Robust Price Optimization in Retail

Ketki Gupte [*1] **Linsey Pang** [*2] **Avinash Thangali** [*2]
Prakhar Mehrotra [2] **Harshada Vuyyuri** [1]

Abstract

At Walmart, our core mission is to help people save money so that they can live better.We accomplish this is by applying downward pressure on our prices in order to increase traffic and sales in our stores. Prior work has developed an automated process for optimal price recommendation (Linsey Pang et al.) including Bayesian Structured Time Series demand forecasting component.In this paper, we seek to extend the previous approach by incorporating robust optimization and an improved demand forecasting scheme with time-series clustering. The improved system is called Robust Price Recommendation System, or PRS+.

1. Demand Forecasting via Bayesian Structured Time Series

An essential component of quantitative price optimization is a model of the relationship between price changes and future demand volume. However, the demand of a particular item in a given price market depends not solely on its price, but also on other factors such as competitor prices, the difference between Walmart's prices and those of the overall market, sales, number of units sold in rest of the market, etc. Seasonality features also play an important role in influencing item demand.To generate the demand-volume forecasts as well as uncertainty estimates, we use Bayesian Structural Time Series (BSTS) (Scott & Varian, 2014),often used for feature selection, time series forecasting, and causal impact inference. In our model, we also impose a *Spike-and-Slab* prior on our regression coefficients, which enables automatic feature selection via parameter shrinkage (Ishwaran & Rao, 2005). We impose an inclusion probability of 1 on the price feature and 0.5 on all others. Additionally we set the elements of the prior mean vector to ± 0.5, with the sign of each element determined by the assumed directionality of the relationship between the corresponding feature and demand volume.Modeling was done at the item-week level. The details of this model can be found in (Linsey Pang et al.). For this improved version (PRS+), we also include time series clustering and cross-price elasticity features directly in our demand forecast model.

[*]Equal contribution [1]WalmartLabs, India [2]WalmartLabs, Sunnyvale, USA. Correspondence to: Ketki Gupte < Ketki.Gupte@walmart.com>, Linsey Pang < linsey.pang@walmart.com>, Avinash Thangali < Avinash.Thangali@walmartlabs.com>, Prakhar Mehrotra < Prakhar.Mehrotra@walmart.com>, Harshada Vuyyuri < Harshada.Vuyyuri@walmart.com>.

Proceedings of the 38th International Conference on Machine Learning, PMLR 139, 2021.

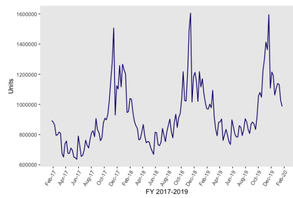
2. Time Series Product Clustering

Different product categories within our suite exhibit different demand structures over time. Figures 1(a)-1(d) show some examples of different temporal structures observed in different products, illustrating the need for specific tailoring of model formulations to different types of products. Developing a separate model for every product category we maintain would be intractable, so in order to accommodate these variations within our forecasts while also maintaining manageability, we group our categories into *clusters* based on the behavior of their demand series structure, and develop our BSTS demand forecasting models on a cluster-wide basis, although they are fit individually to each product. We assume that products within the same category (and by extension, each category cluster) will behave similarly. We make use of the k-means clustering algorithm utilizing a Dynamic Time Warp similarity metric (Giorgino, 2009).

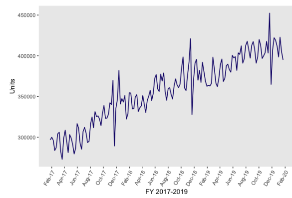
3. Robust Optimization

Despite the success of the traditional optimization frameworks with regression models, in this work, we seek to extend our previous approach by incorporating recent developments in robust optimization(Akihiro Yabe, 2017), replacing our previous formulation of the utility maximization step with a robust equivalent along with a robust quadratic optimization scheme discussed in (Akihiro Yabe, 2017).

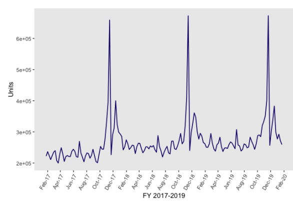
Let us have M products with data available for T time series data points. Let $Y_m = (y_1, y_2, ...y_t)$ be the time series of historical units sold of product $m \in M$ and $X_m = (x_1, x_2, ...x_t)$ be the historical price-series of the same products. We use our BSTS model to forecast the unit demand within given time period. Here we use the matrix form of BSTS with local linear trend and price related



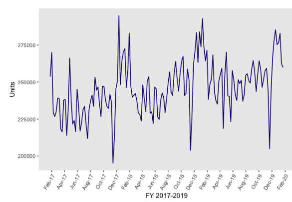
(a) Quadratic trend with seasonality



(b) Linear trend with seasonality



(c) Local level with holiday spikes



(d) Local level trend with seasonality

Figure 1. Examples of Different UPC Demand-Series Structures

features as an example to represent demand model:

$$y_t = Z_t\alpha_t + \epsilon_t = \begin{bmatrix}1 & 0 & x_t\end{bmatrix}\begin{bmatrix}\mu_t\\ \delta_t\\ \beta\end{bmatrix} + \epsilon_t \qquad (1)$$

$$\alpha_{t+1} = T_t\alpha_t + R_t\eta_t = \begin{bmatrix}\mu_{t+1}\\ \delta_{t+1}\end{bmatrix} \qquad (2)$$

$$= \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}\begin{bmatrix}\mu_t\\ \delta_t\end{bmatrix} + \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\eta_{\mu,t}\\ \eta_{\delta,t}\end{bmatrix} \qquad (3)$$

where $Z_t = (1\ 0\ \mathrm{x}_t)^T$, $\alpha_t = (\mu_t,\ \delta_t, \beta_t\ )$, $T_t = \begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$ $R_t = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$ and $\eta_t = (\eta_{\mu,t}, \eta_{\delta,t})^T$, β is the coefficient of regressor, and for generality, we assume it is constant through time x_t (Scott & Varian, 2014) (Kev).

For a given time t , we subtract out the time-series effect by letting $y_t^* = y_t - Z_t\alpha_t$ while keeping price related regressors x_t. In the following, we focus on modelling the uncertainty occurring in estimation of regressor coefficients (i.e. β). We take data points $(x_t, y_t^*) \in (X^T, Y^{*T})$ with one given single time step to illustrate our robust solution. For simplicity, we omit t in the following formulations.

In the demand forecasting model, conditional on spike and slab prior, the posterior distribution of the true coefficient β of the regressor component X is given in the form (Scott & Varian, 2014): $\beta \sim \mathcal{N}(\hat{\beta}, \sigma_\epsilon^2(V^{-1})^{-1})$, where matrix $V^{-1} = X^TX$. At given time t, we can express the forecast of y^* as a linear regression where A is a matrix composed of regressor coefficient β which relates the $X = (x_1, x_2 ... x_m)^T \in R^m$ prices of M products at given time t.

Similar to (Ito & Fujimaki, 2017), suppose we have training points with size $d \in D$ for each product $m \in M$ such that $X = (X_1, X_2, ..., X_d) \in R^{m\times d}$ with sales units as $Y_t = (y_1, y_2, ... y_m)^T \in \mathrm{R}^m$. The regression formulation is expressed as:$Y^* = AX$ we replace true A by it's estimator. This can in turn be simplified as $\hat{A} := \arg\min_A \sum_{d=1}^{D} y_d - Ax_d$. From Proposition 1 in (Ito & Fujimaki, 2017), we obtain $\hat{A}$ which follows matrix normal distribution as in the following: $\hat{A} = A^* + \Sigma^{*\frac{1}{2}} U_{M,M+1} V^{\frac{1}{2}}$ Where $U_{M,M+1}$ is random matrix over $R^{m\times(m+1)}$ and each entry $u_{i,j} \sim \mathcal{N}(0, 1)$. To derive a formulation for robust price optimization, C_λ, the confidence interval of $\hat{A}$ is determined by the estimator of Σ as $\hat{\Sigma}$ and derived as: $\hat{\Sigma} := \frac{1}{D}(Y - \hat{A}V)(Y - \hat{A}V)^T$ and $C_\lambda := \{A|\hat{A} + \hat{\Sigma}^{\frac{1}{2}} U_{M,M+1} V^{\frac{1}{2}}, U \leq \lambda\}$.

Therefore, we define the robust price optimization problem at given time t with matrix normal uncertainty as minimizing the uncertainty while maximizing the revenue as (Ito & Fujimaki, 2017): $f(x) = \underset{x\in X}{max}\ \underset{C_\lambda}{min}\ x^TAx \ \Leftrightarrow\ \underset{x\in X}{min}\ \underset{C_\lambda}{max}\ x^TQx$, where $\hat{Q} := \begin{pmatrix}-\hat{A}\\ 0\end{pmatrix} L_1 := \begin{pmatrix}\hat{\Sigma} & 0\\ 0 & 0\end{pmatrix} L_2 := V^{1/2}$

Using proposition 4 in (Ito & Fujimaki, 2017), for any $x \in X$ and $\gamma > 0$, it holds that $f(x) \leq g(x, \gamma)$ where $g(x, \gamma) := x^T\left(\hat{Q} + \lambda\frac{\gamma M_1 + M_2/\gamma}{2}\right)x$ and $M_1 := L_1^T L_1,\ M_2 := L_2^T L_2, \gamma L_1 X = L_2 X$

To solve equation above,the overall process can be summarized as:

- Extract observation matrix Z_t transition matrix T_t, $\sigma\ \ \epsilon$, for example for a local linear trend model as in eqn(2)
- Generate next state α_{t+1} of bsts at time $t+1$
- Generate prediction of y_t at $t+1$ using state at t
- Generate model intercept and coefficient of regression β from the state α_t and prediction y_{t+1} using matrix form of above equation

$$\beta(i,j) = \begin{cases}\text{model coefficient of cross elasticity for i,j}, & \text{if } i \neq j\\ \text{model coefficient of self price}, & \text{if i = j}\end{cases} \qquad (4)$$

After we obtain obtain intercept and coefficient of regression β from the model we generate Generate $V = XX^T$, $\hat{\Sigma}$ as described in section above). We also generate L_1, L_2, M_1, M_2, $\hat{\mathrm{Q}}$ as described in equations Next step is to check Check $\gamma > 0$, and initialize r $\leftarrow \infty$,$\tilde{\gamma} \leftarrow \gamma_0$

, $\tilde{x} \leftarrow arg\ min_x g(x, \tilde{\gamma})$. We then perform coordinate descent by checking condition $r - g(x, \gamma) > \delta$ and $\gamma \notin \{0, \infty\}$ and updating as

$r \leftarrow g(\tilde{x}, \tilde{\gamma})$

$\tilde{x} \leftarrow argmin_x g(x, \tilde{\gamma})$

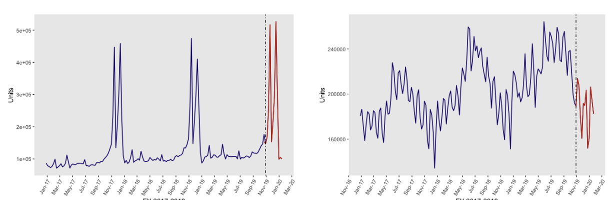
$\tilde{\gamma} \leftarrow argmin_\gamma g(\tilde{x}, \gamma)$

We perform the robust optimization step at each given time-step and update the optimal price obtained at time t to forecast the demand for time $t+1$, thereby improving our forecast. We perform a set of experiments to determine the value of λ in equation which maximizes our revenue.

4. Performance Evaluation

4.1. Forecast Evaluation

We first evaluate our system by demonstrating the accuracy of our forecasting step, augmented with a clustering preprocessing step. Table 1 displays the out-of-sample accuracy of our suite of models applied to 7 different categories and Fig. 2 contains plotsshowing the backtests of our model suite applied to two categoriesin particular within ourFood Consumablesdepartmen. Our models are not only numerically accurate but also capture well the seasonal and holiday effects that drive product sales.



(a) Backtest 1 (b) Backtest 2

Figure 2. Forecasting Performance

$Anonymized$	$Accuracy$	$Mape$	$ModelError$	$Cluster$
Category A	93.2%	6.80%	4 − 7%	$Linear\ Trend$
Category B	91.9%	8.10%	5 − 9%	$Local\ Level$
Category C	91.2%	8.80%	7 − 10%	$Holiday\ Spike$
Category D	90.3%	9.7%	9 − 11%	$Local\ Level$
Category E	89.5%	10.5%	8 − 11%	$Local\ Level$
Category F	88.4%	11.6%	10 − 13%	$Linear\ Trend$
Category G	87.2%	12.8%	11 − 13%	$Local\ Level$

Table 1. Forecast Performance on Price Market 11 for Anonymized Categories in Food and Consumables Department

4.2. Simulation Study

We applied a similar simulated data generation process as that of (Ito & Fujimaki, 2017)(Akihiro Yabe, 2017) to validate robust approach. Fig. 3 a shows experiments result, we randomly generated M true demand models with training datasets of $\{10M, 20M, 30M, 40M,\}$ with $M = 5$. The top horizontal line is the actual revenue yielded from our true data simulation, robust optimization model is executed for 4 different independent groups of products with $\lambda \in \{1, 2, ..., 20\}$. These results from Fig. 3(a) validate the effectiveness of robust approach: (1) The uncertainty of model error decreases when number of training data points increases. (2) Robust solution generates lower model error. (3) Solution also shows $\lambda \in \{2, 3\}$ resulting in much better performance comparing non-robust solution $i.e(\lambda = 0)$

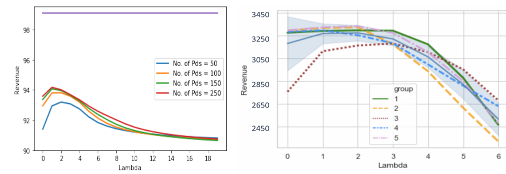


Figure 3. (a)Simulation study (b) Vacuum cleaner sales study

4.3. Vacuum Cleaners Sales Evaluation

The fineline which we analyze is *Upright Vacuum Cleaners*, which consists of 15 different products from different brands.We display the revenues yielded in 5 separate optimization runs on different groups of 15 products each, based on discount levels from their current prices. The values are plotted 3(b) for various values of λ, and their means and 95% confidence intervals are displayed. These charts illustrate how an optimal revenue is yielded with $\lambda = 2$. Table 2 shows the self-price and cross-price elasticities of individual upright vacuum cleaner among 15 products; notice that some products have a high cross price elasticity indicating that consumers view these as easily substitutable, while other with 0 elasticity are not sensitive to price changes. Table 3 illustrates the optimal prices obtained for these products with robust optimization solution. We observe that for λ values of 2 and 3 we take a deeper price cut on some products while keeping the existing prices for most other products. This suggests that we are close to optimal with our original PRS system recommendations but applying a price cut on some products gives higher revenues, illustrating the advantage of our improvements over the old framework.

5. Conclusion

In this paper, we extend our previous price recommendation approach (PRS) by incorporating a robust optimization step, as well as an improved demand forecasting scheme and we call it PRS+. PRS+ is able to take in as input a predetermined strategy aggressiveness parameter, λ, and generate a corresponding set of price recommendations for each item, in accordance with the risk tolerance encoded by λ. To our knowledge, our approach improves upon previous robust

Vacuum Cleaners		$p1$	$p2$	$p3$	$p4$	$p5$	$p6$	$p7$	$p8$	$p9$	$p10$	$p11$	$p12$	$p13$	$p14$	$p15$
Elasticity	$p1$	−0.878	0	0.425	0	−0.334	0	1.206	−0.211	0.056	−1.034	0	−0.202	0	−1.066	−0.045
	$p2$	0	−1.598	0.201	0	0.02	0	0.368	0.86	0.09	1.296	0	0.582	0	−0.785	1.66
	$p3$	0	0	−0.605	0	0.345	0	0.97	0.015	0.231	−0.003	0	0.977	0	−0.903	0.298
	$p4$	0	0	0.846	−0.217	0.263	0	0.086	−0.625	0.42	0.666	0	1.938	0	−1.892	1.12
	$p5$	0	0	1.154	0	−0.745	0	0.779	−0.028	−0.446	0.586	0	−0.367	0	−1.137	−0.348
	$p6$	0	0	1.241	0	0.119	−2.284	0.546	0.134	1.296	−1.936	0	−0.427	0	0.548	0.106
	$p7$	0	0	0.103	0	0.701	0	−0.237	−0.196	−0.063	−0.138	0	0.869	0	0.098	−0.143
	$p8$	0	0	2.437	0	1.075	0	1.049	0.202	0.634	−1.34	0	−2.072	0	0.355	−0.276
	$p9$	0	0	1.72	0	1.46	0	1.104	1.264	−2.44	1.494	0	0.858	0	−0.714	0.086
	$p10$	0	0	0.198	0	0.018	0	0.462	−1.154	1.05	−0.558	0	0.898	0	0.252	0.396
	$p11$	0	0	−0.561	0	0.022	0	0.572	0.766	0.027	0.874	−0.589	0.371	0	−1.746	1.334
	$p12$	0	0	1.598	0	−1.16	0	1.641	−0.506	0.242	−0.968	0	0.603	0	0.304	−0.446
	$p13$	0	0	0.19	0	0.747	0	0.674	0.337	0.791	0.268	0	2.311	−0.642	1.074	−0.458
	$p14$	0	0	0.77	0	0.225	0	1.559	−0.328	1.259	0.555	0	1.899	0	−1.115	−0.118
	$p15$	0	0	0.549	0	0.033	0	−0.184	0.144	0.11	−0.367	0	0.677	0	1.979	−0.086

Table 2. Vacuum Cleaners: Evaluation of the Robust output and price elasticity

Product Discount		$p1$	$p2$	$p3$	$p4$	$p5$	$p6$	$p7$	$p8$	$p9$	$p10$	$p11$	$p12$	$p13$	$p14$	$p15$
λ values	$\lambda = 1$	1	1	1	1	1	1	1	1	1	1	1	1	1	0.6	1
	$\lambda = 2$	0.817	1	1	1	1	1	1	1	1	1	1	1	1	0.6	1
	$\lambda = 3$	0.656	1	1	1	1	0.848	1	1	1	1	1	1	1	0.6	1
	$\lambda = 4$	0.6	1	1	1	1	0.683	1	1	1	1	1	1	1	0.6	1
	$\lambda = 5$	0.6	1	1	1	1	0.6	1	1	1	1	1	1	1	0.6	1
	$\lambda = 6$	0.6	0.951	1	1	1	0.6	1	1	1	1	1	1	1	0.6	0.926

Table 3. Vaccum Cleaners: Optimal product discounts at λ values

price optimization efforts (Akihiro Yabe, 2017) by better accounting for product substitution and retail demand complexities with a Bayesian Structual Time Series forecasting stage, rather than traditional regression methods used in similar robust optimization contexts.

References

State-space modelling. https://kevinkotze.github.io/ts-4-state-space/.

Akihiro Yabe, Shinji Ito, R. F. Robust quadratic programming for price optimization. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence Main track*, pp. 4648–465, 2017.

Giorgino, T. Computing and visualizing dynamic time warping alignments in R: The dtw package. *Journal of Statistical Software*, 31(7):1–24, 2009. doi: 10.18637/jss.v031.i07.

Ishwaran, H. and Rao, J. S. Spike and slab variable selection: Frequentist and bayesian strategies. *Ann. Statist.*, 33(2):730–773, 04 2005. doi: 10.1214/009053604000001147. https://doi.org/10.1214/009053604000001147.

Ito, S. and Fujimaki, R. Optimization beyond prediction: Prescriptive price optimization. *KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1833–1841, 2017.

Linsey Pang, K. G., Thangali, A., and Vuyyui, H. Price investment using. prescriptive analytics and optimization in retail. *KDD '20: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 3136–3144.

Scott, S. L. and Varian, H. Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5:4–23, 2014.