Marketing-Mix Recommendations to Manage Value Growth at P&G Asia-Pacific

Technical Appendix

V. Kumar*
Jia Fan
Rohit Gulati
&
P Venkat

November 2008

* V. Kumar (VK) is the Richard and Susan Lenny Distinguished Chair Professor in Marketing and the Executive Director of the Center for Excellence in Brand and Customer Value Management at J. Mack Robinson College of Business, Georgia State University, Atlanta, GA. Phone: (404) 413-7590; Fax: (404) 413-7585; Email: vk@gsu.edu.

Jia Fan is a doctoral student in Marketing at Georgia State University, Atlanta, GA.

Rohit Gulati is the Manager – Beauty Care CMK, P&G Asia-Pacific, Singapore.

P Venkat is the Associate Director, P&G Asia-Pacific Pvt. Ltd., Singapore.

The authors thank the Editor, Gary Lilien, John Roberts and the reviewers for their valuable comments. The authors also thank the judges of the 2007 INFORMS Practice Prize committee, Robert Leone, S. Sriram, Joseph Pancras, Hongju Liu, Andrew Petersen and Denish Shah for their comments on an earlier version of the manuscript and thank Renu for copyediting the Technical Appendix.

Overview: We first provide the details on how we develop the same-tier and up-tiering models, and then we discuss the model estimation process and the model specification test. Once the models are estimated, we present a sample of parameter estimates and a comparison of actual with fitted data. After that, we show how we calculate various elasticities from the estimates and their confidence intervals. Finally, a list of references related to this appendix is provided.

♦Details on the Same-tier Model and Up-tiering Model

We specify Model A1 as a standard random coefficient model, for each cross-section *i*, the model is

$$Y_i\beta_i + X_i\Gamma_i = \varepsilon_i \qquad \text{---A1}$$

where *i=1,..10*, and Y_i is a *T*N* matrix of endogenous variables, *N=2* is the number of endogenous variables; X_t is a *T*M* matrix of all the exogenous variables, *M=20* is the number of exogenous variables; β_i is a *N*N* matrix of coefficient for Y_i; Γ_i is a *M*M* matrix of coefficient for X_i; and ε_i is a *M*N* matrix of errors. More specifically, these matrices are:

$$Y_i = \begin{bmatrix} Sales_{i1}\, Distribution_{i1} \\ Sales_{i2}\, Distribution_{i2} \\ \ldots\ldots \\ Sales_{iT}\, Distribution_{iT} \end{bmatrix} \qquad \text{---A2}$$

$$X_i = \begin{bmatrix} 1Y03_{i1}\, Y04_{i1}\, Price_{i1}\, (Y03*Price)_{i1}\, (Y04*Price)_{i1}\, (CI)_{i1} \cdots (QSG)_{i1} \\ 1Y03_{i2}\, Y04_{i2}\, Price_{i2}\, (Y03*Price)_{i2}\, (Y04*Price)_{i2}\, (CI)_{i2} \cdots (QSG)_{i2} \\ \ldots\ldots \\ 1Y03_{iT}\, Y04_{iT}\, Price_{iT}\, (Y03*Price)_{iT}\, (Y04*Price)_{iT}\, (CI)_{iT} \cdots (QSG)_{iT} \end{bmatrix} \qquad \text{---A3}$$

Since we don't have $Sales_{it}$ in the equation *(2)*, which means the current sales won't influence the current *Distribution*, and we also want to normalize the *Sales* coefficient to 1, thus in our case

$\beta \equiv \begin{bmatrix} 1 & 0 \\ a_1 & a_2 \end{bmatrix}$, it is fixed over time. ---A4

$\Gamma_i = \begin{bmatrix} \theta_{11}\theta_{12} \\ \theta_{21}\theta_{22} \\ \cdots \\ \theta_{M1}\theta_{M2} \end{bmatrix}$, and we assume $\Gamma_i = \Gamma + U_i$, where $E(U_i) = 0\ \forall i$, $V(U_i) = \Delta \forall i$ ---A5

$\varepsilon_i = \begin{bmatrix} \varsigma_{01}\varsigma_{02} \\ \varsigma_{11}\varsigma_{12} \\ \cdots \\ \varsigma_{T1}\varsigma_{T2} \end{bmatrix}$, and we assume $E(\varepsilon_i) = 0\ \forall i$ and $V(\varepsilon_i) = \Omega_i \otimes I_T$ ---A6

Distribution is a retailer - calculated measure of the percentage of stores in a state that are selling the item weighted by the stores' total turnover. A store is considered in distribution if a sale for the SKU exists in the last month. *Distribution* is measured as

$$Weighted\ Distribution\ of\ \ product\ \ i = \frac{Total\ turnover\ having\ \ product\ \ i}{Total\ turnover}$$

SKU represents the package size in this dataset. For example, *Distribution* for product *i* is

(50+30+20)/ (100+50+30+20) =50%, where 50, 130, and 20 are the total turnover of the stores selling product *i*; 100 is the turnover of the store not selling product *i*. *Distribution* for all the brands and SKUs are defined in the same way.

After specifying the model, we estimate it by a three step procedure discussed below.

Step 1. Two Stage Least Squares Estimation (2SLS)/ Instrumental Variable (IV) estimation

Since we only have one instrument variable in our model, the 2SLS estimators should be equivalent to IV estimators. Because we are most interested in the effects of marketing mix variables on the sales volume after controlling for the endogeneity of *Distribution*, we transform the model into a reduced form (Balestra and Negassi, 1992).

$$Y_i = X_i \gamma_i + \nu_i \qquad \text{---A7}$$

where $\gamma_i = \gamma + \psi_i$ and $\gamma = -\Gamma \cdot \beta^{-1} = -\Gamma \begin{bmatrix} 1 & 0 \\ a_1 & a_2 \end{bmatrix}^{-1}$ ---A8

Step 2. Random Coefficient Simultaneous Equation Estimation

We estimated the individual cross-sections in the previous step, and in this step we obtain the regular RCR estimator by weighted average of $\hat{\gamma}_i$, the cross-section *i=1,2,……S,* where *S=Number of SKU*Number of States*. It has shown that the regular RCR estimator is a weighted average of 2SLS estimator in Balestra and Negassi (1992), and the vectorization form is demonstrated below. The RCR estimator is for the entire data set with all the cross sections. We express the model in the vector form,

$$\begin{bmatrix} vec(Y_1) \\ vec(Y_2) \\ \cdots \\ vec(Y_S) \end{bmatrix} = \begin{bmatrix} I \otimes X_1 \\ I \otimes X_2 \\ \cdots \\ I \otimes X_S \end{bmatrix} vec(\gamma) + \begin{bmatrix} vec(\eta_1) \\ vec(\eta_2) \\ \cdots \\ vec(\eta_S) \end{bmatrix} \qquad \text{---A9}$$

The GLS estimator for the model above is $vec(\hat{\gamma})^{GLS} = \left[\sum_{i=1}^{S} W_i^{-1} \right]^{-1} \sum_{i=1}^{S} \left[W_i^{-1} vec(\hat{\gamma}) \right]$ ---A10

where $\hat{\gamma}$ is the individual 2SLS/IV estimator for each cross-section we obtained in Step 1 and W_i^{-1} is the weight matrix, which has two components: the MSE of the individual $\hat{\gamma}_i$ and the total dispersion of $\hat{\gamma}_i$ from across all the sections, which is $W_i = \Omega_i \otimes (X_i' X_i)^{-1} + \Delta$ ---A11

This weight matrix can be estimated in as in the feasible GLS models.

Hsiao and Pesaran (2004) also suggested this two stage GLS approach to estimate the RCR model with a system of equations. An important contribution of our proposed approach is that it provides a practical way to estimate RCR model with a system of equations, since there is no statistical software that supports RCR with a system of equations to our knowledge, while the 2SLS and GLS estimators are generally attainable by any statistical software.

Now we have both individual 2SLS estimate for each cross-section and the RCR for the pooled data. By calculating the weighted average of these two estimates, we can get our weighted RCR estimates for each cross-section, which will be shown to be a BLUE in the next step.

Step 3. Weighted Estimator Calculation

The GLS estimator above will only provide the common mean for all the cross-sections. Different coefficient estimates for each cross-section would be much closer to the reality and more practical since individual coefficient estimates can provide us insight into how sales volume of each cross-section (e.g., SKU and state combination) responds to its own and competitors' marketing-mix changes, and P&G can also keep track of sales change for each SKU in each state. The two types of variations shown in the Step 2 indicate that the variation of parameter estimates for individual cross-section should be accounted for in the analysis. We develop a new RCR estimator, which is a weighted average of pooled RCR and the individual 2SLS to incorporate the heterogeneity across cross-sections.

The idea of a weighted estimator has been in the literature for a long time. The simplest case is the weighted least squares estimator (WLSE), which gives more strength to the data points with lower variance by using the inverse of variance as the weights. The model fit will be improved by this weighted estimator. Searle and Pukelsheim (1986) have shown that the weighted estimator with the inverse of variance as weights generate the smallest variance among all the weighted estimators in the mixed models.

For the case of a random coefficient model, Kadiyala and Oberhelman (1982) developed a weighted estimator as a weighted average of individual OLS estimator in each cross-section and a RCR estimator for the pooled data. Following their idea, we develop a weighted estimator for our system of equations case and prove that this weighted estimator is a Minimum Mean Square Linear Unbiased Estimator (MMSLUE). Our proposed estimator for section *i* is,

$$vec(\tilde{\gamma}_i) = \underbrace{\hat{\Delta}\hat{W}_i^{-1}}_{p_i} vec(\hat{\gamma}_i) + \underbrace{\hat{\Omega}_i \otimes (X_i{}'X_i)^{-1}\hat{W}_i^{-1}}_{I-p_i} vec\left(\hat{\hat{\gamma}}_i\right) \qquad \text{---A12}$$

Thus, p_i is the weight for individual 2SLS/IV, and $I - p_i$ is the weight for regular RCR. Since both $\hat{\gamma}_i$ and $\hat{\hat{\gamma}}_i$ are unbiased estimators, this weighted estimator is still an unbiased estimator. Intuitively, this weighted estimator follows the same idea as the earlier weighted estimators: giving strength to observations/sections by using their inverse variance as their weights. From A20, we can see that when the MSE of $\hat{\gamma}_i$ is high, more weight is given to $\hat{\hat{\gamma}}_i$ and when the dispersion is high, more strength is given to $\hat{\gamma}_i$ (Kadiyala and Oberhelman 1982). Thus, we get a new group of parameter estimates which vary across cross-sections. The following proof will show that this estimator minimizes MSE among all the weighted estimators including the two special cases: individual 2SLS ($\hat{\gamma}_i$) and the pooled RCR ($\hat{\hat{\gamma}}$).

For the up-tiering models, we followed a similar approach that we used for estimating same-tier models. The only difference is that there are more levels in the hierarchical model since we stack one brand below the other, and the cross-section is defined as the SKU-brand-state combination. We can also show the properties of this estimator is BLUE (Kumar and Fan 2008).

♦ Model Parameter Estimates for Price, Distribution and CI

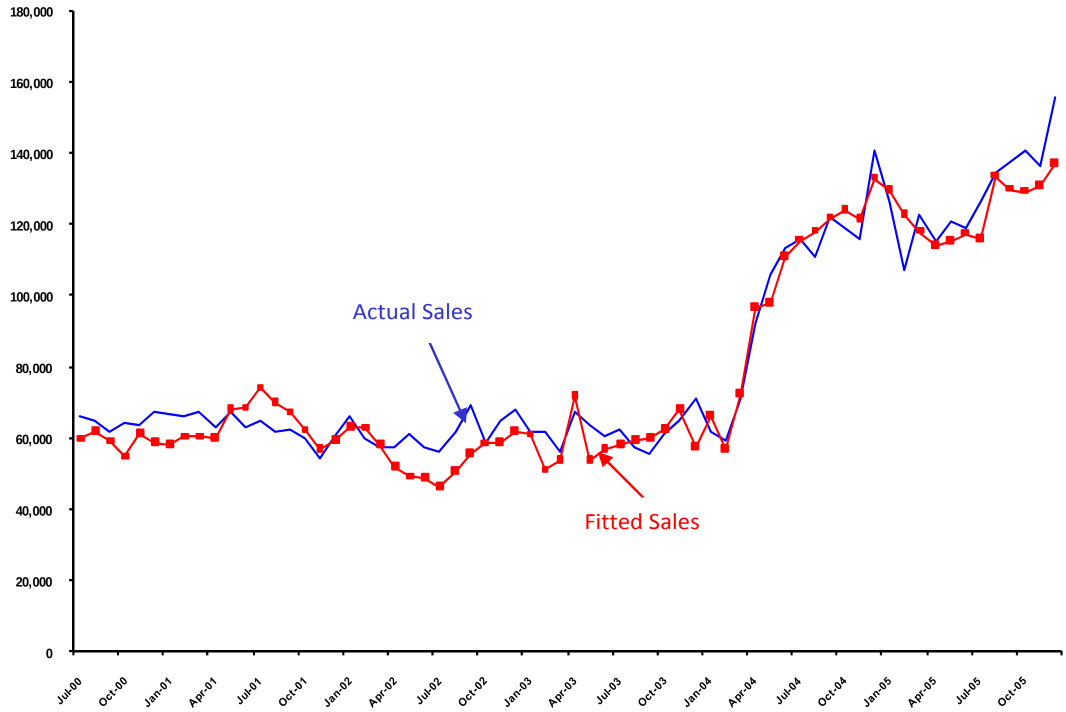
Since our same-tier model is at SKU-state level, and in order to get the weighted RCR estimator, we estimate the individual 2SLS at each SKU-state level first. Because there are so many parameters in the model including the same tier competitors' marketing-mix variables, we demonstrate only the parameter estimates of *Price*, *Distribution* and *CI* for Ariel 500g and compare the 2SLS and RCR models in Table A-1.

Table A-1 Ariel Bags same-tier Model Parameter Estimates from Sales Response Model (Ariel 500g)

State	2SLS			Weighted RCR			State	2SLS			Weighted RCR		
	Price	*Distribution*	*CI*	*Price*	*Distribution*	*CI*		*Price*	*Distribution*	*CI*	*Price*	*Distribution*	*CI*
1	-4.75	60.65	520.08	-5.39	73.6	317.89	**6**	-3.425	58.66	464.53	-5.76	73.44	401.44
	(0.54)	(5.57)	-146.6	(0.26)	(2.88)	-118.28		(0.34)	-5.12	-143.1	(0.26)	-4.99	-135.65
2	-5.665	62.05	477.49	-7.48	72.76	377.69	**7**	-4.345	62.65	501.9	-4.57	67.71	435.76
	(0.25)	-6.33	-137.42	(0.21)	-2.43	-145.7		(0.30)	-5.59	-126.91	(0.29)	-5.34	-101.3
3	-1.365	59.38	498.73	-2.54	68.99	396.65	**8**	-3.425	65.75	499.01	-5.95	69.46	422.92
	(0.34)	-4.76	-155.9	(0.31)	-3.65	134.2		(0.44)	-6.23	-132.5	(0.26)	-5.76	-117.07
4	-1.105	64.17	510.83	-3.93	67.42	402.11	**9**	-0.485	56.93	479.21	-3.58	70.22	415.65
	(0.54)	-5.98	-134.52	(0.34)	-4.76	-122.2		(0.54)	-5.54	-119.73	(0.28)	-4.01	-111.65
5	-1.175	63.59	465.32	3.49	75.69	379.23	**10**	-1.175	59.89	515.79	-2.64	72.48	404.36
	(0.52)	-4.97	-122.34	(0.25)	-3.01	-121.34		-0.33	-4.99	-144.9	-0.26	-3.97	-106.73

From the model estimates shown above, we can see that 2SLS estimates generally have higher standard errors than weighted RCR model estimates. In other words, the weighted RCR estimators are more efficient than 2SLS estimates. The *H* statistics are not significant for the parameter estimates for all the cross-sections (SKU-state combinations). It means that we do not reject the null hypothesis that the weighted RCR model is correctly specified. After applying the same procedure for Up-tiering models, we find that the weighted RCR estimators are also more efficient than the 2SLS estimators. Figure A-2 shows that the predicted sales volume is very close the actual sales volume by using weighted RCR approach.

Figure A-2. Fitted vs. Actual Sales Volume using a weighted RCR model for Ariel Bags*



*The sales volume (in 000SU) has been aggregated to the national level

After estimating the model parameters, we test to see if the model specifications are valid.

♦Details of the Model Specification Test

Due to the endogeneity of Distribution in the model, the two stage least squares estimator $\hat{\gamma}_i$ is the unbiased and consistent estimator for our model. However, under the null hypothesis that our weighted estimation is correctly specified, $\hat{\gamma}_i$ is less efficient than the weighted RCR estimator $\tilde{\gamma}_i$. Thus, we adopt Hausman (1978) specification test to evaluate the appropriateness of the weighted RCR estimator $\tilde{\gamma}_i$ for our model. The basic idea is that if $\tilde{\gamma}_i$ is significantly different from $\hat{\gamma}_i$, it must be biased, thus inappropriate for the model (Hausman, 1978; Leone et al., 1993).

After the specification tests, we prove that the model is valid. We then use the parameter estimates to calculate the elasticities and their confidence intervals. For additional details, please refer to Kumar and Fan (2008).

♦ Calculation of Elasticities and their Confidence Interval

Since we have three parameter estimates for price in the three time zones, and we have parameter estimates at the SKU-state level. We now have to convert the parameter estimates at the state level into national level. For example, Ariel 500g price elasticity in the months after March 2004 (we call it period 3) at the national level is

$$\eta_{P\_500\_prd3\_t} = \frac{\partial Q_{500\_prd3}}{\partial P_{500\_prd3}} * \frac{P_{500\_prd3\_t}}{Q_{500\_prd3\_t}}$$
$$= \frac{\partial \sum_{m=1}^{10} Q_{500\_prd3\_m}}{\partial \overline{P}_{500\_prd3}} * \frac{P_{500\_prd3\_t}}{Q_{500\_prd3\_t}} \qquad \text{----- A13}$$
$$= \sum_{m=1}^{10} \underbrace{\frac{\partial Q_{500\_2005\_m}}{\partial \overline{P}_{500\_2005}}}_{\alpha_{3\_500\_m}} * \frac{P_{500\_prd3\_t}}{Q_{500\_prd3\_t}} = \left[\sum_{m=1}^{10} \tilde{\alpha}_{3\_500\_m}\right] * \frac{P_{500\_prd3\_t}}{Q_{500\_prd3\_t}}$$

where $P_{500\_prd3\_t}$ is the price of 500g in all the states in the month t in period 3 (after March, 2004)

Q_{500_2005} is the national monthly sales volume in period 3, and *m=1.2. ….10* represents the 10 states.

The elasticity before January, 2003 (period 1) is

$$\eta_{P\_500\_prd1\_t} = \sum_{m=1}^{10} (\tilde{\alpha}_{3\_500\_m} + \underline{\tilde{\alpha}_{4\_500\_m}}) * \frac{P_{500\_prd1\_t}}{Q_{500\_prd1\_t}} \qquad \text{----- A14}$$

and the elasticity between January, 2003 and February, 2004 (period 2) is

$$\eta_{P\_500\_prd2\_t} = \sum_{m=1}^{10} (\tilde{\alpha}_{3\_500\_m} + \underline{\tilde{\alpha}_{5\_500\_m}}) * \frac{P_{500\_prd2\_t}}{Q_{500\_prd2\_t}} \qquad \text{----- A15}$$

This is how we calculate the elasticities for each time period using the time-varying coefficient estimates in our model. The three different coefficient estimates for price represent three different price policy periods, and $\frac{P_{500\_prd\_t}}{Q_{500\_prd\_t}}$ reflects the instantaneous market conditions.

We reported the confidence interval for elasticities, which were calculated by the bootstrapping method since it provides a better approximation to the confidence intervals in most situations. In our case, the elasticities are calculated from the linear model estimates. Thus, we follow the Dorfman, Kling and Sexton (1990) bootstrapping procedure to estimate the confidence intervals for elasticities. As our elasticities change over time depending on the ratio of price and sales volume in each time period, we obtain the average confidence interval of the average elasticity to evaluate the overall strategic uncertainty. The detailed procedure is as follows:

i. Calculate residuals by $e_t = Y_t - \hat{Y}_t, t = 1, \cdots T$ from our weighted *RCR* model

ii. Use the residuals computed in *i* to create a distribution ϕ . The residuals are scaled by factor $\left[T/(T-k)\right]^{0.5}$ and each scaled residual is assigned a weight $1/T$

iii. Do a random draw with replacement from the distribution ϕ and generate a vector of new Y_t $\{Y_1^*, \cdots Y_T^*\}$ where $Y_t^* = \text{fitted model} + e_t^*$

iv. Re-estimate the models for each cross-section and get a new group of parameter estimates:

$\{\alpha^*_{0i} \cdots \alpha^*_{ki}\ \beta^*_{0i} \cdots \beta^*_{ki}\}$

v. Create new elasticity estimates for SKU *i*: $\hat{\eta}^*_i = \left(\sum_{m=1}^{M} \alpha^*_{3i}\right) * \frac{P_{avg\_i}}{Sales_{avg\_i}}$, where α^*_{3i} is the new parameter estimate for price in cross section *i*, $\left(\sum_{m=1}^{M} \alpha^*_{3i}\right)$ is the summation across all the states *m*, m=1,2,…10. P_{avg_i} is the average price of SKU i across all the 10 states in December, 2005 and $Sales_{avg\_i}$ is the average sales of SKU i across all the 10 states in December, 2005.

vi. Repeat ii to iv. 1000 times by drawing from ϕ and generate empirical distribution for $\hat{\eta}^*_i$. Assuming the empirical pdf and cdf of $\hat{\eta}^*$ are f and F respectively. Therefore the bias-corrected (using standard normal distribution) percentile 95% CI for $\hat{\eta}^*$ is $\left[F^{-1}(\Phi(2z_0 + z_{0.025})), F^{-1}(\Phi(2z_0 + z_{0.975}))\right]$ where $z_0 = \Phi(F(\hat{\eta}))$ (Dorfman et al. 1990). The application of this procedure results in the availability of the intervals around each price elasticity estimate of each SKU.

<u>Appendix References</u>

Balestra, P. and S. Negassi (1992), "A Random Coefficient Simultaneous Equation System with an Application to Direct Foreign Investment by French Firms," *Empirical Economics*, 17, 205-220.

Biørn, Erik, (1999), "Randcom Coefficient in Regression Equation Systems: The case with unbalanced Panal Data, " University of Oslo, Department of Economics, *Memorandum.*

Dorfman, Jeffrey H., Catherine L. Kling and Richard J. Sexton (1990) "Confidence Intervals for Elasticities and Flexibilities: Reevaluating the Ratios of Normals Case," *American Journal of Agricultural Economics*, Vol. 72, No. 4 (Nov., 1990), 1006-1017

Hsiao, Cheng and M. Hashem Pesaran (2004), "Random Coefficient Panel Data Models," *IEPR Working Papers* 04.2, Institute of Economic Policy Research

Harville, D.A. (1976), "Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects," *Annals of Statistics*, 2, 384-395

Hausman, J. A., (1978), "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1271.

Kadiyala, K. Rao and Dennis Oberhelman, (1982), "Response Predictions in Regressions on Panel Data," *Communications in Statistics: Theory and Methods*, 11 (23), 2699-2714.

Kumar, V and Jia Fan (2008), "A Weighted Random Coefficient Regression Approach for Panel Data," *Working Paper, Georgia State University.*

Leone, Robert P., Dennis Oberhelman and Francis J. Mulhern, (1993), "Estimating Individual Cross-Section Coefficients from the Random Coefficient Regression Model," *Journal of the Academy of Marketing Science*, 21,

Searle, Shayle R. and Friedrich Pukelsheim, (1986), "Effects of Intraclass Correlation on Weighted Averages," *The American Statistician*, 40(2), 103-105

Woodridge, Jeffrey M., (2002), "*Econometric Analysis of Cross Section and Panel Data,*" Cambridge, Massachusetts: MIT Press

♦ A Portion of the Screenshot of Volume and Value Simulator

Target	Current Own Price (per pack)	Monthly Volume (MSU)	Current Monthly Value (INR 000)	Current Tide Overall Volume (MSU)	Current Ariel Total Volume (MSU)	Current Portfolio Volume (Tide + Ariel)
Ariel_SKU500	52.2	59,956.00	37,522.26	242,225.00	213,897.00	456,122.00

Ariel_SKU1000
Ariel_SKU1500
Ariel_SKU20
Ariel_SKU200
Ariel_SKU500

menu.

New Price	New Monthly Volume (MSU)	New Monthly Value (INR 000)	New Tide Overall Volume (MSU)	New Ariel Total Volume (MSU)	New Portfolio Volume (Tide + Ariel)
56.3	59,900.00	40,486.19	250,525.00	216,303.80	466,828.80

Volume change from price change	Price Change Index	New Volume as Index vs. Previous Volume	Monthly Value Change(INR 000)	New Volume as Index vs. Previous Volume	New Volume as Index vs. Previous Volume	New Volume as Index vs. Previous Volume
-56.00	1.08	0.999	2,963.93	1.03	1.01	1.02

	Competition Loss/Gains:	Volume Loss/Gain (MSU)	Market Share of Competitors in the Tier
Super Premium Tier	SurfExcel_SKU500	-33.60	100%
	Total Tier Change	-33.60	
Premium Tier	HENKO_SKU500	-1.40	6.23%
	RIN ADVANCED_SKU500	-8.04	35.88%
	SUNLIGHT_SKU500	-1.06	4.73%
	SURF EXCEL BLUE_SKU500	-2.74	12.23%
	TIDE_SKU500	-9.17	40.92%
	Total Tier Change	-22.40	

New Initiative Dummy	Current ACV	New Monthly Volume with Current ACV(MSU)	New Value with Current ACV
4	36.30	60,230.25	40,486.19

New Value with New Initiative	New Monthly Volume with a New Initiative(MSU)	New ACV	New Monthly Volume with New ACV(MSU)	New Value with New ACV
45,545.00	67,384.59	41.74	64,204.59	43,395.65

Value change from Initiative(INR000)	ACV Change Index	Value change from ACV(INR000)
2,149.35	1.15	2,909.46

Price Change Index	Initiative Change	ACV Change Index
0.999	1.125	1.150

Volume Change from price	Volume change from commercial innovation	Volume change from ACV	Total Volume Change
-56.00	7,484.6	3,974.3	11402.93

Value Change from price (INR 000)	Value change from a commercial innovation/launch (INR 000)	Value change from ACV(INR 000)	Total Monthly Value Change (INR 000)
2,963.93	2,149.35	2,909.46	8,022.74