



Markdown Optimization with Generalized Weighted Least Squares Estimation

Mustafa Hekimoğlu¹

Received: 27 April 2022 / Accepted: 13 November 2022
© The Author(s) 2022

Abstract

Retailers increasingly apply price markdowns for their seasonal products. Efficiency of these markdown applications is driven by the accuracy of empirical models, especially toward the end of a selling season. In the literature, recent sales are recognized to be more important than older sales data for estimating the current period's demand for a given markdown level. The importance difference between the weeks of a selling season is addressed by weighted least squares (WLS) method with continuous weight functions of time. This study suggests a generalization of the weight functions and a method for optimizing their shape and discretization parameters to stimulate the predictive accuracy of models. We find that addressing the importance difference of recent sales observations using our generalized weight functions improves the forecast accuracy by up to 20%, and most of the improvement stems from our weight discretization method.

Keywords Markdown optimization · Weighted least squares · Pricing · Retail

Abbreviations

WLS	Weighted least squares
GWF	Generalized weight functions
GoF	Goodness of fit
AIC	Akaike information criterion
OLS	Ordinary least squares
MAPE	Mean absolute percent error
SKU	Stock keeping unit
WFS	Weight function by Smith et al. (1994)
WFK	Weight function by Keskin and Zeevi (2017)
GH	Greedy heuristic
R^2_{WLS}	Coefficient of determination for WLS (Willett and Singer 1988)
R^2_{WLS-p}	Pseudo coefficient of determination for WLS (Willett and Singer 1988)
ES	Exhaustive search

1 Introduction

Financial performance and stability of retailers mainly depend on their sales performance [16], and retailers focus on increasing their sale rates as much as possible by changing their price and inventory levels. In the retail sector, inventory decisions are made based on demand forecasts before the start of a selling season and actual sales may be lower than forecasted demand due to various factors such as large product variety, existence of competitors, short selling seasons of private labels and fashion merchandise [32]. This leads retailers to end up with unsold products at the end of a selling season. Unsold end-of-season merchandise is a significant problem for fashion retailers, as such products increase material handling costs for leftover inventory that is transferred to warehouses or donated to charities at the end of each selling season to make space for new-season merchandise.

To avoid material handling costs and revenue losses, retailers issue markdowns to stimulate their demand and finish their inventory as much as possible before the end of a selling season. In markdown decisions, remaining inventory level and in-season demand patterns play a crucial role. Usually, demand rates decrease toward the end of a selling season; hence, delaying markdowns can diminish the potential impact of markdowns on a retailer's profit. Similarly, early markdowns can increase discount costs of the retailer

✉ Mustafa Hekimoğlu
mustafa.hekimoglu@khas.edu.tr

¹ Department of Industrial Engineering, Kadir Has University, Cibali Mah., Hisaraltı Cad. No: 17, Fatih, 34083 Istanbul, Turkey

and lead to premature depletion of merchandise. In addition, the existence of markdowns changes characteristic demand patterns of products. Without markdowns, retail demand of a product can be characterized by a concave curve; sales peak in the middle of a selling season and decrease toward the end due to broken assortment and seasonal effects [1]. The existence of markdowns motivates some customers to behave strategically and delay their purchase to the markdown season to exploit lower prices [21]. This leads to lower early-season sales and a shifted demand peak. The effect of markdowns on the daily sales of a typical textile product is depicted in Fig. 1, where daily sales are marked with a line plot, and markdown levels are given with an area graph on the secondary axis over a 210-day selling season (between 10-02-2014 and 09-09-2014). In this example, markdowns increase, and the daily sales rate reaches its peak toward the end of the selling season. In addition, Fig. 1 depicts a hypothetical concave demand curve with a dashed line to emphasize the difference between sales rates with and without markdown applications.

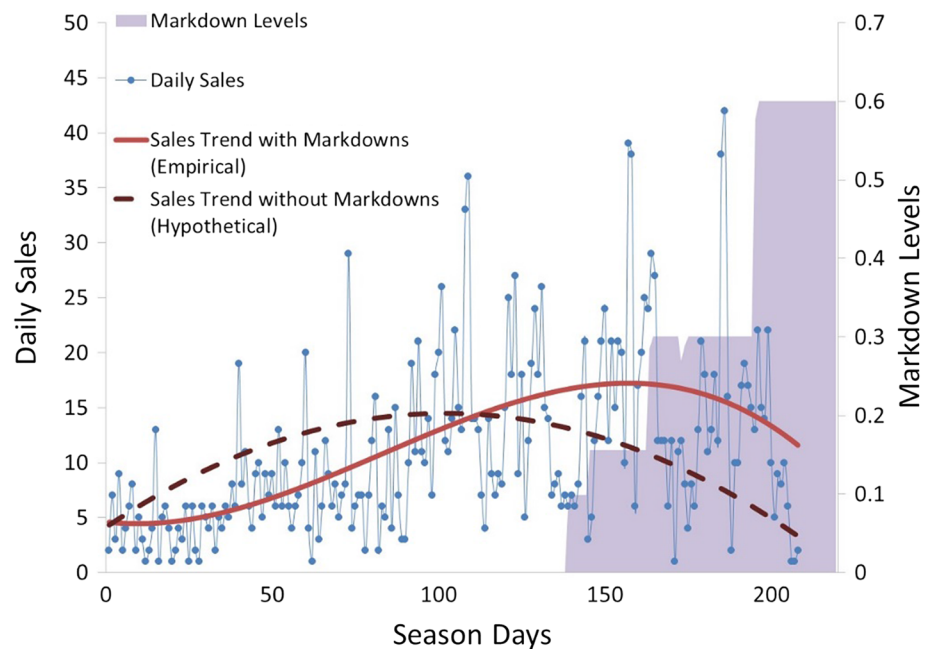
Markdown decisions significantly impact retailers' profitability and financial performance; therefore, they should be managed with a data-driven decision support system. Successful implementation of such a decision support system heavily depends on the accurate estimation of markdown sensitivity of demand over a selling season. This estimation requires a data set including quantity, selling price and inventory availability for each product. For forecasting future demand from historical sales data, recent observations are more reliable compared to the older periods [16, 35]. This importance difference between older and recent weeks of a selling season is operationalized through the weighted

least squares (WLS) method [6, 23, 35]. Scholars consider different mathematical functions to generate sequences of increasing weights toward the most recent observation.

In this paper, we suggest a demand modeling system for markdown optimization by generalizing the weight functions of Smith et al. [35] and Keskin and Zeevi [23] with the dyadic function [11]. Our generalized weight functions (GWFs) allow piecewise constant weight sequences for the parameter estimation of a regression model. To find the best weight sequence, we suggest an optimization procedure that searches over the shape and discretization parameters of GWFs using a goodness of fit (GoF) criterion. In our study, we consider three different GoF criteria, R^2 for WLS, pseudo R^2 for WLS, and Akaike information criteria (AIC), for the weight optimization. The first two are suggested by Willett and Singer [38] for WLS models, whereas the last is typical for choosing a demand model from a candidate set [15]. In addition, we consider a WLS approach from the robust regression literature. We test the contribution of these WLS methods to the prediction power of a demand model using empirical data from a Turkish apparel retailer. The accuracy of our methods is benchmarked against the ordinary least squares (OLS) method in terms of mean absolute percent error (MAPE).

Our results indicate that GWFs increase the accuracy of a multiplicative demand model by up to 20%, and this increase is primarily due to the discretization of weight functions. We find that the weight function by Smith et al. [35] is more effective than the one suggested by Keskin and Zeevi [23] when they are combined with discretization. This can be explained by the fact that Smith et al.'s formulation is more flexible and more suited to log-linear models. Our

Fig. 1 Impact of markdown on in-season sales of a product



experiments with the weight optimization procedure reveal that the AIC leads to more accurate demand models than the other two criteria from the WLS literature. We also find that the robust regression method provides inferior performance for a markdown optimization system. To the best of our knowledge, ours is the first study that provides a comprehensive evaluation of WLS methods for a markdown optimization system.

This paper consists of six sections. The next section reviews the extant literature. Section 3 presents the data set that we used to apply our method. Section 4 gives the details of our model and generalized weight functions. Section 5 presents the results of the study and Sect. 6 summarizes and concludes the paper.

2 Literature Review

The relevant literature for this study consists of two main streams. The first research stream consists of studies focusing on markdown optimization. We consider those markdown studies from the demand modeling perspective. Second, we consider studies on the WLS approach for the parameter estimation of demand models.

The markdown optimization literature includes many studies targeting products with a low end-of-season value, such as apparel, holiday presents, or perishable products. For the apparel sector, Smith et al. [35] consider a markdown optimization problem by considering a two-stage demand modeling approach. Their approach is extended by Smith and Achabal [33] and Caro and Martínez-de Albeniz [6]. In those studies, the log-linear demand model is considered to capture the customers' sensitivity to markdown levels. Studies by Harsha et al. [18] and Chen et al. [9] utilize demand models other than multiplicative equations. Harsha et al. [18] consider a multinomial logit model for customers' preferences and utilize the model predictions for predicting demand, whereas Chen et al. [9] utilize the gradient boosting algorithm combined with a differential equation-based parameter estimation to optimize clearance prices of Walmart over a selling season.

From the demand modeling perspective, interactions between different products are also important for customer demand [14] [13]. Specifically, Cosgun et al. [13, 14] utilize approximate dynamic programming to optimize the markdown levels of multiple products using a linear demand model. Wang et al. [37] consider markdown optimization of holiday items left over after the Christmas period. They state that decreasing values of products require sharp markdowns at sales price to remove inventory from retail stores. Rice et al. [30] compare markdowns with probabilistic selling, which they define

suggesting different options to customers to augment price discrimination, using a stylistic model. They conclude that probabilistic selling has a potential of achieving higher revenue and better inventory utilization. Recently, markdown optimization models have been applied to perishable products, which lose almost the entire value at the end of their shelf lives. In those studies, researchers mainly consider one-time discount approach to stimulate demand [2, 8, 10, 25, 27, 29]. Single discount applications are motivated by the fact that subsequent markdown applications suffer from diminishing rate of returns.

The second relevant research stream consists of the studies on the usage of WLS estimation of parametric demand models. Smith et al. [35] and Smith and Achabal [33] use a two-stage demand procedure to estimate the discount sensitivity of customer demand. They utilize a weight function to assign more importance to recent periods' sales. Their studies are extended by Caro and Gallien [5], who consider a different weighting scheme to assign higher importance to recent observations. Manning and Mullahy [28] consider the WLS approach for log-linear models for health economic estimations in the existence of heteroscedasticity and skewed coefficients. Keskin and Zeevi [23] consider WLS for estimating the parameters of a linear demand function used in price optimization. Their algorithm assigns exponentially decaying weights to old observations, interpreted as information depreciation.

Another interesting usage of WLS-based parameter estimation is robust regression [4, 7, 36] and locally weighted regression [12, 26, 31]. In robust regression studies, WLS estimation is used to discount leverage points in covariate values. In many cases, empirical data might suffer from uncontrollable effects and measurement errors [4], creating bias in the parameter estimation procedure. Researchers suggest various statistical procedures that assign lower weights to leverage points in the parameter estimation of regression models.

The closest studies to our paper are given by Smith et al. [35], Smith and Achabal [33], Caro and Gallien [5], and Keskin and Zeevi [23]. We extend those studies by considering generalized weight functions (GWFs) with a hyper-parameter search to obtain the best model using different quality of fit criteria. Our approach generalizes the WLS methods in the literature by a parametric discretization method. Note that we chose two well-known weight functions from the literature to demonstrate the impact of our approach, whereas it can be applied to other continuous, weight-generating functions. We also compare our GWFs to different WLS algorithms from the robust regression literature [38]. To the best of our knowledge, ours is the first study evaluating robust regression models for a markdown optimization problem.

3 Point-of-Sale Data

We consider point-of-sale data from a large apparel retailer in Turkey. The data set consists of 10,770 rows and 15 columns. The columns of the data set include sales date, sales quantity, product code, product group, gender, collection code, brand code, season code, sales price, and list price. Rows of the data set consist of daily sales of 312 unique product codes sold between 2011 and 2015. The distribution of the products over the years is approximately equal (18, 20, 22, 24, 17 percent). Most of the products (95%) are sold out at the end of their first year on shelf and never returned to stores. Only 16 (out of 312) products are kept on shelf over more than 1 year. Such products are removed from the store and sent to the warehouse until the next year's selling season.

Each product code is a member of a product collection, which is used as the key information for modeling customer demand in our study. Specifically, the retailer makes markdown decisions based on product collections including products with similar markdown sensitivity of demand. In our data set, there are 19 different collection codes. Three of these product collections consist of a single product code, whereas for the rest, each collection includes more than 20 different product codes. A more detailed description of the data set is provided by Yıldız [39].

Furthermore, start dates and durations of markdown periods vary significantly within each collection. We manually break down collections into subcollections to ensure that products with the same start and end points are grouped and subjected to the same markdown levels. We define 81 subcollections and %65 of them include less than three different products. Such a granular approach to the data set is motivated by achieving higher accuracy in our demand modeling study. In this study, we fit a different log-linear regression model to each subcollection. Also, the forecast accuracy of each model is reported at subcollection level to describe the impact of weight optimization on forecasting performance.

3.1 Data Enrichment

In addition to the covariates explained above, we extract additional features proven to be useful for estimating customer demand. One of the extracted features is the number of days since the last change of the markdown level. It is known that customers' response to any price promotion exponentially decreases over time as people get accustomed to the discounted price. This can also be explained with the reference price theory, in which each customer is assumed to have a reference for the product and decides if the selling

price is lower than his or her reference [17]. After the price discount, customer demand surges as there is a significant difference between the selling price and customers' reference prices. Over time, reference price levels are adjusted, so the sales rate goes back to its normal trend. We calculated the average number of weeks since the last markdown update for each product to cover this time-dependent effect.

Seasonal sales trend for retail products is another extracted feature from our data set. Specifically, empirical evidence from the retail management literature suggests that customer demand starts slowing at the beginning of a selling season [32]. Toward the mid-season periods (weeks), sales rates reach its peak and start to decline due to the broken assortment effect [34], or season change. To capture this in-season dynamics we define the following covariate: $T_i^t := \frac{(t_i - t_i^{mid})^2}{\bar{T}_i} \forall i \in \mathcal{P}^k$ where t_i is the week index of the product i , and t_i^{mid} is the middle of the selling season. \bar{T}_i stands for the length of the selling season for product $i \in \mathcal{P}^k$, where \mathcal{P}^k is the set of products in the subcollection $k \in \mathcal{K}$, and \mathcal{K} is the set of all subcollections. Note that the lengths of products' selling seasons might vary due to various factors, such as different introduction dates and in-store retail practices. To address this variability, we used normalization using the length of the selling season. The effect of this covariate is depicted with a dashed line in Fig. 1. In the following section, these two covariates are utilized in a multiplicative regression model to estimate customer demand for given markdown levels using a WLS method.

4 Demand Modeling with WLS Estimation

To model customer demand and its markdown sensitivity, we consider a log-linear demand model and a WLS parameter estimation procedure. In this section, we first present the demand model in Sect. 4.1. Next, the WLS estimation method and a parameter update method are described (Sects. 4.2, 4.3). In the last part of this section, resampling studies are presented (Sect. 4.4).

4.1 Multiplicative Demand Model

Nonlinear, multiplicative regression models are a common approach to establish the relationship between selling price, discount, and customer demand. Using the log transformation on demand, one can estimate the model coefficients using the linear regression methodology. Customer demand is affected by many different factors such as seasonality, the week of a selling season, the existence of special days, and

price discounts. It is crucial to consider all relevant covariates for a successful demand model.

To predict weekly customer demand using the log-linear model, we consider both quantitative and qualitative covariates in this study. Let us define B^i and G^i as the brand and gender codes for the product (code) i . M_t^i is the markdown level given below.

$$M_t^i = \frac{P_0^i - P_t^i}{P_0^i},$$

where P_0^i is the sales price for product $i \in \mathcal{G}^k, k \in \mathcal{K}$ at the beginning of its selling season. W_t^i measures the number of periods since the date of the last markdown change. Using these covariates, we suggest the following model for weekly demand during a markdown period.

$$\log(D_t^i) = \tilde{\beta}_0^i + \tilde{\beta}_1 \log(1 - M_t^i) + \tilde{\beta}_2 W_t^i + \tilde{\beta}_3 T_t^i + \tilde{\beta}_4 B^i + \tilde{\beta}_5 G^i + \tilde{\beta}_6 \log(D_{t-1}^i) + \epsilon_t, \tag{1}$$

for $1 \leq t \leq \tau^i$, where $D_0^i = 0$ and τ^i represents the total number of weeks in the markdown season for product $i \in \mathcal{G}^k, k \in \mathcal{K}$. The model in (1) assumes the first-order autocorrelation for the demand variable, which is a reasonable assumption. In addition, we consider product-specific intercept, denoted by β_0^i , to cover the average sales of each product during a selling season.

For the estimation of model parameters, we run a coefficient estimation procedure combined with the WLS method. Our method consists of three main calculation steps: (1) calculate the best parameters for the GWFs at the beginning of a selling season; (2) at the beginning of each week, calculate the WLS estimate for the coefficients of the model in (1); (3) update the coefficients of the prediction model using the exponential smoothing in Sect. 4.2. The overall algorithm, including these calculation steps, is presented in Algorithm 1.

Algorithm 1 Overall Algorithm for the Parameter Update Procedure

- 1: Optimize parameters of the selected GWF.
 - 2: **while** $1 < t \leq \tau_t$ **do**
 - 3: Estimate $\tilde{\beta}_{0t}^i$ and $\tilde{\beta}_{jt}^i$.
 - 4: Update the coefficients of the prediction model (9)-(10).
 - 5: **end while**
-

Note that the exponential smoothing in Step 4 of Algorithm 1 is suggested by Caro and Gallien [5] and Smith et al. [35] for markdown optimization systems. We extend those

studies with the weight optimization using GWFs presented in the following section.

4.2 Weight Optimization with GWFs

Markdown systems are utilized for a wide variety of products with different characteristics. For most of them, assigning higher importance to recent observations is crucial for a successful implementation. As weight assignment to recent sales data mainly depends on the product characteristics [35], we suggest that the weight vector of WLS should be optimized at the beginning of a markdown period using a parametric GWF. To this end, we consider two different weight functions, by Smith et al. [35] and Keskin and Zeevi [23], and suggest a generalization scheme using the dyadic function. The resulting GWFs are optimized using three distinct GoF criteria. Weight functions and the considered GoF criteria are presented in the following subsections.

4.2.1 Weight Function by Smith et al. [35] (WFS)

The first weight function that we consider is suggested by Smith et al. [35] (from this point, it is referred to as WFS). WFS, denoted by $w_\alpha^S(t)$, yields a weight value for the week t of a selling season with the following function:

$$w_\alpha^S(t) = (1 - \alpha)^{2(\bar{t}-t)}, \tag{2}$$

where \bar{t} is the total length of the selling season, t is the current period (week) and $\alpha \in [0, 1]$ is the *shape* parameter. $w_\alpha^S(t)$ is a continuous, increasing function of t , so it assigns higher weights toward the end of a selling season (Fig. 2a). Also, $w_\alpha^S(t)$ is decreasing in α . When α is close to one, WFS assigns more weights toward the end of the season, whereas values close to zero assume more weights for later weeks. So, one can control the weight assigned to recent sales of a product by manipulating the shape parameter of $w_\alpha^S(t)$ (Fig. 2a).

4.2.2 Weight Function by Keskin and Zeevi [23] (WFK)

Keskin and Zeevi [23] suggest another nonlinear function family, which we call Weight Function by Keskin and Zeevi [23] (WFK) that assigns higher weights to the most recent observations to optimize data-centered pricing strategy of a retailer. They prove that (3) gives the optimal weights to estimate the parameters of a linear demand–price curve. In this study, we evaluate this function family for the weight optimization of a WLS-based parameter estimation method. WFK is given as follows:

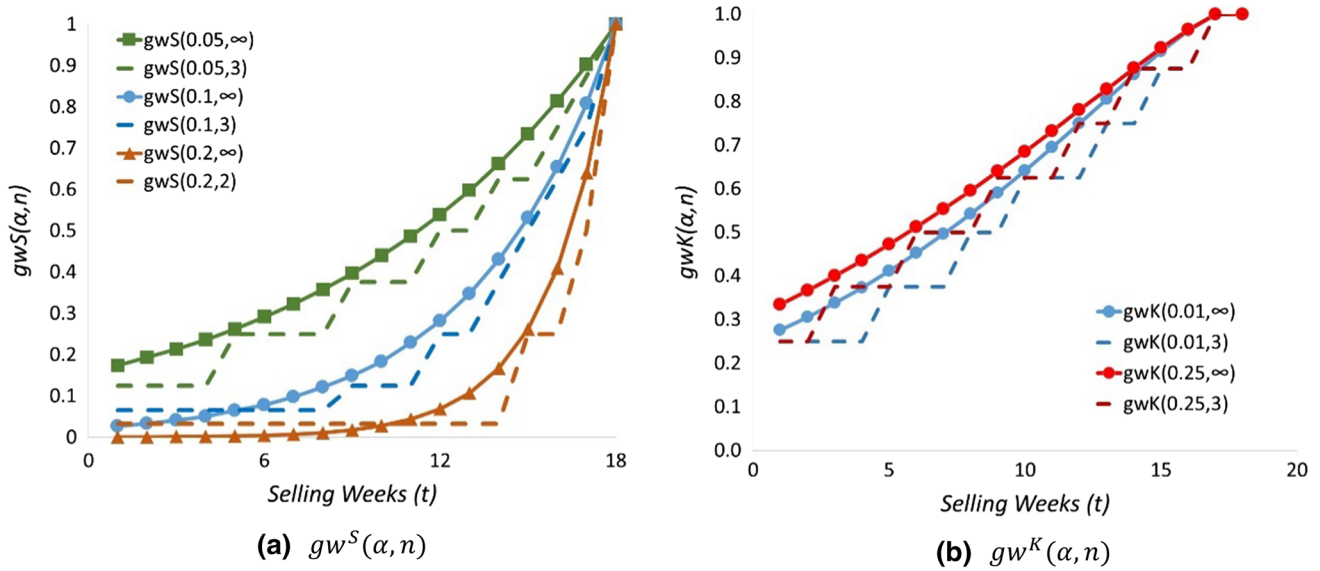


Fig. 2 Weight functions by Smith et al. [35] and Keskin and Zeevi [23] and their generalizations

$$w_\alpha^K(t) = \left(1 - \frac{(\bar{t}-t)}{m^2} + \frac{(\bar{t}-t)^{1-\alpha}}{m^2} \right)^{\frac{1}{\alpha}}, \quad (3)$$

where $m := \lceil \kappa T^{\frac{1}{3}} \rceil$, and α is the shape parameter of the function. In this study, we set κ to 2 and assume T is the largest planning horizon in a subcollection. $w_\alpha^K(t)$ is monotonically decreasing in α . In our study, we optimize the shape parameters of WFS and WFK to obtain the optimum weight sequences for the product after discretizing them with the dyadic function.

4.2.3 Dyadic Function for Discretization

The weight functions in (2) and (3) are continuous increasing functions of t . Utilizing these functions in markdown optimization implicitly assumes that the importance of past sales data continuously decreases for earlier sales date. On the other hand, in some markdown applications, some weeks have the same level of importance from the perspective of discount sensitivity and equal weight values are needed for WLS-based coefficient estimation. This requires the discretization of the weight functions in the form of a step function. To this end, we consider the dyadic function, which approximates any continuous function with a piecewise constant function in any desired accuracy. The dyadic function, denoted by $d_n: [0, \infty) \rightarrow [0, \infty)$ (cf. Çınlar [11]), is defined as follows:

$$d_n(r) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{[(k-1)2^{-n}, k2^{-n})}(r) + n 1_{[n, \infty)}(r), \quad (4)$$

where r is any continuously increasing function and $1_{(a,b)}(r) := 1\{a < r < b\}$ stands for the indicator function. By using the dyadic function with $w_\alpha^S(t)$ and $w_\alpha^K(t)$, we obtained the following GWFs, which are non-decreasing, right continuous, piecewise constant for any $n < \infty$.

$$gw^S(\alpha, n)(t) := d_n(w_\alpha^S(t)), 0 \leq t \leq \bar{t}, n \in \mathbb{N},$$

$$gw^K(\alpha, n)(t) := d_n(w_\alpha^K(t)), 0 \leq t \leq \bar{t}, n \in \mathbb{N}.$$

Recall that $gw^S(\alpha, \infty)(t) = w_\alpha^S(t)$ and $gw^K(\alpha, \infty) = w_\alpha^K(t)$, meaning that the continuous weight functions $w_\alpha^S(t)$ and $w_\alpha^K(t)$ are special cases of $gw^S(\alpha, n)(t)$ and $gw^K(\alpha, n)(t)$, respectively. Also, for every $n \in \mathbb{N}$, we have the following approximation gap of discretization:

$$\|gw^i(\alpha, n)(t) - gw^i(\alpha, \infty)(t)\|_\infty \leq 2^{-n}, \forall t \geq 0,$$

where $i \in \{K, S\}$ and $\|f\|_\infty = \sup_{t \in [0, \bar{t})} |f(t)|$. The approximation gap is bounded by a decreasing function of n , which we refer to as discretization parameter. Realizations of $gw^i(\alpha, n)$, $i \in \{K, S\}$ for different shape and discretization parameters (α and n) are depicted in Fig. 2a and b.

To find the best weight sequence, we utilize a heuristic search over the parameters of $gw^S(\cdot)$ and $gw^K(\cdot)$ functions. Specifically, the search algorithm utilizes a single

dimensional greedy heuristic for $\alpha \in [0, 1]$ for a fixed n , $1 \leq n \leq \bar{n}$. Preliminary experiments on the dyadic function reveal that $n = 10$ approximates the continuous weight function close enough (Fig. 2a). Therefore, we set \bar{n} to 10 and run a single dimension greedy heuristic over $\alpha \in [0, 1]$ for each $n \in \{2, \bar{n}\}$. The search algorithm is restarted from 20 different random points of $\alpha_0 \in [0, 1]$ to find $\alpha^*(n)$. Then comparing different models with $(n, \alpha^*(n))$, $n \in \{1, \dots, \bar{n}\}$ leads to the optimal weight parameters $(n^*, \alpha^*(n^*))$ with respect to a given quality of fit criterion.

4.2.4 Goodness of Fit Criteria for Model Evaluation

To optimize the weight sequences, we consider three alternative GoF criteria for comparing regression models with different weight sequences. The first criterion is suggested as a replacement for the coefficient of determination for parameter estimation with WLS (Willett and Singer [38]). To define the coefficient of determination for WLS, denoted by R^2_{WLS} , Willett and Singer [38] define the diagonal weight matrix \mathbf{W} . The dimension of the matrix is m by m , where m is the length of the dependent variable \mathbf{Y} . i -th diagonal element of \mathbf{W} is defined as $w_{ii} = w_{\alpha,n}(t_i)$, where t_i is the selling week in row i of the covariate matrix \mathbf{X} . Note that \mathbf{Y} is $\log(D_i^i)$, $i \in \mathcal{G}$ in (1). Using \mathbf{W} , we define the transformed variables $\mathbf{Y}^* = \mathbf{W}^{-1/2}\mathbf{Y}$ and $\mathbf{X}^* = \mathbf{W}^{-1/2}\mathbf{X}$. Then,

$$R^2_{WLS} = 1 - \left[\frac{(\mathbf{Y}^* - \mathbf{X}^* \beta^*)^T (\mathbf{Y}^* - \mathbf{X}^* \beta^*)}{\mathbf{Y}^{*T} \mathbf{Y}^* - m \bar{\mathbf{Y}}^{*2}} \right], \tag{5}$$

where $\bar{\mathbf{Y}}^*$ is the average value of \mathbf{Y}^* and β^* is the coefficient of WLS estimation:

$$\beta^* = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}. \tag{6}$$

The second alternative GoF criterion that we consider is the pseudo coefficient of determination for WLS, denoted by R^2_{WLS-p} [38]. This statistic is suggested as a replacement of the classic coefficient of determination in OLS in case of WLS. R^2_{WLS-p} is given below:

$$R^2_{WLS-p} = 1 - \left[\frac{(\mathbf{Y} - \mathbf{X} \beta^*)^T (\mathbf{Y} - \mathbf{X} \beta^*)}{\mathbf{Y}^T \mathbf{Y} - m \bar{\mathbf{Y}}^2} \right]. \tag{7}$$

The motivation behind this metric is to report the fraction of variability explained by the model in the original space, instead of the one transformed with \mathbf{W} matrix [38].

The third alternative metric for comparing regression models is Akaike information criterion (AIC), a common maximum likelihood-based statistic used in different

applications of data mining and statistical machine learning. AIC is suggested for WLS by de Brauwere et al. [15].

In addition to the weight optimization with GWFs and different GoF criteria, we also consider weight schemes from the robust regression literature to benchmark its performance to our GWFs. The robust regression method, which we consider in this study, is summarized in the following section.

4.2.5 Weight Functions of Robust Regression

Robust regression is another stream of research that assigns varying weights to different observations in a data set is robust regression. In this study, we consider a leverage-based weighting scheme that uses the diagonal elements of the projection matrix defined as $\mathbf{P} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ [7]. Observations in the \mathbf{X} -space, which generates too much leverage on the fitted regression curve, is called high-leverage points. To eliminate the effects of such points, Chatterjee and Mächler [7] suggest an iterative algorithm that calculates the weight vector for WLS. To this end, they define

$$w_k^j = \frac{(1 - p_{kk})^2}{\max \left\{ \left| e_k^{j-1} \right|, m^{j-1} \right\}}, \tag{8}$$

as the k -th element of the weight vector at iteration j . In (8), $e_k^j = y_k - X \hat{\beta}^{j-1}$ is the residual value of observation k at iteration j and m^j is the median of (e_1^j, \dots, e_n^j) vector. The iterative algorithm starts with $w_k^0 = 1 / \max \{ p_{kk}, p/n \}$. We compare the impact of the leverage-based weights on the prediction power of the model (1) in Sect. 5.

4.3 Updating Parameters of the Prediction Model

In this study, we suggest a parameter estimation and an update method for estimating weekly customer demand within a markdown season of products. In our model, regression coefficients are estimated using the WLS method with GFWs. These estimated coefficients are used to update the prediction model's coefficients using the exponential smoothing method as follows:

$$\beta_{0t}^i = (1 - \gamma) \beta_{0,t-1}^i + \gamma \tilde{\beta}_{0t}^i, \tag{9}$$

$$\beta_{jt} = (1 - \gamma) \beta_{j,t-1} + \gamma \tilde{\beta}_{jt}, \tag{10}$$

for $j \in \{1, \dots, 6\}$, $i \in \mathcal{G}$ and $t \geq 2$. $\tilde{\beta}_{0t}^i$ and $\tilde{\beta}_{jt}$ are the WLS estimates of the coefficients β_0^i and β_j in (1) for week t . Recall that in our two-stage demand forecasting method, WLS estimates of coefficients are calculated for every week t . Those weekly estimates are used to update the actual model parameters β_{0t}^i and β_{jt}^i to forecast demand for week $t + 1$ for a given

markdown level $M_{t+1}^i, i \in \mathcal{I}$. In our estimations, we set γ parameter to 0.2. Note that our choice of γ is also utilized by Smith et al. [35]. We prefer using updated coefficients $(\beta_{0t}^i, \beta_{jt}^i)$ over WLS estimates $(\tilde{\beta}_{0t}^i, \tilde{\beta}_{jt}^i)$ to control variance of estimates and to avoid overfitting-related problems. By doing so, we accept some extra bias in our demand forecasts to control estimation variance. Our results in Sect. 5.2 reveal that the variance control and overfitting avoidance using exponential smoothing significantly pay off in terms of forecast accuracy.

4.4 Performance Tests with Resampling

Once the coefficient estimation and the parameter update steps are complete, we obtain the demand forecast of period t using β_{0t}^i and β_{jt}^i coefficients. We test our prediction method from two different perspectives. First, we evaluate the accuracy of the model using rolling-origin-recalibration evaluation [3]. Second, we test our model against the possibility of overfitting using a bootstrapping method specifically designed for time series data [20].

4.4.1 Rolling-Origin-Recalibration Method for Model Evaluation

To evaluate the prediction power of forecasting models, different methods are suggested in the literature. One way of classifying these evaluation methods is based on the last observation in the training set, which is defined as forecast origin. For time series data, Bergmeir and Benitez [3] suggest the rolling-origin-recalibration method, which proceeds in two steps: first, for a given forecast origin and a fixed number of observations (forecast horizon) predictions are made and compared with the observations in the test set to calculate deviation; second, the first week of the test set is added to the training set and is taken as the new forecast origin, and the forecast model is recalibrated. In our study, we adopted the rolling-origin-recalibration method for the estimation of markdown-driven customer demand. We take 80% of available data as a training set for each subcollection at the beginning of the evaluation. Starting from the first week of a test set, we generate predictions from our demand model and measure the prediction accuracy. Afterward, the forecast origin is added to the training set and the parameters of the log-linear model are recalculated. The measures that we used for prediction accuracy are explained in Sect. 4.4.3.

4.4.2 Bootstrapping Tests Against Overfitting

Estimating the prediction error of a model and checking against any possible overfitting is challenging for any problem with a limited sample size. It is recognized that

low training set error, high test set bias, and large response variance are primary indicators of overfitting in predictive models [22]. Specifically, the test set error can be decomposed into the sum of prediction bias, variance of model response and irreducible error. To obtain the best prediction accuracy, it is important to find a low-bias and low-variance model. High response variance is an essential indicator of overfitting as small changes in the training set lead to large fluctuations in model estimates and lower test set accuracy.

Bootstrapping is suggested for estimating model accuracy and response variance when the data set size is limited. Bootstrapping method relies on simulating a new data set from the original data by random sampling with replacement. Bootstrap samples can be used to estimate the variance of a model's response or its prediction bias [19]. The standard bootstrapping method of sampling with replacement creates biased estimates of population features when it is applied to time series data with possible autocorrelation. In such a case, moving block bootstrapping with overlapping samples is suggested by Kunsch [24] and Li and Maddala [20]. In this method, the data set is divided into blocks of size l . For a data set of $\{x_1, x_2, \dots, x_n\}$ and $l = 3$, moving bootstrap blocks are $\{\{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}, \{x_3, x_4, x_5\}, \{x_4, x_5, x_6\}\}$. To apply bootstrapping, b of these blocks is sampled with replacement. Kunsch [24] suggested the following formula for the optimal block size, l^* .

$$l^* = n^{1/3} \rho^{-2/3},$$

where $\rho = (1 - \phi^2)/\phi$, ϕ is the coefficient of AR(1) process and n is the sample size of the original data. Li and Maddala [20] suggest that $b = n/l^*$ for the number of bootstrap blocks.

In this study, we applied the moving block bootstrapping method for our markdown demand modeling problem. We keep the weeks with no markdown in the training set and create bootstrapping blocks starting from the first week of the discount season. Each bootstrap sample is combined with the training set to estimate prediction error and variance of model response using the rolling-origin-recalibration scheme described above. For the block size calculation, we estimate $\hat{\phi}$ and l^* values by fitting a separate AR (1) model to log-demand observations of each subcollection.

4.4.3 Measures for Prediction Accuracy

The accuracy of our model's forecasts is tested using the evaluation method in Sect. 4.4.1. At each time point, the difference between the forecasted and actual sales value is

evaluated using mean absolute percent error (MAPE) for each subcollection k , denoted by $MAPE^k$. The description of subcollection is given in Sect. 3.

$$MAPE^k = \frac{1}{T_k} \sum_{t=1}^{T_k} \left[\frac{1}{N^k} \sum_{i=1}^{N^k} \frac{|\hat{D}_t^{i,k} - D_t^{i,k}|}{D_t^{i,k}} \right], \tag{11}$$

where N^k stands for the number of SKUs and T_k is the length of time series in subcollection $k \in \mathcal{K}$. Model performances are presented in terms of subcollection averages, e.g., $MAPE_{agg} = \sum_{k \in \mathcal{K}} MAPE^k / |\mathcal{K}|$ in the following section.

Recall that our subcollection-based rolling-origin-recalibration scheme and moving block bootstrapping tests are specifically modified to address the needs of markdown optimization systems, which are usually run for a group of products that have similar discount sensitivity. Each markdown update requires manual labeling for existing products with new price and discount tags, and grouping products significantly lowers the operational complexity of markdown systems (Caro and Martínez-de Albéniz [6]).

5 Results and Discussion

This study suggests a weight optimization procedure for the WLS-based estimation of a demand forecasting model. We suggest two different GWFs and three GoF criteria for the weight optimization, which are evaluated using empirical sales data of a Turkish apparel retailer. We present the performances of these weight optimization alternatives relative to the models with OLS parameter estimation. Using resampling tests, we investigate the impact of the following factors of the weight optimization system on the prediction accuracy:

- A) Weight discretization and optimization.
- B) Goodness-of-fit criteria for weight optimization.
- C) A single-stage parameter update procedure.
- D) Optimization of weight sequences at every period.

To understand the impact of (A), we compare WFS and WFK functions with and without the discretization in (4). For each subcollection weight vectors from $gw^S(\alpha^*(\infty), \infty)$, $gw^S(\alpha^*(n^*), n^*)$, $gw^K(\alpha^*(\infty), \infty)$ and $gw^K(\alpha^*(n^*), n^*)$ functions are utilized to estimate model coefficients in (1). In addition, we consider the leverage-based weight determination method in Sect. 4.2.5. Impact of (B) is considered with the implementation of three different criteria for the weight optimization. The main experimental design that includes these WLS components with 42 experiments is presented in Table 1. In addition, the experimental factors in Table 1 are extended with a single-stage parameter update procedure

Table 1 Main different experimental factors

Weight optimization αn	GoF criteria	Interpretation
OLS	∞	Benchmark
GH	ES R^2_{WLS}	Joint optimization
GH	ES R^2_{WLS-p}	Joint optimization
GH	ES AIC	Joint optimization
GH	∞ R^2_{WLS}	Optimization over α
GH	∞ R^2_{WLS-p}	Optimization over α
GH	∞ AIC	Optimization over α
{0.1,0.15,0.2}	∞	Continuous GWFs with constant α
{0.1,0.15,0.2}	ES R^2_{WLS} R^2_{WLS-p} AIC	Optimization over n
LB		Leverage-based weights

($\gamma = 1$) (C) and weight optimization at every period (D). The total amount of experiments considered in our study is 126. The implementation details and the results of our analyses are discussed in the following subsections.

5.1 Impact of Weight Optimization with Discretization

To evaluate the impact of the weight optimization on prediction accuracy, we optimize GWFs' shape and discretization parameters. The shape parameter, α , is optimized with the greedy heuristic (GH) over [0.1, 0.6] for a given $n \in \{1, \dots, \bar{n}\}$. This search leads to the best weight parameter for a given n , denoted by $\alpha^*(n)$. Then the best parameter set is found using an exhaustive search (ES) over $n \in \{1, \dots, \bar{n}\}$ values, where $\bar{n} = 10$, as it is almost equal to the continuous weight function (2). In the parameter optimization of weight vectors, we conduct runs with the three GoF criteria in Sect. 4.2.4.

In addition to the joint optimization, we investigate the impact of optimization over n for a given $\alpha \in \{0.1, 0.15, 0.2\}$ (Table 1). These values of α are suggested by [35] for markdown optimization of an apparel retailer. Therefore, our experimental design extends their work by comparing their setup with new features. Also, the performances of $\alpha^*(\infty)$, continuous weight functions (2) and (3), are considered by setting $n = \infty$ in $gw^S(\alpha, n)$ and $gw^K(\alpha, n)$. In addition to the weight functions, we evaluate the robust regression method in a markdown setting. The results of our experiments are given in Table 2, where the optimized weight function parameters are indicated with *.

Our results reveal that the best performance is obtained from the joint optimization over α and n using the AIC criterion. It creates 21% of additional accuracy in terms $MAPE_{agg}$ compared to OLS. The improvement over OLS

Table 2 Relative performance of WLS and WFK methods in terms of $MAPE_{agg}$ for $\gamma=0.2$ (performance of OLS is set to 100, WFK is expressed in brackets)

Weight optimization		Optimization criteria		
<i>A</i>	<i>N</i>	R^2_{WLS}	R^2_{WLS-p}	AIC
$\alpha^*(n^*)$	n^*	> 1000 (95.6)	84.4 (92.1)	78.9 (89.7)
$\alpha^*(\infty)$	∞	> 1000 (99.2)	130.0 (96.2)	112.1 (95.8)
0.10	n^*	83.6 (91.4)	85.5 (90.5)	85.1 (88.5)
0.15	n^*	88.6 (92.1)	83.5 (90.7)	87.4 (88.6)
0.20	n^*	113.6 (93.1)	82.7 (91.1)	81.1 (88.8)
0.10	∞		87.0 (93.1)	
0.15	∞		90.0 (93.9)	
0.20	∞		105.8 (94.8)	

occurs in 66% of 41 subcollections (65% of 282 products) considered in the study. Interestingly, the second-best accuracy is obtained from the optimization over n with respect to AIC for a constant α . We find no statistically significant difference between the best two performances. Therefore, we conclude that the main contribution to forecast accuracy stems from the WLS with a constant α , $\alpha < 1$ and its discretization (Table 2). The importance of WLS is mainly because the main drivers of retail demand, such as competitor’s discounts, and broken-assortment effects, are prominent at later weeks of selling seasons whereas demand at the beginning of a selling season is usually not affected by exogenous factors. Discretization is essential since it significantly lowers response variance and test set errors. Note that the relative unimportance of the search over α is also consistent with the study by Smith et al. [35], who claims that α is mainly related to the product characteristics, e.g., basic textile or a trendy design, and does not change over the markdown period. However, ours is the first study revealing the benefit of WLS when it is combined with discretization.

Furthermore, we find that R^2_{WLS} and R^2_{WLS-p} statistics perform significantly worse than the AIC statistic in the weight optimization procedures. The superior performance of AIC stems from the choice of n^* values of 2 and 3 in the optimization compared to the other two GoF statistics. In the joint optimization (first row in Table 2), AIC chooses $n^* \leq 3$ in 78% of cases, whereas this ratio drops to 70% and 49% with R^2_{WLS-p} and R^2_{WLS} statistics, respectively. Similar qualitative observations hold for n^* for $\alpha = 0.2$ (fifth row in Table 2).

Table 2 also reveals that the leverage-based weights [7] are not useful for demand forecasting in markdown applications. The inferior performance of the leverage-based

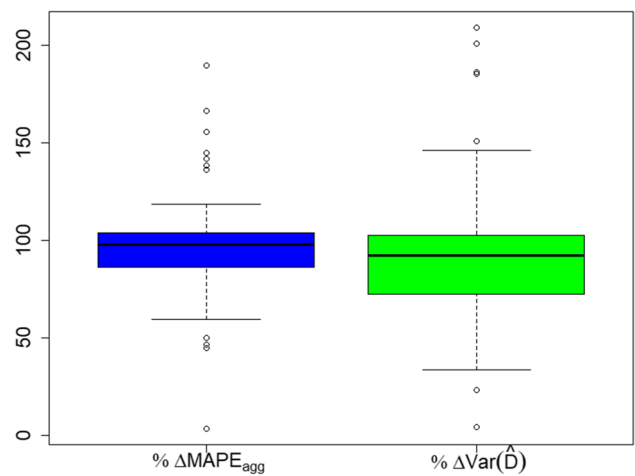


Fig. 3 Box plot for average test error and response variance

weights (> 1000 relative $MAPE_{agg}$ compared to OLS) suggests that the robust optimization is not a good candidate for WLS-based markdown optimization. Robust regression diminishes the impact of end-of-season markdowns and associated demand surges by decreasing their weights in the data set. Note that our conclusions are only limited with our empirical data (around 300 SKUs) and should be tested in other studies.

The positive impact of discretization on prediction accuracy is also confirmed by our bootstrapping study described in Sect. 4.4.2. For each subcollection, we estimate prediction error, $MAPE_{agg}$, and response variance $\widehat{Var}(\widehat{D}^i)$ using the moving block bootstrapping for the regression model (1) trained with $gw^S(0.1, \infty)$ and $gw^S(0.1, n^*)$, respectively. Then the impact of discretization is quantified by percent reduction in $MAPE_{agg}$ and $\widehat{Var}(\widehat{D}^i)$, denoted by $\% \Delta MAPE_{agg}$ and $\% \Delta Var(\widehat{D}^i)$

$$= 100 \frac{MAPE_{agg}(0.1, n^*) - MAPE_{agg}(0.1, \infty)}{MAPE_{agg}(0.1, \infty)} \quad \text{and} \quad \% \Delta Var(\widehat{D}^i)$$

$$= 100 \frac{Var(\widehat{D})(n^*, \infty) - Var(\widehat{D})(0.1, \infty)}{Var(\widehat{D})(0.1, \infty)}$$

The results of the bootstrapping study are depicted in Fig. 3.

Our results indicate that the discretization decreases test error and model variance by 5% and 6% on average. This means the discretization contributes to forecasting accuracy while reducing response variance. Hence, it prevents the model from overfitting the training set.

5.2 Impact of Parameter Update with Exponential Smoothing

Another critical element of our WLS-based demand model is the calculation of prediction coefficients using exponential smoothing. The procedure relies on estimating the regression coefficients at every period and using them to update

Table 3 Relative performance of WLS (WFK) methods with single-stage parameter update ($MAPE_{agg}$ of OLS is set to 100)

Weight parameters		Optimization criteria		
A	N	$RWLS_{WLS}^2$	R^2_{WLS-p}	AIC
$\alpha^*(n^*)$	n^*	> 1000 (149.7)	> 1000 (150.7)	359.5 (152.3)
$\alpha^*(\infty)$	∞	> 1000 (99.2)	> 1000(94.6)	314.2 (95.4)
0.10	n^*	185.2 (105.8)	184.4 (105.6)	184.5 (105.7)
0.15	n^*	121.5 (105.9)	118.2 (105.0)	119.3 (105.2)
0.20	n^*	113.0 (106.3)	105.9 (104.8)	211.3 (105.3)
0.10	∞		217.6 (148.7)	
0.15	∞		138.7 (149.5)	
0.20	∞		113.5 (149.5)	

prediction coefficients with the exponential smoothing in (9) and (10). The update parameter γ controls the speed of the model’s adaptation to the changing discount sensitivity of apparel demand. In this part of the study, we conduct experiments in Table 1 by setting γ to 1 to evaluate the performance of a single-stage parameter calculation procedure. $\gamma = 1$ implies that the estimated regression parameters are equal to the prediction coefficients. The results of the single-stage experiments are presented in Table 3 in terms of the percent decrease in $MAPE_{agg}$ values.

Our results imply that the performance of the procedure with no parameter update is significantly worse compared to the results in Table 2. The best performance is 5% higher $MAPE_{agg}$ than the OLS, which is obtained with the optimization over n via R^2_{WLS-p} (Table 3). The accuracy difference between the two procedures stems from rapidly changing coefficients of the log-linear model, which indicates the existence of some level of overfitting due to extra bias in the demand predictions. These findings indicate that it is beneficial to utilize smoothing procedures when successive predictions from the demand model are required. Our findings extend Smith et al. [35], Smith and Achabal [33], Caro and Martínez-de Albéniz [6], who omit evaluating the performance of the single-stage parameter update. Furthermore, the performance of continuous weight functions is significantly inferior to the discretized weights for the case of $\gamma = 1$. This observation, which is similar to the results in Table 2, also indicates the importance of the weight discretization.

5.3 Impact of Weight Optimization at Every Period

In this study, we suggest a demand modeling system with WLS estimation where the shape and discretization parameters of the GWFs are optimized at the beginning of a markdown season. This practice is motivated by the fact that the WLS estimation addresses customers’ changing discount

Table 4 Relative performance of WLS (WFK) methods with periodic weight optimization and $\gamma = 0.2$ ($MAPE_{agg}$ of OLS is set to 100)

Weight parameters		Optimization criteria		
A	N	R^2_{WLS}	R^2_{WLS-p}	AIC
$\alpha^*(n^*)$	n^*	> 1000 (95.6)	84.7 (87.1)	80.0 (86.3)
$\alpha^*(\infty)$	∞	> 1000 (99.2)	89.8 (94.6)	87.0 (95.4)
0.10	n^*	84.4 (94.4)	84.9 (87.2)	84.4 (86.9)
0.15	n^*	90.6 (93.8)	83.1 (87.1)	84.1 (86.8)
0.20	n^*	115.2 (93.4)	82.2 (87.1)	79.2 (86.8)
0.10	∞	105.8 (93.1)		
0.15	∞	90.0 (93.9)		
0.20	∞	87.0 (94.8)		

sensitivity over the markdown period and the shape of the weight function is mainly driven by the customers’ purchase behavior. In this part of the study, we also evaluate a WLS-based system design where the parameters of GWFs are optimized at every period using a GoF criterion given in the experimental design in Table 1. The results of these experiments are reported in Table 4.

The results of our periodic weight optimization are similar to those shown in Table 2. The highest two accuracy rates are obtained from AIC with the joint optimization and optimization over n , respectively. Comparing these results with the ones in Table 2 indicates that executing the weight optimization at every period does not contribute to the accuracy of the markdown system. This finding is consistent with the assertion that the shape parameter of the WLS method depends on the product’s characteristic [35].

5.4 Impact of Markdown Applications on Demand

In addition to the impact of different WLS features on accuracy, we investigate the effect of markdown applications on total demand in a selling season for the apparel industry. To this end, we consider two models with the highest accuracy: discrete weight functions with $\alpha = 0.2$ with optimization over n using AIC criterion ($gw^S(0.2, n^*)-AIC$), and continuous weight function with $\alpha = 0.1$ ($gw^S(0.1, \infty)$). These two model configurations are benchmarked to the results of the OLS model. In all of these models, $\gamma = 0.2$. The impact of markdown on demand is estimated by calculating the percent difference between list price and demand, denoted by $\hat{D}_t^i(0)$, and estimated demand for the empirical markdown levels at week t :

$$100 * \frac{\hat{D}_t^k(0) - \hat{D}_t^k}{\hat{D}_t^k}, \forall k \in \mathcal{K}.$$

The results of these models indicate that markdown applications during the discount season increases demand by

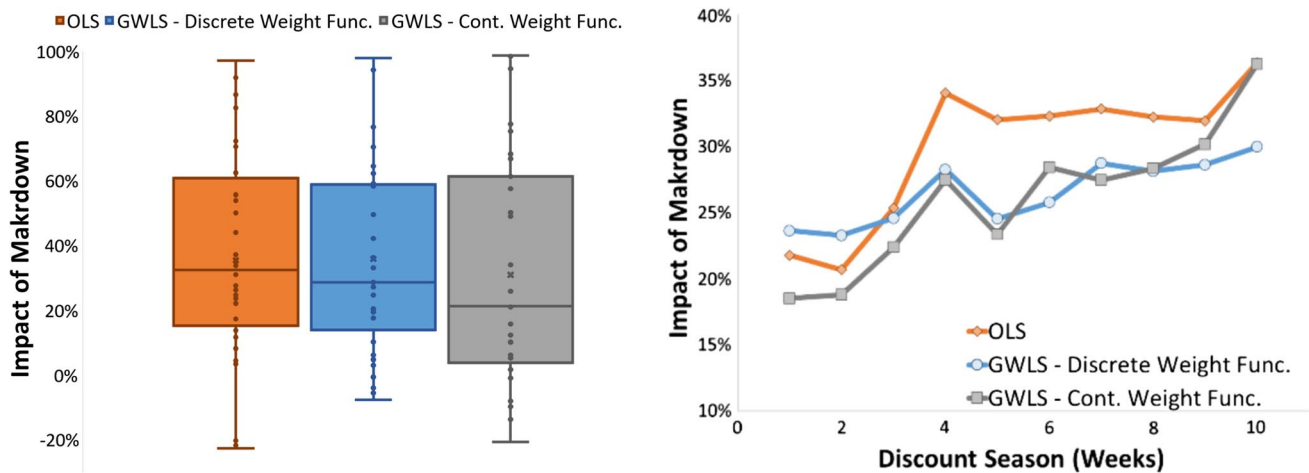


Fig. 4 (Left) Impact of markdown on total demand in discount season. (Right) average impact of markdown over discount season

35% on average. Although there is no significant difference between the three models in terms of the average markdown, our results indicate that OLS and continuous weight function lead to a larger variance for markdown impact estimates (Fig. 4, left).

Furthermore, we find that the impact of markdown significantly increases over the discount season (Fig. 4, right). This increase can be observed from the results of the three models (Fig. 4, left).

6 Conclusion

In the retail sector and the marketing literature, product characteristics are recognized to be the primary drivers of the sales performance of products. Some products, such as fashion clothes or holiday presents, are sold over a single selling season, and they significantly lose their value over time. At the end of a season, leftover inventory must be taken to storage rooms or warehouses to make space for the new-season collection. To stimulate demand and control the associated costs of leftover inventory, retailers apply price markdowns toward the end of selling seasons to maximize their revenue or sell-through rates for the same inventory. Performances of the markdown systems depends on the accuracy of demand predictions.

In this study, we consider a WLS-based demand modeling problem for markdown optimization systems. We take two different weight functions from the literature and generalize them using the dyadic function. The shape and discretization parameters of GWFs are optimized using three different GoF criteria to obtain the best weight sequences. Performances of different weights are compared using point-of-sale data of a Turkish apparel retailer.

We find that assigning higher importance to later weeks of a season pays off significantly, especially with the discretization of weight functions and a simple optimization over the discretization parameter. In contrast, the optimization over the shape parameters of GWFs does not create any additional impact. The importance of the WLS method stems from the prominence of the main drivers of retail demand, e.g., broken assortment effect or competitors' discounts, at the later stages of a selling season. Hence, these periods should be considered with higher weights than the more stable initial weeks of a season. The benefit of discretization stems from the division of the sales history of a product into sub-intervals with increasing weights over time. Our method provides an easy and effective way of achieving this through the dyadic function. It can be applied to other continuous weight functions to increase the forecast accuracy of demand models.

This study can be extended in multiple directions. First, we aim to perform discretization for different products and with different weight function families. Second, we are planning to consider the joint optimization of WLS and markdown levels over a selling season.

Acknowledgements We are thankful to our reviewers for their helpful comments and fruitful discussions.

Author Contributions Mustafa Hekimoğlu is responsible for writing the paper, developing models, compiling results, and writing the paper.

Funding No external funding was used for this study.

Availability of Data and Materials The data set used in this study belongs to a commercial organization which prohibits public sharing of the raw data. We are happy to share Excel files including model statistics upon request.

Declarations

Conflict of Interest There is no conflict of interest.

Ethical Approval and Consent to Participate Hereby, I declare that there is no need for ethics approval or participation consent for the research presented in this paper.

Consent for Publication Hereby, I allow International Journal of Computational Intelligence Systems to publish my article upon the revision process.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agrawal, N., Stephen, A.S.: Retail supply chain management. Springer (2015)
- Azadi, Z., Sandra, D., Eksioğlu, B.E., Palak, G.: Stochastic optimization models for joint pricing and inventory replenishment of perishable products. *Comput. Ind. Eng.* **127**, 625–642 (2019)
- Bergmeir, C., Benítez, J.M.: On the use of cross-validation for time series predictor evaluation. *Inf. Sci.* **191**, 192–213 (2012)
- Billor, N., Chatterjee, S., Hadi, A.S.: A re-weighted least squares method for robust regression estimation. *Am. J. Math. Manag. Sci.* **26**(3–4), 229–252 (2006)
- Caro, F., Gallien, J.: Clearance pricing optimization for a fast-fashion retailer. *Oper. Res.* **60**(6), 1404–1422 (2012). <https://doi.org/10.1287/opre.1120.1102>
- Caro, F., Albéniz, V.M.-d: Fast fashion: business model overview and research opportunities. In: Retail supply chain management, pp. 237–264. Springer (2015)
- Chatterjee, S., Mächler, M.: Robust regression: a weighted least squares approach. *Commun. Stat. Theory Methods* **26**(6), 1381–1394 (1997)
- Chen, S.-P., Chen, C.-Y.: Dynamic markdown decisions based on a quality loss function in on-site direct-sale supply chains for perishable food. *J. Oper. Res. Soc.* **72**(4), 822–836 (2021)
- Chen, Y., Mehrotra, P., Samala, N.K.S., Ahmadi, K., Jivane, V., Pang, L., Shrivastav, M., Lyman, N., Pleiman, S.: A multiobjective optimization for clearance in Walmart brick-and-mortar stores. *INFORMS J. Appl. Anal.* **51**(1), 76–89 (2021)
- Chua, G.A., Mokhlesi, R., Sainathan, A.: Optimal discounting and replenishment policies for perishable products. *Int. J. Prod. Econ.* **186**, 8–20 (2017)
- Çınlar, E.: Probability and stochastics, vol. 261. Springer Science & Business Media (2011)
- Cleveland, W.S., Devlin, S.J.: Locally weighted regression: an approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **83**(403), 596–610 (1988)
- Coşgun, O., Kula, U., Kahraman, C.: Analysis of cross-price effects on markdown policies by using function approximation techniques. *Knowl.-Based Syst.* **53**, 173–184 (2013)
- Cosgun, O., Kula, U., Kahraman, C.: Markdown optimization for an apparel retailer under cross-price and initial inventory effects. *Knowl.-Based Syst.* **120**, 186–197 (2017)
- de Brauwere, A., De Ridder, F., Pintelon, R., Elskens, M., Schoukens, J., Baeyens, W.: Model selection through a statistical analysis of the minimum of a weighted least squares cost function. *Chemom. Intell. Lab. Syst.* **76**(2), 163–173 (2005)
- Fisher, M.: Or forum-rocket science retailing: the 2006 philip mccord morse lecture. *Oper. Res.* **57**(3), 527–540 (2009)
- Guler, M.G., Bilgic, T., Gullu, R.: Joint inventory and pricing decisions with reference effects. *IIE Trans.* **46**(4), 330–343 (2014)
- Harsha, P., Subramanian, S., Uichanco, J.: Dynamic pricing of omnichannel inventories: honorable mention—2017 M&SOM practice-based research competition. *Manuf. Serv. Oper. Manag.* **21**(1), 47–65 (2019)
- Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H.: The elements of statistical learning: data mining, inference, and prediction, vol. 2, pp. 1–758. Springer, New York (2009)
- Hongyi Li, G.S., Maddala.: Bootstrapping time series models. *Economet. Rev.* **15**(2), 115–158 (1996)
- Huang, Z., Huang, L., Zhao, Y., Meng, X.: Money-back guarantee in the presence of strategic customer behavior. *Int. J. Prod. Econ.* **239**, 108191 (2021)
- James, G., Witten, D., Hastie, T., Tibshirani, R.: An introduction to statistical learning, vol. 112. Springer (2013)
- Keskin, N.B., Zeevi, A.: Chasing demand: learning and earning in a changing environment. *Math. Oper. Res.* **42**(2), 277–307 (2017)
- Kunsch, H.R.: The jackknife and the bootstrap for general stationary observations. *Ann. Stat.* **17**, 1217–1241 (1989)
- Li, R., Teng, J.-T.: Pricing and lot-sizing decisions for perishable goods when demand depends on selling price, reference price, product freshness, and displayed stocks. *Eur. J. Oper. Res.* **270**(3), 1099–1108 (2018)
- Loader, C.: Local regression and likelihood. Springer Science & Business Media (2006)
- Mahmoodi, A.: Joint pricing and inventory control of duopoly retailers with deteriorating items and linear demand. *Comput. Ind. Eng.* **132**, 36–46 (2019)
- Manning, W.G., Mullahy, J.: Estimating log models: to transform or not to transform? *J. Health Econ.* **20**(4), 461–494 (2001)
- Onal, M., Kundakcioglu, O.E., Jain, S.: An EOQ model with deteriorating items and self-selection constraints. *Optim. Lett.* **14**(6), 1491–1518 (2020)
- Rice, D.H., Fay, S.A., Xie, J.: Probabilistic selling vs markdown selling: price discrimination and management of demand uncertainty in retailing. *Int. J. Res. Mark.* **31**(2), 147–155 (2014)
- Ruppert, D., Wand, M.P.: Multivariate locally weighted least squares regression. *Ann. Stat.* **22**, 1346–1370 (1994)
- Smith, S.A.: Clearance pricing in retail chains. In: Retail supply chain management, pp. 387–408. Springer, Boston (2015)
- Smith, S.A., Achabal, D.D.: Clearance pricing and inventory policies for retail chains. *Manage. Sci.* **44**(3), 285–300 (1998)
- Smith, S.A., Agrawal, N.: Optimal markdown pricing and inventory allocation for retail chains with inventory dependent demand. *Manuf. Serv. Oper. Manag.* **19**(2), 290–304 (2017)

35. Smith, S.A., McIntyre, S.H., Achabal, D.D.: A two-stage sales forecasting procedure using discounted least squares. *J. Mark. Res.* **31**, 44–56 (1994)
36. Street, J.O., Carroll, R.J., Ruppert, D.: A note on computing robust regression estimates via iteratively reweighted least squares. *Am. Stat.* **42**(2), 152–154 (1988)
37. Wang, H., Li, S., Luo, J.: Optimal markdown pricing for holiday basket with customer valuation. *Int. J. Prod. Res.* **56**(18), 5982–5996 (2018)
38. Willett, J.B., Singer, J.D.: Another cautionary note about r^2 : its use in weighted least-squares regression analysis. *Am. Stat.* **42**(3), 236–238 (1988)
39. Yıldız, S. C.: Markdown optimization in apparel retail sector. Master's thesis, Işık University, İstanbul (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.