Article can be accessed online at http://www.publishingindia.com

Is Your New Product Really Boosting our Sales? An Econometric Model to Quantify the Cannibalization Effect

Vamse Goutam*

Introduction

As a manufacturer introduces a new product to compete in the market, it is competing with the already existing products of the same manufacturer. Cannibalization is an extremely important concept in marketing – a decrease in sales, revenue or market share of an already existing product of the category due to the introduction of a new product. Most top manufacturers use their already existing brand names to circumvent if not all but some barriers for entry of a new product as a part of line/brand extension strategy, believing that an already successful brand would augment the new product's sales. But, gaining movement at the expense of the parent product is not a real success. A new product is considered successful when it would have created its own market either by creating new customer base (market expansion) or through shift of customers from the competitor's base (customer switching from the competitors brand to the new product). But, there is another possibility as well – the new product might actually intrude into the existing market of the older products, thus eating away the sales of the parent product, in such a case the sales incurred by the new product is not a true incremental sales. In fact it is very critical to acknowledge that for most of the new products some amount of cannibalization is inevitable. However one could take initiatives to keep the cannibalization as minimal as possible, but to expect no cannibalization would be utopian.

Though, cannibalization has an inherent negativity, but when carefully planned and executed, it could be extremely effective, by ultimately growing the market. For example, in the chocolate market in Britain, the decline in Kit Kat's share was arrested by the launch of a new chunkier bar, which undoubtedly cannibalized the market for the original. Firms, also intentionally cannibalize their own products by producing marginally improved products, this is more common with mobile industry. The idea is to persuade the customers to buy a new upgraded version which is always expensive than the original.

Thus, understanding and quantifying cannibalization becomes a very important factor in measuring the success rate of a new product. This paper concentrates on how one could leverage econometrics in the field of marketing to quantify the cannibalization effect of a new product on the already existing products and the portfolio of the same manufacturer.

For the purpose of quantifying the cannibalization impact, we use the SUR (seemingly un-related models) which allows us to account for the fact that the sales of a particular product within a store will not just depend on its price and its own other features but will also depend on the interplay effects between the target and the other products within a store. SUR is a system of equations with multiple equations instead of one single equation with each equation representing an individual business relationship. These individual equations can be estimated with OLS as well but what makes SUR more efficient is that SUR lets the individual equations to be linked and interact with each other. A popular example of SUR is demand estimation for example:

$$Y_{pepsi,t} = X_{pepsi,t\beta pepsi} + E_{pepsi,t}$$

* Bangalore, India Vamse.kobe09@gmail.com

$Y_{coke,t}=X_{coke,t}\,\beta_{coke}+E_{coke,t}.$

Where Ypepsi is the quantity demand (sales) for Pepsi, X pepsi represents vector of regressors such as price of Pepsi, its promotion etc. and "t" represents time index.

The approach used in this paper gives us an absolute sales that was cannibalized by new product from its parent brand or from the portfolio as a whole. This enables the manufacturer to understand the percentage contribution by the new product in the category growth in terms of incrementality and thus one would know if the new innovation has actually fetched the desired results or not (its success rate).

Methodology and Model Form

Seemingly Unrelated Regression (SUR) is a system of linear equations that is not simultaneous. The concept was first considered by Zellner in 1962. The conceptual idea behind SUR is that each and every individual linear equation in the system is conceptually related to one another and aside this conceptual relationship the equations do not hold any outward relationships. Each equation in the SUR system models a different dependent variable and the vector of regressors need not to be same. For instance consider the below system of equation with three individual equations (Three dependent variables are considered as a group because they bear a conceptual relationship, but relationship, but other than this conceptual relationship the variables hold no connection):

$$Y_{t,1}=X_{t,1}\,\beta_1+U_{t,1} \qquad (1)$$

$$Y_{t,2}=X_{t,2}\,\beta_2+U_{t,2} \qquad (2)$$

$$Y_{t,3}=X_{t,3}\,\beta_2+U_{t,3} \qquad (3)$$

At first glance the above three equations seem unrelated - each individual equation has different vector of regressors and different dependent variables. Hence one would typically claim to have them estimated separately. Of course if the three equations are unrelated then they need to be estimated separately, but if these three equations hold a relationship that is not explicitly controlled by the regressors but brought forward by the their error terms ($U_{t,i}$) then it would be more efficient to have all three estimated jointly. The idea is better understood for a single equation with serial correlation, if $U_{t,1}$ is correlated with $U_{t,2}$ and $U_{t,2}$ with $U_{t,3}$ then the knowledge of $U_{t,2}$ and $U_{t,3}$ can help us in better accuracy. Thus, if the error terms of the three equations are correlated (estimated variance covariance matrix becomes non – singular) then we would have a more efficient estimator by estimating all the equations jointly.

SUR is a two-step generalized least squares (GLS) estimator for the model where the estimated variance covariance matrix of the error terms in the system is non-singular. On many applications in economics, business and management, the assumption of non-singularity may be violated and that is also the case with estimating cannibalization.

In order to quantify the cannibalization impact of a new product over the already existing products one needs to consider the fact that the sales of a particular product within a store will not just depend on its price and its own other features but will also depend on the interplay effect between the target and the other products within a store and SUR allows to account for this fact.

A large US confectionary manufacturer has introduced a new product differentiating it by a new flavor and would like to understand the cannibalization impact of the new product over the already existing products in the portfolio. Thus, wanting to understand if the sales incurred by the new product was actually an incremental sale or cannibalized from its parent products.

The portfolio consists of four products excluding the new product. For simplicity we call them product1, product2, product3, product4 and the new product. Thus, we would want to understand what percentage of the new product sales was actually rooted out from the already existing four products in the portfolio and what percentage of the new products sales was a true incremental sale.

Prior to the SUR system framework it is better to look at one single equation to get a better understanding on the model form. Assuming that we are looking at only one of the products in the portfolio – product 1. The equation will take the below form:

$$S_1=\alpha(Calculated\ \ Price_1)^{\beta 1}(PI_1)^{\beta 2}e^{\beta 3(Feature1)+\beta 4(Display)+\beta 5(Feat\&Disp)+\beta 6(Growthterm)}$$

1 1

Where,

S1 is the sales of product1.

Calculated price of product1 is an indication of what the price points in the store be in the absence of price discounts. The whole idea of calculated price is to see whether the price discounts are effective or not.

PI1 (Price Index) is a ratio of actual price (this will always be lesser than or equal to the calculated price of the product) observed in the retail store for product1 to calculated price of product1 – actual price is price after discounts. Thus, as the PI increases the gap between the calculated price and the actual price reduces which means lesser discounts at the store hence lesser sales indicating that the actual price index has a negative relationship with the sales.

Feature, Display and Feature and Display are dummy variables capturing the promotional activities with an objective to boost the sales of product1, for simplicity we refer feature, display and feature and display as the promotional variables. Since the investment in the promotional variables is made with an expectation to boost the sales over and above the base sale, we expect these variables in the model to show a positive relationship with the sales of the product.

Growth term is the variable of interest here, this is the variable that captures the impact of the cannibalization – the impact of a new product being introduced on the already existing products and the βs are the coefficients to be estimated. The growth term is defined as a ratio of the new product sales to the portfolio sales as a whole. In simple words the variable represents share of the new product in the portfolio.

The above model is in a multiplicative form such that it allows for the interactions between variables in order to account for the fact that the sales of a product is not determined only by its standalone features (Price and other promotional variables) but also that these features interact with each other to incur sales. The above multiplicative model is transformed in order to get an estimable form – we double - log linearize the above multiplicative model and the below is the new model form:

$$ln(S_1) = ln\alpha + \beta_1 ln(Calculated\ Price_1) + \beta_2 ln(PI_1) + \beta_3 Feature_1 + \ldots + \beta_6 Growth\ Term_1 + \varepsilon_1$$

With four already existing products in the portfolio, the SUR system will consist of four independent equations like the above each modeling one of the four already existing products in the portfolio against its own price points, promotions and the growth term. The model form will look as below:

$$ln(S_{1,t}) = ln\alpha_1 + \beta_1 ln(CalculatedPrice_{1,t}) + \beta_2 ln(PI_1) + \beta_3 Feature_{1,t} + \ldots\ldots + \beta_6 GrowthTerm + \varepsilon_{1,t} \quad \text{-- (1)}$$

$$ln(S_{2,t}) = ln\alpha_2 + \beta_1 ln(CalculatedPrice_{2,t}) + \beta_2 ln(PI_2) + \beta_3 Feature_{2,t} + \ldots\ldots + \beta_6 GrowthTerm + \varepsilon_{2,t} \quad \text{-- (2)}$$

$$ln(S_{3,t}) = ln\alpha_3 + \beta_1 ln(CalculatedPrice_{3,t}) + \beta_2 ln(PI_3) + \beta_3 Feature_{3,t} + \ldots\ldots + \beta_6 GrowthTerm + \varepsilon_{3,t} \quad \text{-- (3)}$$

$$ln(S_{4,t}) = ln\alpha_4 + \beta_1 ln(CalculatedPrice_{4,t}) + \beta_2 ln(PI_4) + \beta_3 Feature_{4,t} + \ldots\ldots + \beta_6 GrowthTerm + \varepsilon_{4,t} \quad \text{-- (4)}$$

The SUR system will now have four independent double–log linearize equations with the same growth term repeated in all the equations.

Data

Data is a cross sectional data that has been simulated for1000 stores over a period of 78 weeks. This is a point of sales data where the sales information along with the data on the independent variables have been simulated for each and every individual store for each week. The 78week time period is actually divided into two segments, 52 weeks prior the new product launch and 26 weeks post the launch. The idea here is to have at least one whole year of data on the already existing products in the portfolio, sothat the model has enough data points to learn and capturethe pattern of these products and also give enough timefor the new product to stabilize in the market.

Start Period:02/22/2014

End Period: 08/15/2015

Total # of Stores:1000

Total # of Weeks:78

Total # of Observations:78,000

Initial Diagnosis

High promotional support (Display and Price Discounts)

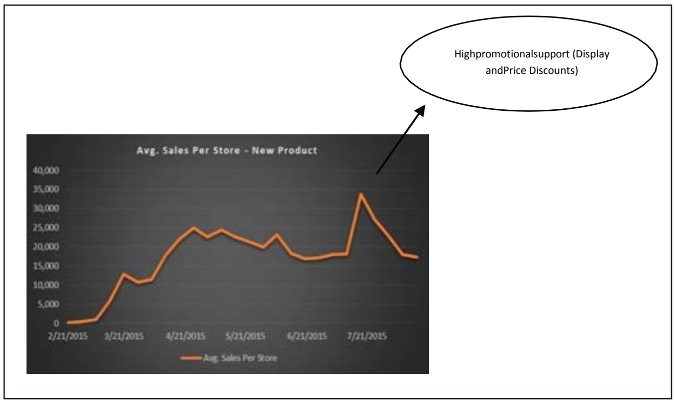

Figure 1.1: (New Product Sales per Store by Week)

Figure 1.2: (Prodcut 1 Sales per Store by Week)

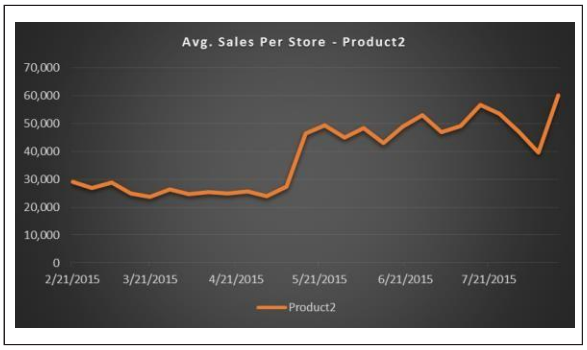

Figure 1.3:

Figure 1.4:

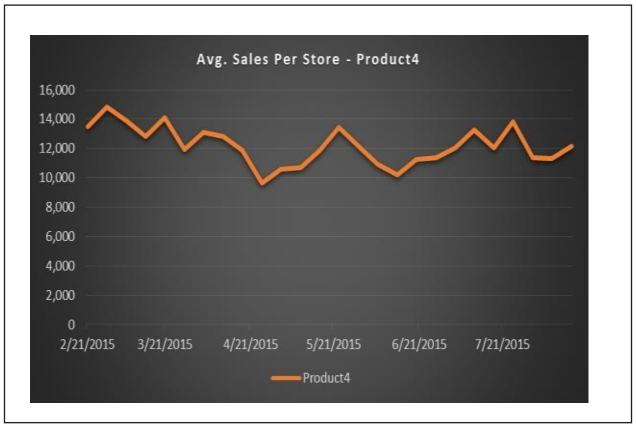

Figure 1.5. (Product 4 Sales per Store by Week)

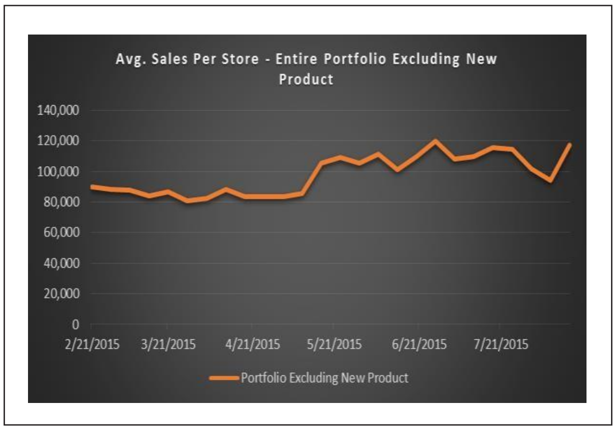

Figure 1.6. (Entire Portfolio Sales per Store Excluding New Product)

Looking at the initial diagnosis (Sales Per Store by Week) the first thing that one could claim is that – "distributional gains and merchandising support drove the sales of the new product – see Figure 1.2", there has been a combination of promotional activity (In – Store Display and 30% price discount) incurred during the stabilization period, which supported the launch of the new product.

Looking at the plots of the already existing products in comparison to the new product one could claim that the cannibalization effect within the portfolio has been really contained to minimal and this could have been carried out with appropriate merchandising support by the manufacturer for the already existing products in the portfolio.

The trend of the already existing products in the portfolio is either neutral or upward moving trend except product 4 – see Figure 1.5, which shows a slight negative sloping trend, but it is too early to conclude that this downward sloping trend is indeed a cannibalization effect by the new product, however with the above plots it is now safe to claim that the cannibalization has been really minimal. Figure 1.6 – Sales per store indicates that even at the portfolio level there has been almost no cannibalization effect, even if it had incurred it has been contained to very minimal. This is even more evident with the below plot – Figure 1.7 –

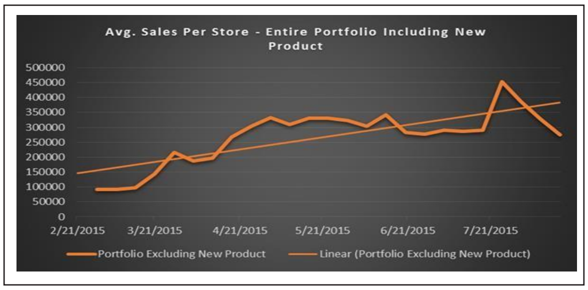

Figure1.7: (Entire Portfolio Sales per Store by Week Including New Product)

If Figure 1.6 is compared with Figure 1.7, one could claim that the new product launch has actually been incremental to the portfolio and indeed incurred incremental sales.

For modeling SAS was used and procedure PROC SYSYLIN was implemented, below is a

sample code of procedure PROC SYSLIN:

procsyslin data = Model_Datasuroutest=est_model out=pred_model;

model Sales1 = cal_price1 PI2 display1 feature1 feat_disp1 growth_term seasonality trend;

model Sales2 = cal_price2 PI2 display2 feature2 feat_disp2 growth_term seasonality trend;

model Sales3 = cal_price3 PI3 display3 feature3 feat_disp3 growth_term seasonality trend;

model Sales4 = cal_price4 PI4 display4 feature4 feat_disp4 growth_term seasonality trend;

run;

In the above code there are four model statements one for each product in the portfolio except the new product. Below is the list of significant variables of the model results for all four equations, one thing to note here is that the below results are output from seemingly unrelated regression estimation. PROC SYSLIN provides both the outputs the OLS and the SUR but the below results are only SUR as they are the ones we are concerned with:

Product1		
Variable	*Estimates*	*Pr>|t|*
Intercept	(0.024)	0.030
Price	(0.405)	0.025
PI	(0.003)	<.0001
Seasonality	0.187	0.023
CaniblizationEffect	(0.044)	<.0001
Promo1(Display)	0.114	0.102
Promo2(Feature)	0.127	0.003
TREND	0.094	<.0001

Product2		
Variable	*Estimates*	*Pr>|t|*
Intercept	(0.113)	<.0001
Price	(0.188)	0.002
PI	(0.470)	<.0001
Seasonality	0.286	0.183
CaniblizationEffect	(0.099)	0.022
Promo1(Display)	0.035	<.0001
Promo2(Feature)	0.153	<.0001
Promo3(Featureand	0.337	<.0001
TREND	0.039	<.0001

Product3		
Variable	*Estimates*	*Pr>|t|*
Intercept	0.044	<.0001
Price	(0.592)	0.002
PI	(0.518)	<.0001
CaniblizationEffect	(0.076)	0.022

Product3		
Variable	*Estimates*	*Pr>|t|*
Promo1(Display)	0.346	<.0001
TREND	0.439	<.0001

Product4		
Variable	*Estimates*	*Pr>|t|*
Intercept	0.242	0.030
Price	(0.432)	0.025
PI	(1.456)	<.0001
Seasonality	0.187	0.023
CaniblizationEffect	(0.380)	<.0001
Promo1 (Display)	0.075	0.102
Promo2 (Feature)	0.266	0.003
TREND	0.998	<.0001

Across models we see that the cannibalization term has become significant and negative. This indicates that there has been some amount of cannibalization effect because of the new product.

As the above results are from a linear model form and in order to quantify the exact amount of the cannibalized sales and the incremental sales of the new product we need to use the above results and roll them back to their real space i.e. multiplicative model form.

We use the above cannibalization effect estimates and calculate the actual cannibalized sales by the new product for all the 4 products individually and then for the entire portfolio. Due to confidentiality of the data with the client, the entire calculation is not shown in this paper using the entire data, however in the appendix using product 3 a sample calculation is shown, so that the reader gets an idea on how the quantification is done after the modeling stage.

Insights and Actions

- The New product launch has been successful in generating portfolio incrementality, hence driving the brand and the portfolio growth.
- Focus on building strong distribution in early time periods along with merchandising support (evident from Figure 1.1) as it enables strong trial to allow for repeat purchases within the first year of launch.
- Of all the four products the new product tends to have cannibalized product4 the most with 12.13% decrease in its sales. Hence, the downward trend that was observed during the initial diagnosis can now be attributed to the new product launch.
- The overall % of cannibalized sales that the new product accounted is 23% with 77% incremental.
- The 23% of the cannibalization by the new product is distributed as 1.04%, 6.76%, 2.82% and 12.13% for product 1, product 2, product 3 and product 4 respectively.
- Cannibalization sample calculation:
- Assume that fora particular product the growth term estimated through PROC SYSLIN is**-0.076** and we are working with onlyten weeks of data.
- Given that, the original multiplicative model has been linearized, thereforethe estimates needs to berolled back to theoriginal space. Below is given theformulathat we useto roll back the growth term
- *((Exponential(***-0. 076***(growth term))) -1)*– This is calculated for each and every single week. Please look at the below table for further explanation.

Please note the numbers in the above table are only for explanatory, hence were simulated. These numbers were not used as part of the any project that has been executed.

((Exponential(-0. 076(growth term))) -1)

((Sales of the existing product/ (1+Weekly Lift))

	Growth TermCoeficient	-0.076					
Week	Lift	SalesofExisitngProd	NewProductSales	ExisitngProductsalesifnonewproduct	SalesCannibalized	NewProd %cannibalisticto b	Growth Term
Week1	0%	32,106	0	32,106	0	0%	0.00
Week2	0%	31,350	250	31,369	19	8%	0.01
Week3	0%	32,007	2,017	32,162	155	8%	0.06
Week4	-2%	30,706	9,818	31,466	760	8%	0.32
Week5	-4%	33,191	19,006	34,677	1,486	8%	0.57
Week6	-5%	30,199	19,145	31,699	1,500	8%	0.63
Week7	-5%	31,041	22,696	32,827	1,786	8%	0.73
Week8	-4%	36,040	21,632	37,734	1,693	8%	0.60
Week9	-6%	32,886	28,170	35,113	2,227	8%	0.86
Week10	-6%	33,907	28,678	36,173	2,266	8%	0.85
Overall		323,434	151,411	335,326	11,892	8%	

In the above table the cell highlighted in green is the growth term estimate that was obtained through SUR model.

Column 1 (Week)-states the week number,

Column 2 (Weekly Lift)-states the percentage of cannibalization effect– This is calculated using the below formula

((Exponential(-0.076)*(growth term))– 1),

Column 3 -Sales ofthe Existing product, Column 4 – Sales of the new product launched,

Column 5 – Existing Product sales if the new product has not been launched– This is calculated using the below formula

((Sales of the existing product/ (1+Weekly Lift))

Column 6 – Sales Cannibalized– This is the difference betweenexisting product sales (column 3)

andexisting product sales if the new product hasnot been launched (column 5),

Column 7 & Column 8 - Percentage cannibalization incurred by the new product in the above example it is 8% and growth term (independent variable that actually went into the model).

In the given example, we have estimated the cannibalization sales to be 8% then the incremental sales of the new product would be 100% - 8% = 92%.

Similarly the cannibalization effect is calculated for all the products in the portfolio individually and then the cannibalized sales from all the existing products is summed up across products/equations in the system to arrive at the overall cannibalized sales within the portfolio.