad (3b)$$

= fraction of base volume of item u that has merchandising type k during week t (dimensionless).

$$\ell^k_{ut} = \Sigma_r w^k_{urt} / \Sigma_r s_{0urt} m^k_{urt} \qquad (3c)$$

= ratio of total incremental volume of an item u having merchandising k during t to base volume with k in the same set of item-store-weeks (dimensionless).

To see that these are the consistent measures we seek, first note that, if R comprises just a single store, r, (3b) and (3c) reduce to m_{kurt} and ℓ^k_{urt}, as they should. Next, substituting (1) into (2a) and using the definitions (2c), (3), and some algebra yields

$$s_{ut} = s_{0ut} [1 + \Sigma_k \ \ell^k_{ut} m^k_{ut}] \qquad (4a)$$

which is the aggregate form of (1a). Finally, substituting the appropriate definition,

$$s_{ut} = v_{0ut} d_{ut} [1 + \Sigma_k \ \ell^k_{ut} m^k_{ut}] \qquad (4b)$$

Thus the aggregate measures satisfy the same model as the store data but in aggregate variables. Furthermore, by construction, the aggregate measures of sales exactly account for all the sales of their lower level constituents.

2.3 Product Lines

Consider next the task of aggregating across items to form a brand or product line, holding week constant. Let

$P = \{u\} =$ a set of items that form a product line P.

$$s_{Pt} = \Sigma_{u \epsilon P} s_{ut} \qquad (5a)$$

= sales volume of product line P in the market during week t (equivalent units).

$$s_{0Pt} = \Sigma_{u \epsilon P} s_{0ut} \qquad (5b)$$

= base volume of product line P in the market during week t (equivalent units).

$$d_{Pt} = \Sigma_{u \epsilon P} d_{0ut} \qquad (5c)$$

= average number of products (items) of line P in distribution in the market during week t.

$$v_{0Pt} = s_{0t}/d_{Pt} \qquad (5d)$$

= average base volume per product in distribution for product line P in the market during week t (equivalent units).

Analogously to the previous cases, define measures of merchandising activity and lift for the product line:

$$w^k_{Pt} = \Sigma_{u\epsilon P}\, w^k_{ut} \qquad (6a)$$

$$m^k_{Pt} = \Sigma_{u\epsilon P}\, s_{0ut}\, m^k_{ut} \,/\, \Sigma_{u\epsilon P}\, s_{0ut} \qquad (6b)$$

= fraction of base volume of product line P that has merchandising type k during week t

$$\ell^k_{Pt} = \Sigma_{u\epsilon P}\, w^k_{ut} \,/\, \Sigma_{u\epsilon P}\, s_{0ut}\, m^k_{ut} \qquad (6c)$$

= ratio of total incremental volume of P with merchandising k during t to base volume with k in the same set of item-store-weeks.

To see that these are consistent measures, note that if P contains just a single product, u, (6a) and (6b) reduce to m^k_{ut} and ℓ^k_{ut}, as they should. Further, substituting (4a) into (5a) and using the definitions (5) and (6) yields

$$s_{Pt} = s_{0Pt}\, [1 + \Sigma_k\, \ell^k_{Pt}\, m^k_{Pt}] \qquad (7a)$$

which is the aggregate form of (4). Finally, substituting (5d),

$$s_{Pt} = v_{0Pt}\, d_{Pt}\, [1 + \Sigma_k\, \ell^k_{Pt}\, m^k_{Pt}] \qquad (7b)$$

Thus aggregate measures over a product line satisfy the same response model as the store and market data but in aggregate product line variables that account for all lower level sales.

2.4 Time aggregates

The basic rhythm of a supermarket is weekly. Many customers are in the habit of shopping once a week. Feature advertising appears weekly. Special displays are set up each week and, although some may last longer, a weekly cycle paces the store's planning. For these reasons the data companies collect and report most store data with a week as the time unit.

Although merchandising within a given week may affect subsequent weeks because some customers stock up on a product, this phenomenon is obscured in store data by normal household purchase cycles, which, for most products, are several weeks or more. Although one group of customers may buy a product in a given week, a different set is likely to buy it in the following one. In addition there is a tendency for retailers to separate promotions of the same product with periods of other activity.

As a result, the first order effect of merchandising is to produce a separate response in each week that it takes place. This means that an important measure of the aggregate effect over a multiple week period will be a simple sum of the activities in individual weeks.

Starting with measures at the item-market-week level, define:

$T = \{t\} =$ a set of W time periods forming a multi-week period T.

$$s_{uT} = \Sigma_{t \epsilon T}\, s_{ut} \tag{8a}$$

= sales volume of item u in the market over time period T (equivalent units).

$$s_{0uT} = \Sigma_{t \epsilon T}\, s_{0ut} \tag{8b}$$

= base volume of item u in the market over time period T (equivalent units).

$$d_{0uT} = \Sigma_{t \epsilon T}\, d_{0ut}/W \tag{8c}$$

= average distribution of u in the market during period T.

$$v_{0uT} = s_{0uT}/d_{0uT} \tag{8d}$$

= average base volume per product in distribution for item u in the market during period T (equivalent units).

Notice that we have chosen to define distribution, when it is cumulated over weeks, as an average rather than a sum. This is fairly arbitrary and is done because an average seems more intuitively meaningful than a sum as a measure for distribution.

In a manner similar to the previous cases, define measures of merchandising activity and lift for the time period:

$$m^k_{uT} = \Sigma_{t \epsilon T}\, s_{0ut}\, m^k_{ut} \,/\, \Sigma_{t \epsilon T}\, s_{0ut} \tag{9a}$$

= fraction of base volume of item u that has merchandising type k during period T

$$\ell^k_{uT} = \Sigma_{t \epsilon T}\, w^k_{ut} \,/\, \Sigma_{t \epsilon T}\, s_{0ut}\, m^k_{ut} \tag{9b}$$

= ratio of total incremental volume with merchandising k during T to base volume with k in the same set of item-store-weeks.

To see that these are consistent measures, note that, for T comprising a single week, (9a) and

(9b) reduce to m^k_{ut} and ℓ^k_{ut}, as desired. Further, substituting (4a) into (8a) and using the definitions (8) and (9) yields

$$s_{uT} = s_{0uT} [1 + \Sigma_k \ell^k_{uT} m^k_{uT}] \tag{10a}$$

which is the aggregate form of (4a). Finally, substituting (8d),

$$s_{uT} = v_{0uT} d_{uT} [1 + \Sigma_k \ell^k_{uT} m^k_{uT}] \tag{10b}$$

Thus the aggregate measures for multi-week period T satisfy the same model as the store and market data but in multi-week variables that account for all single week sales.

2.5 General Case

The algebra of aggregation can be generalized and separated from the particular application. Given (1) A = {a} = a set of indices for lower level variables that are to be aggregated into higher level variables to be indexed by A, and (2) nonnegative variables s_a, s_{0a}, m^k_a, and w^k_a such that

$$s_a = s_{0a} + \Sigma_k w^k_a ;$$

then, letting

$$\ell^k_a = w^k_a / s_{0a} m^k_a \text{ if } m^k_a > 0;$$
$$= 0 \qquad \text{otherwise}$$

it follows that

$$s_a = s_{0a} [1 + \Sigma_k \ell^k_a m^k_a] .$$

Define higher level measures in terms of the lower level variables

$$s_A = \Sigma_{a\epsilon A} s_a$$

$$s_{0A} = \Sigma_{a\epsilon A} s_{0a}$$

$$w^k_A = \Sigma_{a\epsilon A} w^k_a$$

$$m^k_A = \Sigma_{a\epsilon A} s_{0a} m^k_a / \Sigma_{a\epsilon A} s_{0a}$$

$$\ell^k_A = \Sigma_{a\epsilon A} w^k_a / \Sigma_{a\epsilon A} s_{0a} m^k_a .$$

It follows algebraically that, for these higher level variables,

$$s_A = s_{0A} [1 + \Sigma_k \ell^k_A m^k_A]$$

and also that the new set of variables s_A, s_{0A}, m^k_A, and w^k_A are all non-negative and satisfy

$$s_A = s_{0A} + \Sigma_k w^k_A .$$

This last expression shows that the aggregation process is recursive. Therefore, we can take any level of product aggregation and create a higher one in a consistent manner. For example, a set of product lines can be aggregated into a product "portfolio". Furthermore, having aggregated on one dimension, we can aggregate on another and obtain parallel formulas.

3. Empirical comparisons

Tables 2 and 3 compare integrated measures with conventional ones for the Ocean Spray products considered previously. Data is for Total US Food sales in the period ending 13 July 1997, as reported by IRI.

Distribution. Tables 2a and 2b compare distribution measures for one-week and 52-week periods.

One-week. Our integrated measure is *average number of items in distribution*, which is to be compared with the conventional *%ACV with distribution.* For an individual item, the two are identical except for the decimal point. They are just two interpretations of the same number, as may be seen in Table 2a. The difference comes at aggregation, exemplified here by the product line, Total Cranberry Drinks. The integrated *average number of items in distribution* is simply the sum over the items and equals 5.83. This seems like a useful summary of how widely available in stores are the 8 items of the product line. Such a measure can be watched from week to week as marketing tactics unfold.

The conventional, non-additive measure, which considers the product line to be present whenever any of its items are present, has the value 99.7%. This is a high value, crowding 100%, that does not seem to add much information not already contained in the individual items, since one them has 98.2% distribution. Of course, the conventional measure will sometimes be interesting and useful since it describes a different aspect of the data. However, if a choice must be made between the two measures in the interests of brevity, the integrated version seems to tell more. The conventional measure could remain in the background, obtainable by drill down.

52-weeks. Time aggregation illustrates the utility of the integrated measures as norms. We can make such statements as, "In the current week, the average numbers of items in distribution for Cran Cherry and Cran Currant are ahead of their 52-week values but Cran Strawberry is behind. Furthermore, the product line, Total Cranberry Drinks, is ahead." By contrast, the conventional measure, *%ACV with distribution*, because of the way it is defined, can only increase with aggregation and so cannot perform the analogous role. For example, the *%ACV with distribution* of Total Cranberry Drinks for the current week is 99.7%, which is behind (rather than ahead of) the 52-week value of 99.9%. These numbers also show how saturation toward 100% reduces the meaningfulness of the conventional measure.

Distribution

Product (64 oz bottles)	1 week — Conventional: % ACV with Distribution (%)	1 week — Integrated: Average Number of Items	52 weeks — Conventional: % ACV with Distribution (%)	52 weeks — Integrated: Average Number of Items
OCEAN SPRAY CRAN- COCKTAIL	98.2	0.982	98.9	0.984
OCEAN SPRAY CRAN- APPLE	94.8	0.948	96.9	0.956
OCEAN SPRAY CRAN- CHERRY	44.8	0.448	46.1	0.428
OCEAN SPRAY CRAN- GRAPE	93.6	0.936	94.4	0.934
OCEAN SPRAY CRAN- RASPBERRY	91.0	0.910	93.8	0.922
OCEAN SPRAY CRAN- STRAWBERRY	80.6	0.806	86.7	0.842
OCEAN SPRAY CRAN- KIWI	8.8	0.088	24.7	0.089
OCEAN SPRAY CRAN- CURRANT	71.1	0.711	82.9	0.575
OCEAN SPRAY --TOTAL CRAN DRINK	99.7	5.829	99.9	5.730
	(a)		(b)	

Merchandising (display-only)

Product (64 oz bottles)	1 week — Conventional: % ACV with Display - only (%)	1 week — Integrated: % Base Volume with Display - only (%)	52 weeks — Conventional: % ACV with Display - only (%)	52 weeks — Integrated: % Base Volume with Display - only (%)
OCEAN SPRAY CRAN- COCKTAIL	4.6	6.00	13.6	6.78
OCEAN SPRAY CRAN- APPLE	5.0	5.82	12.5	5.91
OCEAN SPRAY CRAN- CHERRY	1.2	3.21	2.6	2.41
OCEAN SPRAY CRAN- GRAPE	2.8	3.50	7.8	4.05
OCEAN SPRAY CRAN- RASPBERRY	3.5	4.67	8.1	4.47
OCEAN SPRAY CRAN- STRAWBERRY	2.4	3.42	4.9	2.92
OCEAN SPRAY CRAN- KIWI	3.7	44.78	8.3	25.40
OCEAN SPRAY CRAN- CURRANT	1.7	2.72	5.2	3.50
OCEAN SPRAY --TOTAL CRAN DRINK	14.0	5.13	30.7	5.31
	(c)		(d)	

Table 2. Conventional and integrated measures of distribution and display-only for one and 52 weeks. Conventional measures can only increase with 52-week aggregation whereas the 52-week integrated measure can act as a norm for judging performance in an individual week.

Merchandising. The integrated measure for merchandising activity is quite different from the conventional one. Its construction achieves consistency and also makes it a relevant norm across levels of aggregation. Tables 2c and 2d show one-week and 52-week comparisons.

One-week. Looking first at the conventional measure, *%ACV with display-only*, for the product line, Total Cranberry Drink, we find its value to be 14.0%. This is several times that for any individual item. Therefore, 14.0% does not provide a norm against which the individual items can be judged. Neither is it related in any simple way to the incremental sales generated by display-only conditions. In contrast, the integrated measure, *% of base volume with display-only*, has a value 5.13% for the product line. This provides a reference point for comparing the performance of the individual items. For example, only 3.50% of the base volume of Cran Grape has display-only during this week. Furthermore, because of the way 5.13% for the product line is constructed, it combines with other data in a consistent way to generate the total incremental sales attributed to display-only.

52-week. A similar situation holds for the 52-week time aggregate. The conventional aggregate neither provides a standard for single week performance nor tells how display-only worked to generate extra sales for the item. As an illustration, the 52-week value for Cran-Cocktail is 13.6% for the conventional measure. Although this may be compared with its single week value of 4.6%, the comparison seems less useful than the corresponding one for the integrated measure. In the latter case, the 52-week value of 6.78% can act as a standard and we see that the latest single week value of 6.00% is noticeably below the 52-week average. Furthermore, the integrated measure combines with lift factors and base volume to play back the incremental sales attributed to display-only conditions.

Consistency of the Integrated Measures. Table 3 illustrates how the integrated measures fit together. The numbers across each row consistently follow model (7b), i.e.:

base volume = (base volume per item in distribution) x (number of items in distribution)

incremental volume from display-only = (base volume) x (lift for display-only) x (fraction of base volume having display-only)

These relationships hold both for individual items and for the product line as a whole. The same calculation would also apply to the other types of merchandising not shown. Sales volume is then the sum of base volume and the incremental volumes for the several types of merchandising. Table 3a shows 1-week numbers and Table 3b 52-week aggregates. The two tables are similar, except that the volume numbers are roughly 52 times larger in the 52-week case.

The integrated measures fit together in this way at all levels of aggregation across products, geographies, and time periods. We know of no set composed entirely of conventional measures that has the same property. Within the integrated set, three of the measures are, in fact, conventional: *base volume*, *incremental volume*, and *lift*. Another, *distribution*, is the same at the item-week level. The measure that is most non-standard is

Product (64 oz bottles)	Average Number of Items	Base Volume per Item (eq. units)	Base Volume (eq. units)	Lift Due to Display Only	% Base Volume with Display Only (%)	Incremental Volume Due to Display Only (eq. units)
OCEAN SPRAY CRAN- COCKTAIL	0.982	123,648	121,422	0.73	6.00	5,351
OCEAN SPRAY CRAN- APPLE	0.948	55,448	52,565	0.69	5.82	2,106
OCEAN SPRAY CRAN- CHERRY	0.448	32,734	14,665	0.95	3.21	447
OCEAN SPRAY CRAN- GRAPE	0.936	58,201	54,476	0.68	3.50	1,301
OCEAN SPRAY CRAN- RASPBERRY	0.910	54,329	49,439	0.97	4.67	2,229
OCEAN SPRAY CRAN- STRAWBERRY	0.806	32,931	26,542	0.88	3.42	801
OCEAN SPRAY CRAN- KIWI	0.088	22,511	1,981	1.60	44.78	1,417
OCEAN SPRAY CRAN- CURRANT	0.711	20,444	14,536	1.99	2.72	785
OCEAN SPRAY --TOTAL CRAN DRINK	5.829	57,579	335,628	0.84	5.13	14,437

(a) 1 week

Product (64 oz bottles)	Average Number of Items	Base Volume per Item (eq. units)	Base Volume (eq. units)	Lift Due to Display Only	% Base Volume with Display Only (%)	Incremental Volume Due to Display Only (eq. units)
OCEAN SPRAY CRAN- COCKTAIL	0.984	7,079,248	6,965,980	1.10	6.78	517,921
OCEAN SPRAY CRAN- APPLE	0.956	3,28,598	3,182,140	1.08	5.91	204,069
OCEAN SPRAY CRAN- CHERRY	0.428	1,865,729	798,532	1.44	2.41	27,639
OCEAN SPRAY CRAN- GRAPE	0.934	3,389,390	3,165,690	.99	4.05	126,856
OCEAN SPRAY CRAN- RASPBERRY	0.922	3,156,963	2,910,720	1.36	4.47	176,613
OCEAN SPRAY CRAN- STRAWBERRY	0.842	1,943,812	1,636,690	1.15	2.92	55,087
OCEAN SPRAY CRAN- KIWI	0.189	1,231,652	109,617	2.02	25.40	56,283
OCEAN SPRAY CRAN- CURRANT	0.575	1,588,941	913,641	1.30	3.50	41,729
OCEAN SPRAY --TOTAL CRAN DRINK	5.730	3,435,080	19,683,010	1.15	5.31	1,206,197

(b) 52 weeks

Table 3. Integrated measures are consistent across aggregations over products and times periods. In each row incremental volume due to display-only is the product of base volume with display-only, percent of base volume with display-only, and lift due to display-only.

merchandising activity. This variable knits the others together and permits the simple decomposition of sales volume into its components at all levels of aggregation.

4. Conclusions

In order to understand and discuss market performance, managers and analysts need summary measures of sales, distribution and merchandising. Aggregation must be possible over time, geographic areas, and products. Good measures of merchandising and distribution are ones that are intuitively meaningful and fit together in consistent ways.

We have presented a class of integrated measures that start with information routinely provided by data suppliers: the decomposition of sales into base and incremental volume, as attributed to mutually exclusive and exhaustive types of merchandising. The decomposition is expressed in a deterministic, accounting-like model at the item-store-week level. Each of its variables is then aggregated analytically to store groups, product lines, and multi-week periods. The model retains its algebraic form at each level of aggregation.

The advantages of the integrated measures are (1) *interpretability*: in certain instances the integrated measures seem more meaningful than conventional ones, (2) *utility as norms*: higher levels of aggregation provide reference values for judging performance at lower levels, (3) *transparency*: a simple mental model of how the measures fit together is the actual model and works consistently at all levels of aggregation, and (4) *extensibility*: new aggregations over products, store groups, and time periods can be created analogously without difficulty.

It is not suggested that the underlying model captures all merchandising issues of interest and, in fact, examples of missing phenomena have been cited. Differences in the measures show up across all dimensions, since individual stores and items respond in different ways. Marketers will interpret these differences in light of their knowledge about events taking place in the market and will draw inferences about the effectiveness of their programs. Differences and anomalies will trigger further analysis. Thus aggregate measures will often be the starting point for drilling deeper into the immense detail offered by scanner data and will stimulate other types of statistical studies.

ACKNOWLEDGEMENT

The author thanks Gordon Armstrong and Eugenie Murray-Brown of Ocean Spray Cranberries, Inc. for data, helpful discussions, and good suggestions.

REFERENCES

Abraham, M.M. and L.M. Lodish (1993), "An implemented system for improving promotion productivity using store scanner data," *Marketing Science*, 12, 248-269.

Allenby, G.M. and P.E. Rossi (1991), "There is no aggregation bias: why macro logit models work," *Journal of Business and Economic Statistics*, 9, 1-11.

Christen, M., S. Gupta, J.C. Porter, R. Staelin, and D.R. Wittink (1997), "Using market-level data to understand promotion effects in a nonlinear model," *Journal of Marketing Research*, 34, 322-346.

Foekens, E.W., P.S.H. Leeflang, and D.R. Wittink (1994), "A comparison and exploration of the forecasting accuracy of nonlinear models at different levels of aggregation," *International Journal of Forecasting*, 10, 245-261.

Gupta, S., P. Chintagunta, A. Kaul, and D.R. Wittink (1996), "Do household scanner panels provide representative inferences from brand choices: a comparison with store data," *Journal of Marketing Research*, 33, 383-398.

IRI (1997), *InfoScan Measure Selection Guide*, Information Resources, Inc., Chicago.

Link, R. (1995), "Are aggregate scanner data models biased?" *Journal of Advertising Research*, 35, RC8-12, (Sept-Oct).