

## A Model to Improve the Estimation of Baseline Retail Sales

Kurt Jetta

*TABS Group, Shelton, Connecticut*

Erick W. Rengifo \*

*Fordham University, New York, New York*

---

### Abstract

This paper develops more accurate and robust baseline sales estimates (sales in the absence of price promotion) using a dynamic linear model (DLM) enhanced with a multiple structural change model (MSCM). We first discuss the value of utilizing aggregated (chain-level) vs. disaggregated (store-level) point-of-sale (POS) data to estimate baseline sales and to measure promotional effectiveness. We then present the practical advantage of the DLM-MSCM modeling approach using aggregated data, and we propose two tests to determine the superiority of a particular baseline estimate: the minimization of weekly sales volatility and the existence of no correlation with promotional activities in these estimates. Finally, we test this new baseline against the industry standard ones on the two measures of performance. Our tests find the DLM-MSCM baseline sales to be superior to the existing log-linear models by reducing the weekly baseline sales volatility by over 80% and by being uncorrelated to promotional activities.

*Keywords:* dynamic linear models, multiple structural change model, consumer packaged goods, marketing, sales, promotions, baseline sales

JEL Classification codes: M30, M31, C01, C11

---

In the United States, the consumer packaged goods industry (CPG) accounts for over \$500 billion in annual retail sales according to ACNielsen and at least twice that worldwide. It is well documented that retailer price promotions (defined as a temporary reduction in retailer price for a specific set of products for a specific period of time) account for the largest share of CPG firms' marketing budget (Cannondale, 2007), and that percentage has grown consistently over time. Industry estimates peg the amount of annual spending on retailer price promotions at about \$50-75 billion annually in the United States (about 15-20% of factory sales according to Accenture) and over \$100 billion worldwide<sup>1</sup>.

The CPG industry has one of the most extensive information infrastructures of any industry. Most U.S. retail outlets are able to track the sales of virtually every product that is sold in the store with the use of scanners. These scanners can read the Universal Product Code (UPC) on each product. The UPC is matched to information that describes dozens of characteristics about the product: manufacturer, brand, product type, flavor, weight, count size, and so on. The in-store scanner data are augmented by household-level scanning data (panel data) from over 100,000 U.S. households. The panel data are used to generate even more granular information on the consumer purchasing process. Two major firms, Information Resources (IRI) and

ACNielsen, have created a multibillion dollar industry by collecting much of this information and selling it to manufacturers, retailers, and other interested parties.

Armed with this information, manufacturers, retailers, and academics have developed extraordinarily detailed models to measure the effectiveness of promotions and other marketing tactics like consumer advertising, price changes, and public relations. A common denominator of all these models is that in order to determine the effectiveness of a given marketing tactic, one needs to determine first the benchmark baseline sales level (i.e., the expected sales in the absence of a particular marketing variable like price promotion). It is worth noting that the baseline sales are simply the counterfactual of sales activity in the hypothetical case of no promotions for a period of time.

In this paper, we propose a new model to estimate baseline sales and compare it to the two models that are considered to be the industry and academic standard: Scan\*Pro (Wittink, Addona, Hawkes & Porter 1988) and PromotionScan (Abraham & Lodish, 1993). Scan\*Pro and PromotionScan were developed in conjunction with ACNielsen and IRI. Both are log-linear models that provide estimates of baseline sales and sales response as a function of specific retailer promotional tactics such as price discounts, feature ads, and displays. According to Bucklin, and Gupta (1999) and to Hanssens, Parsons, and Schultz (2000), both models are fundamentally similar.

While there have been no formal academic challenges to the validity of the model, there are obvious data limitations in terms of quality and availability. The use of disaggregated data could potentially have measurement errors. Moreover, CPG practitioners and consultants generally recognize that the baseline sales generated by these models are flawed in that they yield “phantom” spikes.<sup>2</sup> They show increases in baseline sales exactly concurrent with promotional activity when the expectation is that no such spike should occur. In Figure 1, we show one example of regular phantom spikes. Later in this paper, we explain why baseline sales are supposed to be relatively stable estimates over time and why baseline estimates should be uncorrelated with promotional activity.

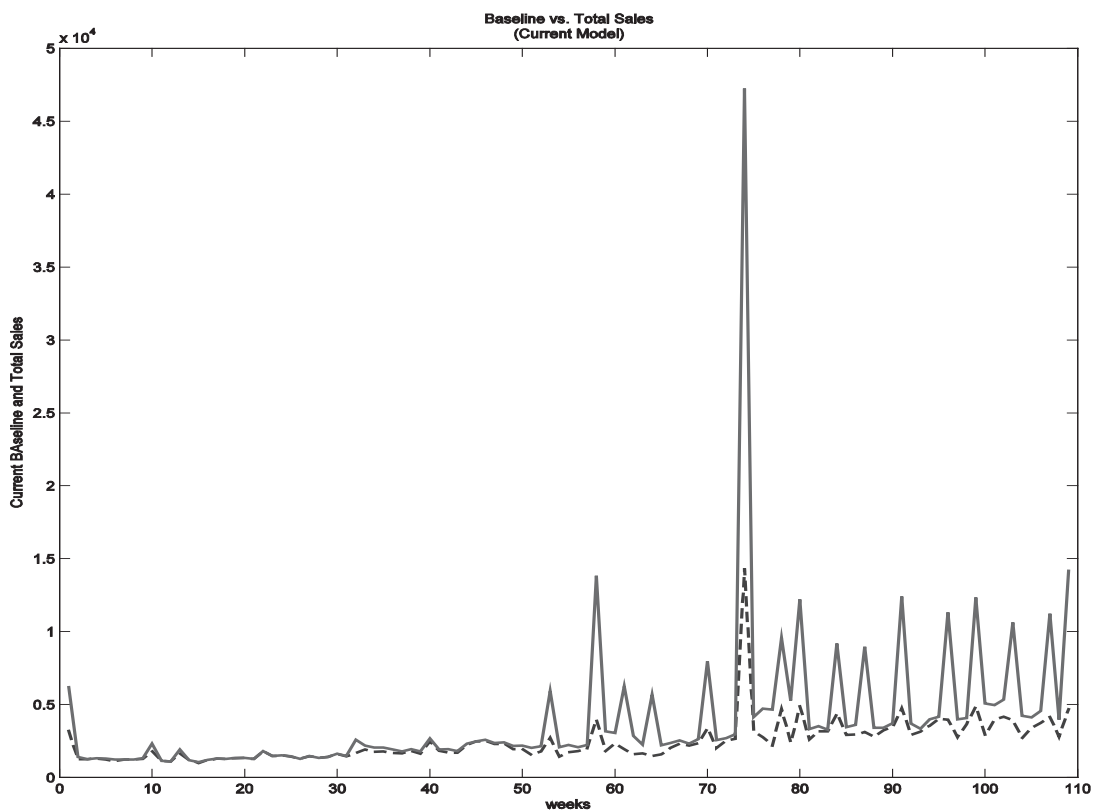


Figure 1. The sales (solid line) and the actual estimate of the baseline sales generated using Scan\*Pro (dashed line) for an adult personal care product.

This paper has two main contributions to make to the literature: a methodological and a practical one. The methodological contribution of this paper is the introduction of a method that leads to a more robust, less costly, and more accurate estimate of baseline sales. This contribution to the existing literature is important because any measure of promotion performance depends directly on the baseline sales estimate. Flaws in the existing baseline model understate the incremental sales impact of price promotions and overstate the overall level of baseline sales. We implement the new baseline model using two econometric techniques: the dynamic linear model (DLM) based on Ataman, Mela, and Van Heerde (2007) which we improved with the use of the inclusion of a dummy variable to flag promotional activity (Jetta, 2008) as well as the multiple structural change model (MSCM) of Bai, and Perron (2003).

On the empirical side, this paper proposes to make several important contributions to the field. First, a better baseline estimate will help managers make better spending decisions on their promotion budgets. Second, the baseline method will be extendable to a broader section of retailers to include club stores and category killers (like Home Depot and Staples). Third, the baseline model can reach thousands of small to mid-sized manufacturers that cannot afford the significant investment required to purchase baseline estimates from the major syndicated data suppliers. Fourth, the DLM-MSCM is new in marketing applications, and this paper adds these useful tools in econometric analysis to the body of knowledge in marketing research.

The rest of this paper is structured as follows. The next section contains a discussion of the use of aggregated versus disaggregated data, a description of the actual baseline model and its flaws, and a presentation of some desirable properties that any baseline sales should have. A presentation of the econometric techniques used in constructing the new baseline sales follows. The next section shows the empirical results obtained, and the final section contains conclusions and future research ideas.

## **The Baseline Sales**

Fundamental to the analysis of any marketing tactic is the concept of baseline sales. In order to determine whether a causal variable generated some effect on sales, the analyst needs a reasonable estimate of what sales would have been without the existence of the causal variable (the counterfactual). Therefore, a baseline sale is defined as an estimate of sales in the absence of specific promotional activity for a specific product and for a determined period of time.

In this section, we present a brief discussion about the use of aggregated (chain or market-level) versus disaggregated (store-level) data, and we point out the reasons one should prefer to work with aggregated data. Later, we present the actual baseline model and its flaws. Finally, we introduce two tests that a desirable baseline sales model should satisfy.

### ***Aggregated vs. disaggregated data for baseline sales modeling***

Starting with Wittink et al. (1988) and Abraham, and Lodish (1993), the use of the disaggregated data standard was established, and the research paradigm that only disaggregated data should be used for marketing model estimation persists to this day. The research on the issue (Christen, Gupta, Porter, Staelin, & Wittink, 1997; Foekens, Leeflang, & Wittink, 1994; Van Heerde, Leeflang, & Wittink, 2002) maintained that there was a significant risk of parameter estimation bias by using aggregated data in nonlinear models. (Scan\*Pro and PromotionScan are log linear.) This bias would imply, for example, that the estimate of the percentage increase in sales from display activity using aggregated data might be overstated.

These authors concede, however, that the use of aggregated data holds several appealing properties in the areas of cost, availability, modeling flexibility, processing time, and overall compliance and acceptance by practitioners. Christen et al. (1997) suggested a debiasing procedure that can be used for market-level data. They mentioned nothing, however, about the more important issue of chain-level aggregation. Therefore, there has been almost no use of the debiasing procedure in other literature, and the conventional wisdom remains that disaggregated data are always optimal for modeling.

A further discussion of the practical shortcomings of disaggregated data is in order. Most importantly, disaggregated data are not aligned with the standard of management accountability, as are aggregated data, either at the chain or market level. The research tools have not been developed to predict and explain results at that level. While a role clearly exists for the use of disaggregated data, the initial discovery process should occur at the group (aggregated) level to determine the total effects of programs in which managers are most

interested. Unfortunately, the disaggregation paradigm means that most promotional researchers have overlooked aggregated effects entirely.

The second major limitation of disaggregated data is in the area of cost and availability. Currently, very few parties have access to such data. In the academic world, there are just a few databases—University of Chicago Dominick’s Database, the Stanford Basket Dataset, and a database recently released by IRI—with this information. This static universe of data available for research limits the opportunity to check the robustness of existing results and to test new hypotheses. From both a commercial and academic standpoint, there are significant processing constraints to modeling store-level data unless there is a costly computer hardware investment in processing the massive database. This constraint is the reason most econometric models in the literature are built on databases of 30 stores or fewer. Even then, the DLM used by Ataman et al. (2007) took several weeks to process a 30-store database. This precludes any meaningful analysis of a 6000+ store chain like CVS for all but the most powerful of hardware and software.

From a modeling standpoint, disaggregated data have only a marginal advantage over aggregated data. Van Heerde et al. (2002) stated that the primary reason for using disaggregated data is to ensure that there is no estimation bias of parameters when the independent variables are heterogeneous. Accordingly, as long as marketing activity is implemented homogeneously, there is very little risk of biased estimation. Furthermore, even with heterogeneous marketing activity, the magnitude of the bias depends on the percentage of stores promoted: the bias decreases as the percentage of stores promoted becomes larger (Van Heerde et al., 2002). From a practical standpoint, most chains execute advertisements and price reductions homogeneously. That means every store within a chain receives the same marketing stimulus. For example, for the adult personal care category studied in this paper, 86% of the 34008 observations with some level of feature activity had all commodity volume (ACV) percentages of 80% or more. This observation is particularly valid for the United States and Canada.<sup>3</sup>

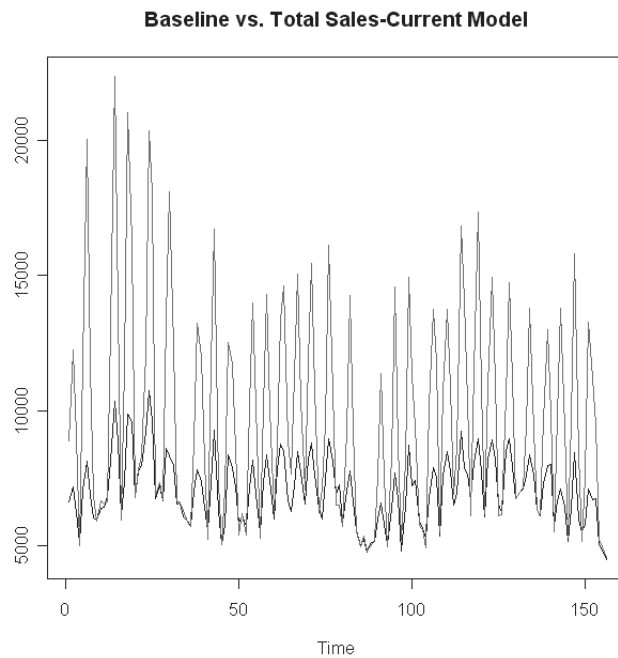
In summary, disaggregated data contain severe practical and quantitative limitations that preclude them from being the sole or even primary data source for marketing research. Particularly given the homogeneity of most marketing stimuli, aggregated chain-level data should be appropriate for most applications. However, we left for future research the use of our model with disaggregated data.

### ***The Existing Baseline***

Bucklin, and Gupta (1999) pointed out that many practitioners believe the baseline measure to be an actual number, when, in fact, it is a modeled measure. A modeled measure presents difficulties in determining whether the measure is accurate, because there are never any actual data to validate it against. The first benchmark of measurement is intuition and judgment. In other words, does the baseline appear to measure sales in the absence of promotion? In discussions with dozens of practitioners over the years, many expect baseline sales to be relatively stable, similar to sales trends they see during sustained periods without sales promotion.

A baseline sales estimate can range in sophistication from the back-of-the-envelope guess to complex, econometric models that require a great deal of data input and computer processing power. In the CPG industry, the two industry standard models are Scan\*Pro (Wittink et al., 1988) and PromotionScan (Abraham & Lodish, 1993). Both are log-linear models that are fundamentally similar (Bucklin & Gupta, 1999; Hanssens et al., 2000). Both models regress the log of unit sales against log price and dummy variables for other promotional effects such as display or feature activity.

One can observe that the resulting baseline sales from these models exhibit much variation and high correlation with promotional activity; that is, we can observe that the baseline sales exhibit phantom spikes concurrent with promotional activity. If we assume that these baseline sales are correct, it would be better for sales managers not to do *any* promotional activities because just by not doing them, the sales would naturally increase. Figure 2 shows an example of this lack of stability and high correlation.



*Figure 2.* The sales (solid line) and the actual estimate of the baseline sales generated using PromotionScan (dashed line) for a frozen food product.

Van Heerde et al. (2002) set out the original version of the Scan\*Pro model. This model is nonlinear, hence the authors' concern about parameter bias. Taking the natural log of this model provides the opportunity to conduct ordinary least squares regression on the data. The authors imply that the model is simple, as it was the first step in an evolutionary model building process. Later, they expanded this model by introducing dynamics either through time-varying parameters or via the inclusion of leads and lags. Their model also incorporates cross-brand promotional effects from numerous brands. In the latest published version of this model, the dependent variable can be a function of hundreds of independent variables once all the cross-brand and timing variables are considered.

Van Heerde et al. (2002) presented a baseline sales model based on their extended model where they included four weeks of leads and lags to the original model in order to “accommodate the illusive post-promotion dip.”<sup>24</sup> Their estimated baseline sales also show several sharp dips and spikes. Within a 10-week period, the baseline deviates by about 12% around the median level for the period. They contended that this dynamic effect is “consistent with expectation” because promotional lifts tend to reduce post-period baseline sales. However, they did not mention any explanation for the phantom spikes observed in their baseline sales.

Other models have been used in the literature to calculate baseline sales, but none has been offered as a formal alternative to the industry-standard log-linear models. Nijs, Marnik, Dekimpe, Steenkamp, and Hanssens (2001) and Pauwels, Hanssens, and Siddarth (2002) both developed baseline sales models using a Vector Autoregressive with Exogenous variables (VARx) model where baseline sales are implied from the sales forecast for time  $t$ . Then, they used impulse response functions for each promotion to gauge the incremental effects for periods  $t$ ,  $t+1$ ,  $t+2$ , and so on. Ataman et al. (2007) used a dynamic linear model (DLM) to estimate baseline sales in a model for decomposing the effects of various marketing mix elements in new brands. Both of these models are confined to specific academic applications.

### ***Validity Standards for Baseline Sales***

After analyzing sustained periods of no promotional activity for certain brands, we see that baseline sales are relatively stable over time, in the absence of any major structural shift in sales (e.g., increased retail distribution) or seasonal growth. Moreover, there is no reason to expect high correlation between the expected sales in the absence of promotions with specific promotional activity except in cases where manufacturers consistently execute other marketing programs not tracked in the data-gathering process (e.g., Free Standing

Insert (FSI) or single-week TV advertising) during those promotion weeks. However, instances where manufacturers are able to do this *consistently* are rare. Based on this, we argue that a superior baseline sales model implies that promotions and baseline sales should be contemporaneously uncorrelated.

In this vein, we first show that weeks without promotions do, in fact, have less sales variability than those weeks with promotional activity for a specific retail chain. Specifically, we test the following hypothesis:

*H0*: Weekly sales variability during promotion weeks is greater than or equal to the weekly sales variability of weeks without promotions (after controlling for structural shifts and seasonality).

$$\sigma_{L(PROMO)} \geq \sigma_{L(NONPROMO)}$$

where,  $\sigma_{L(NONPROMO)}$  ( $\sigma_{L(PROMO)}$ ) is the standard deviation of the natural log differences of sales during weeks without promotion (with promotion). It is expected that the null hypothesis will not be rejected.

We then test the hypothesis to determine which baseline model (log-linear industry standards vs. DLM-MSCM) has lower sales variability:<sup>5</sup>

*H1*: The existing baseline model has volatility that is greater than or equal to the volatility of the proposed baseline model.

$$\sigma_{L(Be)} \geq \sigma_{L(Bn)}$$

where,  $\sigma_{L(Be)}$  ( $\sigma_{L(Bn)}$ ) is the standard deviation of the natural log differences of sales for the existing (new) model baseline model. It is expected that the null hypothesis will not be rejected.

Finally, we see whether the baseline sales present phantom spikes concurrent with promotional activity, that is, whether the baseline sales are correlated with promotional activity. Formally, we test the following two hypothesis:

*H2*: The current baseline sales estimate is contemporaneously correlated with promotional activity.

$$\text{Correlation}(B_{ert}, \varphi_{rt}) \neq 0$$

where,  $B_{ert}$  is the baseline sales estimate for the existing model (e) in retailer (r) at time (t) and  $\varphi_{rt}$  is the promotion activity of retailer (r) at time (t). It is expected that the null hypothesis will not be rejected.

*H3*: The new baseline sales estimate is contemporaneously correlated with promotional activity.

$$\text{Correlation}(B_{nrt}, \varphi_{rt}) \neq 0$$

where,  $B_{nrt}$  is the baseline sales estimate for the new model (n) in retailer (r) at time (t) and  $\varphi_{rt}$  is the promotion activity of retailer (r) at time (t). It is expected that the null hypothesis will be rejected.

Note that in each test, we are ensuring that the equality is always placed in the null hypothesis. This increases the power of the test, that is, the likelihood of rejecting a null hypothesis that should be rejected.

## Econometric Implementation

This section presents the econometric techniques used to create the new baseline sales. The first subsection introduces the dynamic linear model (DLM) used by Ataman et al. (2007) as well as a method to determine the promotional dummy using the technique proposed by Jetta (2008). The second subsection presents the technique developed by Bai, and Perron (2003) to detect multiple structural changes (MSCM), and the third subsection describes briefly our implementation methodology.

## Dynamic Linear Models

DLM is a modeling technique pioneered by West, Harrison, & Migon (1985) to address time series problems. The technique uses a Bayesian approach to provide probability estimates to each observation in a time series.

From a marketing modeling perspective, Ataman et al. (2007) offered the following advantages of DLM. First, it has greater statistical efficiency with parameter evolution and explanation in one step. Second, there is no need for pre-steps (like unit root testing) or assumptions on the distribution of error terms. This gives DLM an advantage over Kalman Filter, which requires the assumption of normally distributed error terms. Third, parameters update immediately as new data become available. Fourth, missing data are accommodated relatively easily by using estimates from prior periods for imputation in the missing data. Fifth, the technique allows for subjective information. Prior expectations can be overridden to accommodate anomalies in the data. Sixth, the model accommodates longitudinal as well as cross-sectional heterogeneity.

The disadvantages of DLM involve issues related to the implementation of the model rather than any statistical weakness. Specifically, DLMs can be extremely processing intensive, where models can take days or even weeks to run. Another minor disadvantage is that few software packages include the DLM.<sup>6</sup>

The Ataman et al. (2007) DLM model is as follows:

$$Sales_t = \alpha_t + \beta_t * PIndex_t + v_t \quad (1)$$

where Equation 1 is referred to as the observation equation and,  $PIndex_t$  is a Price Index at time t, and

$$\alpha_t = \lambda * \alpha_{t-1} + \omega_t \quad (2)$$

$$\beta_t = \beta_{t-1} + \varepsilon_t \quad (3)$$

Equations 2 and 3 are known as state equations. Observed sales for a given week are a function of a dynamic baseline component  $\alpha$  at time t and a dynamic promotion response evolution defined by  $\beta$  at time t. It is evident that this is a more parsimonious model than the log-linear models used in the industry. This model leaves open the possibility of additional exogenous variables, but as a first-generation model it is much simpler.

In general, Equations 1 and 2 can be written as

$$Sales_t = [1 \quad PIndex_t] \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} + v_t$$

or

$$y_t = F_t' \theta_t + v_t \quad (4)$$

and

$$\begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \omega_{1t} \\ \omega_{2t} \end{bmatrix} \quad (5)$$

The core DLM equations are then:

$$(y_t | \theta_t) \sim N(F_t' \theta_t, V_t) \quad (6)$$

$$(\theta_t | \theta_{t-1}) \sim N(G_t' \theta_{t-1}, W_t) \quad (7)$$

$$(\theta_0 | I_0) \sim N(m_0, C_0) \quad (8)$$

where  $I_0$  is the initial prior information at time 0, including  $F_t$ ,  $G_t$ ,  $V_t$  and  $W_t$ . Moreover, at any future time  $t$  the available information set is  $I_t = \{Y_t, I_{t-1}\}$ . The posterior for some mean  $m_{t-1}$  and variance matrix  $C_{t-1}$  is given by

$$(\theta_{t-1} | I_{t-1}) \sim N(m_{t-1}, C_{t-1}) \quad (9)$$

It can be shown that the prior at time  $t$  is

$$(\theta_t | I_{t-1}) \sim N(a_t, R_t) \quad (10)$$

where,  $a_t = G_t m_{t-1}$  and  $R_t = G_t C_{t-1} G_t' + W_t$ . With this, the one step ahead forecast is given by:

$$(y_t | I_{t-1}) \sim N(f_t, Q_t) \quad (11)$$

with  $f_t = F_t' a_t$  and  $Q_t = F_t' R_t F_t + V_t$ . Thus, the posterior at time  $t$  is

$$(\theta_t | I_t) \sim N(m_t, C_t) \quad (12)$$

where  $m_t = a_t + A_t e_t$ ,  $C_t = R_t - A_t Q_t A_t'$ ,  $A_t = R_t F_t Q_t^{-1}$  and,  $e_t = y_t - f_t$

While Equations 1 to 3 provide a good starting point for a general baseline model, the inclusion of the price index variable presents some potential problems. The price is prone to measurement error: some retailers deduct promotion discounts off the entire shopping order and do not assign them to a specific product. Additionally, many promotional vehicles such as in-ad coupons, rebates, and loyalty card discounts have a history of tracking difficulties. From a retailer perspective, some stores may lower prices on a local basis for competitive reasons without the typical promotional support like shelf tags. Other retailers have nontraditional methods of handling buy one/get one free (BOGO) consumer deals. Often, both items will be scanned at full revenue with some other code denoting the BOGO offer. Although there is no evidence of systematic problems with price tracking in the syndicated data (except for a few isolated retailers), with so many potential shortcomings, even for own brand promotion response, using the price index as an exogenous variable does not appear to be optimal.

Based on the previous mentioned shortcomings of the price index, we use instead a dummy variable ( $\phi$ ) approach to account for promotional weeks (i.e., the variable takes on the value of 1 if it is a promotion week and 0 otherwise). An additional potential problem arises here: sometimes, a clear identifier of a promotional activity during a week is not readily available.<sup>7</sup> Under this circumstance, we use the technique proposed by Jetta (2008) to determine the values that the dummy should take on. The model runs through several ordinary least squares iterations to refine the fit of the model by minimizing the model's standard error or maximizing its coefficient of determination. This calibration exercise flags any observation week where there is an abnormal deviation in weekly sales change or where the absolute sales level is significantly above the overall average.<sup>8</sup>

An additional advantage of Jetta's (2008) procedure is that whereas the price index variable and other explanatory variables have problems with respect to acquisition costs and availability beyond CPG products carried in food/drug/mass (ex Wal-Mart), the dummy variable technique has no such problems. It eliminates the data acquisition costs just by using weekly unit sales, also eliminating all other causal inputs. Additionally, the measure is available to all retailers where scanner data are available.

Accordingly, the DLM model that we are going to use has the following observation equation:

$$Sales_t = \alpha_t + \beta_t \phi_t + \gamma_t I_t + v_t \quad (13)$$

where,  $\phi_t$  is the promotion dummy and  $I_t$  represents another category-specific dummy such as a seasonality dummy. From here, the sales are a function of the dynamic baseline sales ( $\alpha_t$ ), promotional activity the ( $\beta_t$ ) and other explanatory variables ( $I_t$ ). By construction, our baseline sales figure captures the unit sales in the absence of promotions. The respective state equations are as follows:

$$\alpha_t = \lambda_t \alpha_{t-1} + \omega_{1t} \quad (14)$$

$$\beta_t = \delta_t \beta_{t-1} + \omega_{2t} \quad (15)$$

$$\gamma_t = \rho_t \gamma_{t-1} + \omega_{3t} \quad (16)$$

Equation 14 presents the baseline evolution lift parameter. In Equation 13, we replace the price index as an explanatory variable with Promo dummy variable  $\phi$ . Equation 15 shows the dynamics for the lift parameter ( $\beta$ ) and permits us to test for promotional wear-out effects over time. Equation 16 could potentially include category-specific dummies to control for seasonality, for example.

### ***Multiple structural change model***

A weakness of the DLM introduced previously is that it is not able to capture structural changes in sales. Structural changes occur when the demand for a given product increases as a result of factors not directly related to promotions. From a practical standpoint, these structural changes are usually related to major increases or decreases in item-level distribution for a promoted brand. The main distinction between a structural change and a promotion shift is that the first implies a permanent movement while the latter implies a temporary one.<sup>9</sup>

In order to capture this behavior, we complement the DLM with a technique proposed by Bai, and Perron (2003) that allows us to capture multiple structural changes that could potentially be present. A detailed discussion of this model is presented in Appendix B.

### ***Methodology***

We estimate our new baseline sales using a two-step procedure: First, we determine the structural changes in the data (if any) following Bai, and Perron (2003), and second, we implement piece-wise the DLM to model the unit sales as a function of a constant (the baseline), a dummy variable that captures promotional activity, and another dummy variable to capture the seasonality of the series. We apply piece-wise DLM to the regimes that were found in the first stage of the methodology.

### ***Empirical Application***

In this section, we compute our new baseline sales using the econometric techniques described above. For this application, we use aggregated data for adult personal care products and frozen foods. We present our new baseline sales estimate and test it in the framework of the hypotheses.

### ***Data description***

We use aggregated data at the retail chain level for two categories: adult personal care product and frozen foods. The data were gathered at weekly intervals from each of the major syndicated data suppliers (one category from IRI and one category from ACNielsen). The data span 109 weeks (from 4/30/2006 to 5/25/2008) and 125 weeks (from 8/27/2005 to 1/21/2008) for adult personal care products and frozen foods, respectively. We present the basic analytical grouping as a data class, which is all weekly observations for a specific category, within a specific retailer for a specific brand. We have 312 data classes in adult personal care products and 247 data classes in frozen foods.

### ***Stationarity of the data***

Even though the DLM model does not require stationarity of the time series data, it is a necessary condition for the multiple structural change model of Bai, and Perron (2003). Each data class was tested for both level and trend stationarity. In total, there were 559 data classes across two categories (312 adult personal care products and 247 frozen foods). We conduct the Augmented Dickey-Fuller Unit Root test on each data class to

test for stationarity in levels and considering a deterministic trend. The unit root results show that retail sales data represent a trend stationary process, as 95.4% of the data classes did not have a unit root.<sup>10</sup>

## The New Baseline Sales

In this section, we present and test the new baseline sales estimates and compare them with the existing ones. The tests are performed under the hypotheses stated in Section 2. That is, an improved baseline estimate should exhibit low week-to-week sales variability and no contemporaneous correlation with promotional activity.

First, we test  $H_0$  where the null hypothesis is that weekly sales variability for high promotion weeks is greater than or equal to the weekly sales variability for low promotion weeks. To test this hypothesis, each weekly observation within each data class was divided into one of four quartiles based on the percentage units on any promotion (PUAP) that week: class 1 (0-25%), class 2 (25-50%), class 3 (50-75%), and class 4 (75%-100%). The PUAP is a measure directly pulled from the data supplier with no other manipulation to the figures. Table 1 provides the analysis of variance results by quartile.

Table 1  
Standard Deviation of Unit Sales

Chain	(% Unit on promotion)	Adult personal care			Frozen foods		
		Obs.	Standard deviation	<i>p</i> -value	Obs.	Standard deviation	<i>p</i> -value
Quartile I	0-25%	15,478	9,849.0		18,685	1,972.7	
Quartile II	25-50%	3,755	11,554.0	0.00	4,774	2,566.3	0.01
Quartile III	50-75%	2,935	14,937.1	0.00	3,871	2,532.9	0.01
Quartile IV	75-100%	5,244	15,157.6	0.00	3,545	2,968.4	0.00

*Note.* This table presents the standard deviation of unit sales according to the percent units on any promotion (PUAP) during a given week (broken into quartiles from lowest to highest values) and the *p*-values of the *F*-test for equality of variances. The null hypothesis, in all cases, is that the variances of any quartile relative to Quartile I are equal. The results show that the standard deviations of all quartiles respect to Quartile 1 are significantly different at the 5% confidence level.

Table 1 shows that in all the cases and for both categories (adult personal care products and frozen foods), in order to operationalize  $H_0$ , we use the *F*-test for equality of variances. We compare the variance of each quartile with respect to the first one (the case with no promotional activity). The results in Table 1 show that the null hypothesis of equality of variance is rejected at a 5% significance level for all the cases. Moreover, we can see that the variance of promotional weeks is larger than the one in weeks without promotions; the variance of unit sales during low-promotion weeks is significantly different (and low) compared to highly promoted weeks (class 3 and 4). This result goes in hand with our empirical observation and is the basis for testing hypothesis  $H_1$ .

In order to test hypothesis  $H_1$ , we compare the variance of our new baseline sales with the existing industry standard models. The null hypothesis ( $H_1$ ) is that the volatility of the industry standard baseline sales model is greater than or equal to the volatility of the new model. Figure 3 shows the results of our comparison using a histogram. This figure depicts the difference of the standard deviation in the log difference of our proposed baseline sales minus the standard deviation in the log difference of the existing baseline sales. We can see a dramatic reduction (over 80%, on average) in the variability in weekly baseline sales estimate using our new baseline model.

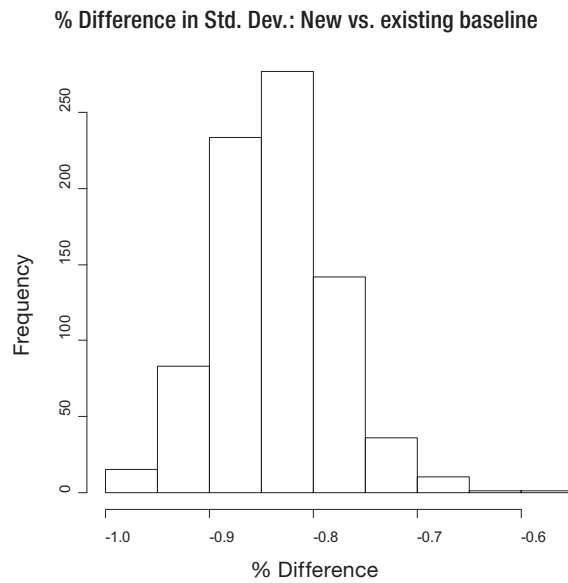


Figure 3. Histogram presenting the frequencies of the difference of variability between our new baseline sales and the existing baseline sales model.

We also performed a test based on the four data classes described before, that is, we tested the null hypothesis of equal variances of the new baseline sales and the Scan\*Pro/PromotionScan, one per data class. In these cases, 100% of the variances were significantly different at the 5% level. Moreover, this difference is significantly larger in heavily promoted weeks (Class 4).<sup>11</sup> In conclusion, we do not reject the null hypothesis that the volatility of the existing baseline is greater than or equal to the new baseline.

The final two tests measure the existence of contemporaneous correlation between promotional activity and baseline sales. By inspection, it appears that the major source of the variability in the existing baseline measures is due to this correlation, which we have referred to as the phantom spike. *H2* will test whether this phantom spike exists on a consistent basis for all data classes.

We perform this test using the natural log differences in the weekly baseline sales for all weeks. We code each weekly observation as follows: first week of promotion and other promotion week. We make this distinction because the first week of promotion in a multiweek promotion usually exhibits a much higher sales increase (lift) than subsequent weeks. The test will capture whether there may be a specific class of promotional observations where this correlation exists. Of course, this will not be an issue in single-week promotions.

Table 2 presents the average, maximum, and minimum simple correlation between each individual baseline sales observation and its respective promotional activity; that is, when  $\phi=1$ . Moreover, this table also presents the pooled *p*-value of the *t* test under the null of significant correlation. We constructed the pool by using all available data (regardless of product) in a given group and performed the *t* test for significant correlation.<sup>12</sup>

Table 2  
*Correlation Between the Industry Baseline Sales and Promotional Activity*

Week	Scan*Pro		PromotionScan	
	Av.corr/max/min	<i>p</i> -value	Av.corr/max/min	<i>p</i> -value
First-week promo	0.81/0.96/0.75	0.41	0.85/0.99/0.72	0.46
Other-week promo	0.74/0.88/0.69	0.37	0.79/0.92/0.74	0.42

*Note.* This table presents the average, the maximum, and minimum correlation between the industry baseline sales (Scan\*Pro and PromotionScan) and promotional activity. The data were divided into two groups: first-week of promotion and other promotion week. This table also presents the pooled *p*-value of the *t* test under the null of significant correlation. The results show that there is significant correlation between the existing baseline models (Scan\*Pro and PromotionScan) and promotional activity for both groups.

Observing the results in Table 2, we can appreciate the high level of significant positive correlation between both the existing baseline sales and the promotional activities. This feature is not desirable because there is no reason to expect an increase in sales in the absence of any promotional activity. This high, positive and significant correlation of the existing baseline sales with promotion activities is evident by observing the existence of spikes and dips throughout, as shown before in Figure 1.

We next performed the same test for the DLM-MSMC model. From Table 3, it is clear that that there is no significant correlation between our new baseline sales and promotional activity at the 5% significance level for both categories under analysis.<sup>13</sup>

Table 3  
Correlation Between Baseline Sales and Promotional Activity for Products

Week	Adult personal care		Frozen foods	
	Av.corr/max/min	<i>p</i> -value	Av.corr/max/min	<i>p</i> -value
First-week promo	0.12/0.15/0.05	0.03	0.09/0.14/0.07	0.01
Other-week promo	0.08/0.12/0.05	0.02	0.06/0.11/0.03	0.00

*Note.* This table presents the average, the maximum, and the minimum correlation between the baseline sales and promotional activity for each individual product in each of the categories that we analyze (adult personal care and frozen foods). The data were divided into two groups: first-week of promotion and other promotion week. This table also presents the pooled *p*-value of the *t* test under the null of significant correlation. The results show that there is no significant correlation between our new baseline and promotional activity for both groups (first-week promo and other-week promo).

Based on the results presented here, we do not reject *H2* that the correlation between promotional activity and baseline sales for the existing model is different from zero. Conversely, we reject *H3* that the correlation between promotional activity and the new baseline estimate is different from zero.

Furthermore, in Figure 4, we can observe (a) actual sales versus new baseline (upper left), (b) existing baseline versus new baseline (upper right), (c) actual sales versus existing baseline (lower left), and (d) actual sales versus fitted sales from DLM. This figure represents one data class from the frozen foods category.

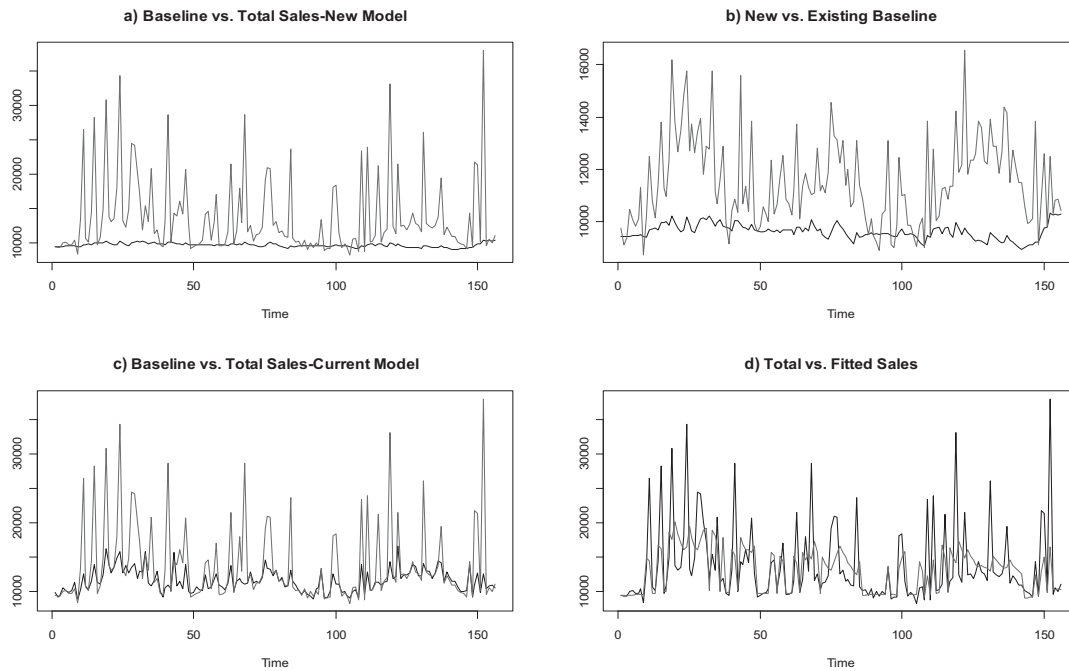


Figure 4: Plots of (a) actual sales vs. new baseline (upper left); (b) existing baseline vs. new baseline (upper right); (c) actual sales vs. existing baseline (lower left), and (d) actual sales vs. fitted sales from DLM.

We can visually see what was proven analytically: the new baseline is less volatile than the existing one (upper right panel), and the existing baseline sales highly covariate with actual promotion sales (lower left panel). This figure also presents the fitted unit sales created using the DLM model.

Finally, we present Figure 5 in order to address the problem of structural changes that motivated us to use the model of Bai, and Perron (2003). The main problem with our DLM specification is that is not able to capture structural changes (see upper left panel of Figure 5), for example, when a product experiences a permanent change in its units sold due to changes in distribution. We deal with this empirical issue by using in a first step the multiple structural change model (MSCM) of Bai, and Perron (2003). The upper left graph in Figure 6 shows the results for the same product but with the baseline model computed using the DLM-MSCM.

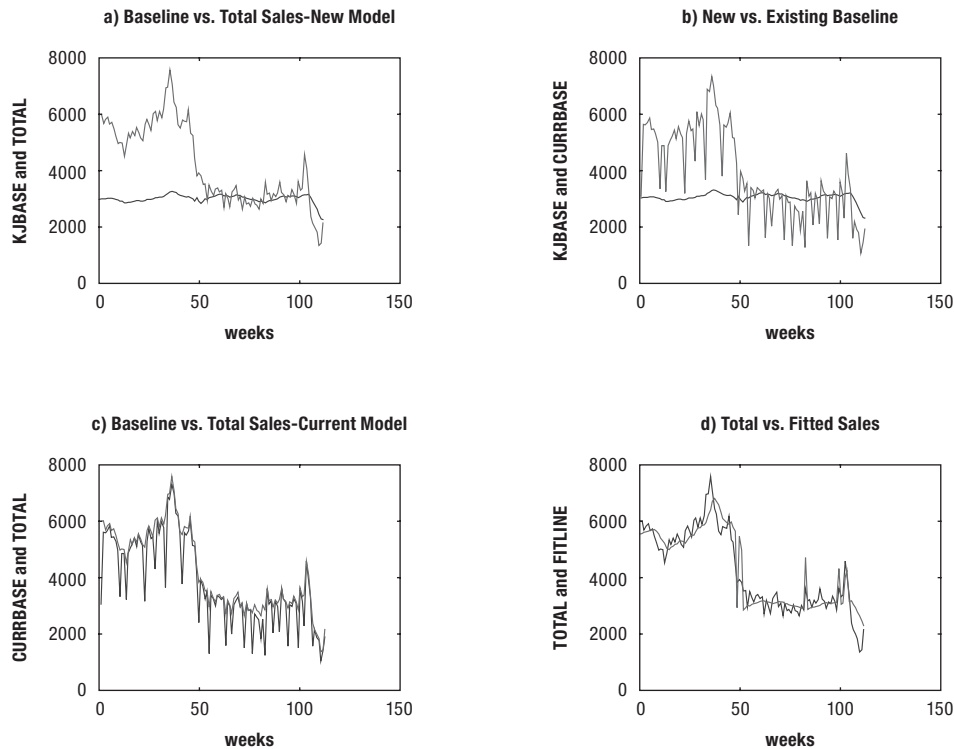


Figure 5: Plots of (a) actual sales vs. new baseline (upper left), (b) existing baseline vs. new baseline (upper right), (c) actual sales vs. existing baseline (lower left), and (d) actual sales vs. fitted sales from DLM. This figure corresponds to a product that belongs to the adult personal care category. Observe that our baseline computed using only the DLM model is not able to capture structural changes that can be present in the data (upper right).

A comparison of figures 5 and 6 makes it clear that our new baseline model, implemented by the use of the MSCM of Bai, and Perron (2003) and the DLM is able to capture that characteristic of the data quite accurately.<sup>14</sup> It is worth noting that the average  $R^2$  of the fitted values of all the data classes without (with) the MSCM is 0.71 (0.86). Apparently, just looking at this high average  $R^2$  without the MSCM implies that, just by using the DLM, we have a significantly good fit of the data. However, this high average  $R^2$  (0.71) is due to the fact that for the data that we have available, almost 90% of the data classes do not present structural changes in unit sales. Thus, if we only consider the remaining 10% of data classes (i.e., products with structural breaks in unit sales), the average  $R^2$  using *only* the DLM drops to 0.52. In the same vein, if we just consider this group with structural breaks and compute the baseline sales with the DLM-MSCM, the  $R^2$  jumps to 0.92. Thus, we dramatically improve the fit of our data using the mixture of these two models.

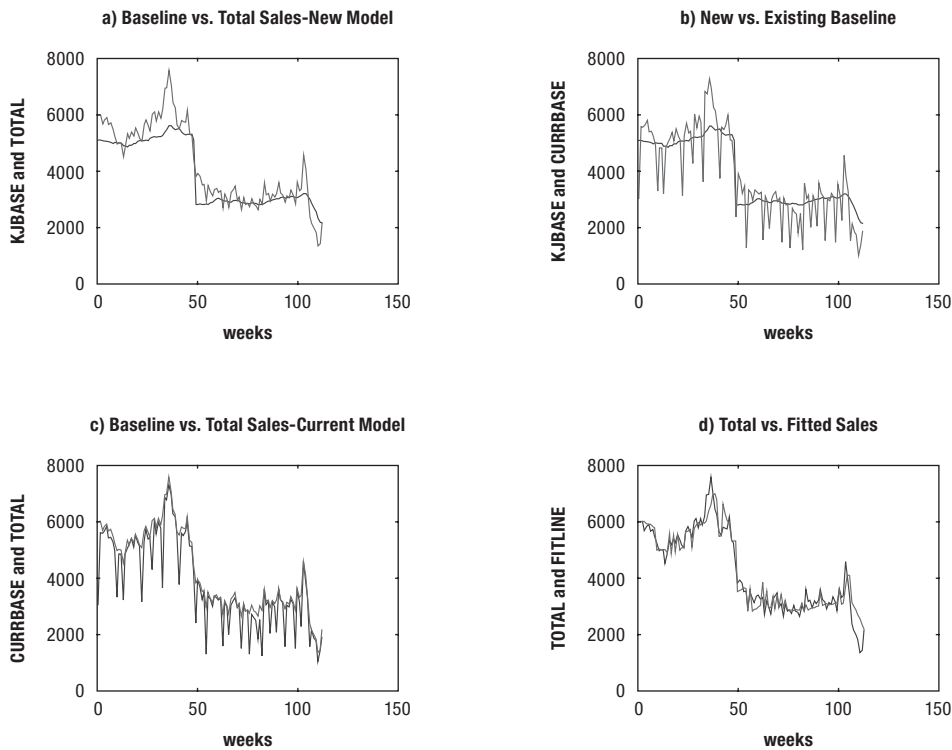


Figure 6: Plots of (a) actual sales vs. new baseline (upper left), (b) existing baseline vs. new baseline (upper right), (c) actual sales vs. existing baseline (lower left), and (d) actual sales vs. fitted sales from DLM. This figure corresponds to a product that belongs to the adult personal care category and the industry baseline used corresponds to the Scan\*Pro. Observe that our baseline computed using the DLM-MSCM model is now able to capture structural changes that can be present in the data (upper right).

## Conclusions and Future Research

In this paper, we introduce a new technique to compute the baseline sales. This new technique consists of the use of a dynamic linear model (based on Ataman et al., 2007) complemented with a multiple structural change model proposed by Bai, and Perron (2003). This new baseline sales model has many highly desirable properties for being considered as the expected sales in the absence of promotional activities: it has low sales variability (after controlling for seasonality), and it is almost not correlated with promotional activities. Moreover, this new baseline model is able to capture structural changes that could be present in certain products after controlling for seasonality and other predictable patterns.

We checked these desirable properties not only for our new proposed baseline model but also for the actual industry standards, the Scan\*Pro and PromotionScan. Our findings show that the industry benchmarks lack both these properties: they have high volatility, and they are highly correlated with promotional activities.

We presented an empirical application, studying two main categories: adult personal care products and frozen foods. Using aggregated data of 312 data classes in adult personal care products and 247 for frozen foods, we show how our baseline sales model is able to capture a more reliable expectation about the sales in the absence of promotional activities, after controlling for seasonality. We can also observe from our results that the new model strategy perfectly captures structural changes in the data.

In summary, these tests provide compelling evidence of the superiority of the baseline sales using the DLM-MSCM compared to the existing industry standards. First, we demonstrated that a week without promotions shows a lower level of sales variability than promotion weeks, particularly for chain-level data. Next, we showed that the new DLM-MSCM greatly reduced the level of volatility in the weekly baseline estimates. On average, the reduction in variability is around 80%. Finally, we demonstrated that the existing baseline sales based on a log-linear model exhibited high correlation with promotion weeks (+75%) for chain-level data. Meanwhile, our new DLM-MSCM has no significant correlation with promotional activities. Moreover, our baseline model

is able to capture structural changes present in some products. In addition to the quantitative benefits of this model, it also has the advantage of not being reliant on an expensive data-gathering infrastructure for causal measures, and it can be extended to any retailer and trade class which gathers weekly point-of-sale data.

Two potential limitations of this research are that it reflects only two categories and that the research was done only in the U.S. market. Future research involves testing this model for other CPG categories in order to generalize the results. Additionally, this model should be conducted in European markets where there is a belief by some that homogeneity of retailer promotional stimulus cannot be assumed for some chains. Finally, we expect to use our model with disaggregated data if available. In general, this research has demonstrated a new approach that greatly improves the baseline model accuracy.

## References

- ACNielsen (2008). Consumer Packed Goods Reports. <http://www.nielsen.com/us/en/insights/reports-downloads.html>
- Abraham, M., & Lodish, L. (1993). An implemented system for improving promotion productivity using store scanner data. *Marketing Science*, 12(3), 248-269.
- Ataman, B., Mela, C., & Van Heerde, H. (2007). Building brands (Working Paper). Marketing Dynamics Conference, University of Groningen, The Netherlands.
- Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18, 1-22.
- Bucklin, R., & Gupta, S. (1999). Commercial use of UPC scanner data: Industry and academic perspectives. *Marketing Science*, 18(3), 247-273.
- Cannondale Associates (2007). Industry Trends. [http://www.cannondaleassoc.com/research\\_fs.htm](http://www.cannondaleassoc.com/research_fs.htm)
- Christen, M., Gupta, S., Porter, J. C., Staelin, R., & Wittink, D. R. (1997). Using market-level data to understand promotion effects in a nonlinear model. *Journal of Marketing Research*, 34, 322-334.
- Foekens, E. W., Leeflang, P. S. H., & Wittink, D. R. (1994). A comparison and exploration of the forecasting accuracy of nonlinear models at different levels of aggregation. *International Journal of Forecasting*, 10, 245-261.
- Hanssens, D., Parsons, L. J., & Schultz, R. L. (2000). *Market response models: Econometric and time series analysis*. Norwell, MA: Kluwer Academic Publishers.
- Jetta, K. A. (2008). *A theory of retailer price promotions using economic foundations: It's all incremental* (Unpublished doctoral dissertation). Fordham University, New York.
- Nijs, V., Marnik, R., Dekimpe, G., Steenkamp, J. E. M., & Hanssens, D. (2001). The category-demand effects of price promotions. *Marketing Science*, 20(1), 1-22.
- Pauwels, K., D. Hanssens, & S, Siddarth (2002), "The Long-Term Effects of Price Promotions on Category Incidence, Brand Choice, and Purchase Quantity," *Journal of Marketing Research*, Vol. 39, 421-39.
- Van Heerde, H. J., Leeflang, P. S. H., & Wittink, D. R. (2002). How promotions work: SCAN\*PRO-based evolutionary model building. *Schmalenbach Business Review*, 54, 198-220.
- West, M., Harrison, J., & Migon, H. (1985). Dynamic generalized linear models and Bayesian forecasting. *American Statistical Association*, 80, 77-83.
- Wittink, D. R., Addona, M., Hawkes, W., & Porter, J. C. (1988). *SCAN\*PRO: The estimation, validation and use of promotional effects based on scanner data* (Working Paper).

## Footnotes

- 1 Authors' estimations based on ACNielsen and Cannondale numbers.
- 2 Judgment is based on dozens of formal and informal interviews of practitioners in CPG. Several of the interviewees are available to discuss their assessments upon request.
- 3 It is important to note that some industry experts consider that feature advertising in Europe is implemented heterogeneously by several major retailers and disaggregated data would be more appropriate in those instances.
- 4 The authors call it the "illusive post-promotion dip" because they acknowledge that the dip, which is widely accepted to be true, is rarely evident from inspection of aggregated POS data. For more details on the issue, see Jetta (2008).
- 5 It is important to note that we compute the standard deviation of the baseline sales after controlling for structural shifts and seasonality. Thus, the volatility left can be basically attributable to promotional activities of some sort.
- 6 We have programmed the DLM and the MSCM using Matlab and Gauss.
- 7 Sometimes, promotional activity data are available at an extra cost. If this is the case, the construction of the dummy variable is straight forward: 1 if it is a promotional week and 0 otherwise.

- 8 For more information about this method, see Jetta (2008) and Appendix A for a brief description of it.
- 9 Recall that temporary movements are captured directly by Equation 13.
- 10 Due to space limitations, we do not write all the results. However, they are available upon request. For similar results we refer the reader to Jetta (2008).
- 11 Due to space restrictions, we do not present the results here. They are available upon request.
- 12 We also performed the t-test at the individual level. Results show that approximately 95% of the industry baseline sales of individual products in the adult personal care products are significantly correlated with promotional activity for the first group (first-week promo) and 93% are significantly correlated with promotional activities for the last group (other-week promo).
- 13 See Jetta (2008) for another test based on a linear regression analysis.
- 14 Additional figures and proofs are available upon request.

## **Author Note**

The authors would like to thank Luc Bauwens, Andreas Heinen, Duncan James, Praveen Kopalle, Dominick Salvatore and Hrishkesh Vinod for helpful suggestions.

Kurt Jetta obtained his Ph.D at Forthan University and is President and Founder of TABS Group.

Erick Rengifo obtained his Ph.D at Catholic University of Louvain, Belgium.

\* Correspondence concerning this article should be addressed to: [rengifomina@fordham.edu](mailto:rengifomina@fordham.edu)

## Appendix A: The Endogenous Promo Dummy Variable

We propose the endogenous PROMO dummy variable ( $\phi$ ) as an alternative to the costly acquisition of the percentage of units on any promotion (PUAP) measure provided by the data supplier, a measure that is currently the industry's accepted standard for detecting the presence of meaningful promotional activity. This dummy variable is calibrated to flag any observation week where there is an abnormal deviation in weekly sales change or where the absolute sales level is significantly above the overall average. The model runs through several iterations to refine this variable in order to minimize the model standard error or to maximize its coefficient of determination ( $R^2$ ).<sup>13</sup>

To measure the accuracy of our method, we compare this PROMO dummy to the PUAP provided by the data supplier. We run simple regressions where the data provided by the data supplier were treated as the dependent variable and the PROMO dummy variable as the explanatory one. In the case of the adult personal care product (frozen food), the average  $R^2$  is 96% (94%), the coefficients were significant at the 5% significance level in all the cases, and also in all the cases, the  $p$ -values of the  $F$ -test were smaller than 0.05. The minimum  $R^2$  for adult personal care product (frozen foods) is 89% (91%), showing a robust result.

We can thus observe that there is a high level of convergence between the two measures of promotional activity, meaning that our estimated PROMO variable tracks very closely the syndicated values. With more than 90% accuracy in capturing high levels of promotional activity, the endogenous promotional calculation provides a viable substitute for the expensive causal measure infrastructure. So with confidence in the validity of the estimated promotional dummy,  $\phi$ , we can proceed with our new baseline model.

## Appendix B: The Multiple Structural Change Model

In this paper we propose to use the DLM and complement it with the multiple structural change model (MSCM). The main idea is first to detect structural changes in the data. Once this step is done, we apply piece-wise DLM to the resulting regimes. Bai, and Perron (2003) defined a multiple linear regression with  $n$  breaks ( $n+1$  regimes) as follows:

$$y_t = x_t' \beta + z_t' \gamma_j + \mu_t \quad t = T_{j-1} + 1, \dots, T_j; \quad j=1, \dots, n+1$$

where  $y_t$  is the dependent variable observed at time  $t$ ,  $x_t$  and  $z_t$  are vectors of covariates,  $(p \times 1)$  and  $(q \times 1)$ , respectively. The vectors of covariates are  $\beta$  and  $\gamma_j$ .  $\mu_t$  is the disturbance term at time  $t$ . The break points are identified by  $T_1, \dots, T_n$  and are treated as unknown variables. The unknown regression coefficients are estimated together with the break points when  $T$  observations on  $(y_t, x_t, z_t)$  are available. As the authors mentioned, this is a partial structural change model because  $\beta$  is not subject to changes and it is estimated for the complete sample. Setting  $p=0$ , gives rise to the pure structural model.

The estimation method is based on the least squares principle. For each  $n$ -partition, the associated least-squares estimates of  $\beta$  and  $\gamma_j$  are obtained by minimizing the sum of squared residuals:

$$(\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}\gamma)'(\mathbf{Y} - \mathbf{X}\beta - \mathbf{Z}\gamma) = \sum_{i=1}^{n+1} \sum_{t=T_{i-1}+1}^T [y_t - x_t' \beta - z_t' \gamma_i]^2$$

The authors showed that the break point estimators are global minimizers of the objective function.<sup>13</sup> For the estimation procedure, they propose the use of an algorithm based on a dynamic programming principle that allows the computation of estimates of the break points as global minimizers of the sum of squared residuals.