



Dynamic pricing using flexible heterogeneous sales response models

Philipp Aschersleben¹ · Winfried J. Steiner¹

Received: 31 March 2022 / Accepted: 12 February 2024 / Published online: 29 March 2024
© The Author(s) 2024

Abstract

We combine nonparametric price response modeling and dynamic pricing. In particular, we model sales response for fast-moving consumer goods sold by a physical retailer using a Bayesian semiparametric approach and incorporate the price of the previous period as well as further time-dependent covariates. All nonlinear effects including the one-period lagged price dynamics are modeled via P-splines, and embedding the semiparametric model into a Hierarchical Bayesian framework enables the estimation of nonlinear heterogeneous (i.e., store-specific) immediate and lagged price effects. The nonlinear heterogeneous model specification is used for price optimization and allows the derivation of optimal price paths of brands for individual stores of retailers. In an empirical study, we demonstrate that our proposed model can provide higher expected profits compared to competing benchmark models, while at the same time not seriously suffering from boundary problems for optimized prices and sales quantities. Optimal price policies for brands are determined by a discrete dynamic programming algorithm.

Keywords Sales response models · Functional flexibility · Store heterogeneity · Price dynamics · Price optimization · Discrete dynamic programming

1 Introduction

Estimating price response functions to support pricing decisions is a highly relevant topic in the marketing literature and in marketing research practice. A price response function or, more generally speaking, a sales response function relates the sales of a brand to own and competitive brand prices, to accompanying marketing activities

✉ Philipp Aschersleben
philipp.aschersleben@tu-clausthal.de

Winfried J. Steiner
winfried.steiner@tu-clausthal.de

¹ Department of Marketing, Clausthal University of Technology, Julius-Albert-Str. 2, 38678 Clausthal-Zellerfeld, Germany

(like feature or display advertising), and/or to further covariates accounting for time- or store-specific effects. In this paper, we focus on sales response modeling for fast-moving consumer goods that are sold by physical retailers and for which scanner data are nowadays widely available. Publications in this research area have particularly focused on the following dimensions:

First, the specification of the ‘correct’ functional form for the relationship between sales and prices was dominated by strictly parametric modeling until the 2000s, starting with simple linear regression (lin-lin) models and followed by non-linear parametric models, especially multiplicative (log-log) and exponential (log-lin) models. To overcome the problem that parametric models can largely fail to approximate the true functional form inherent to real data (Härdle 1990; Van Heerde 1999; Leeftang et al. 2000), researchers started to apply more flexible nonparametric techniques, which are able to explore the functional shape directly from data instead of assuming a predefined parametric functional form (Hanssens et al. 2002). There is very clear evidence from many studies that these nonparametric methods can (greatly) improve the predictive model performance as well as expected profits over parametric modeling.¹ According to Van Heerde et al. (2002), managers should rely on models which provide the most accurate predictions. Besides, more flexible nonparametric estimation methods have also become established in choice modeling, see, e.g., Abe (1991), Abe (1995), Abe et al. (2004), or Boztuğ et al. (2014).

Second, it is well-known that the aggregation of store-level data across stores of a retail chain leads to biased estimates of the effects of marketing activities if these effects differ for individual stores (as one would expect) but even if marketing effects are homogeneous across stores (e.g., Krishnamurthi et al. 1990; Christen et al. 1997). Further, if marketing effects were different across individual stores of a retailer, pooling store-level data into a homogeneous model should also cause a bias in the effect estimates. Although empirical findings are not unequivocal, heterogeneous store-level sales response models that enable the estimation of store-specific effects often provided more or at least not less accurate sales predictions compared to their homogeneous (pooled) counterparts. More importantly, only heterogeneous sales response models allow for store-specific optimal pricing, also known as micro-marketing pricing (Montgomery 1997). There is further empirical evidence that accounting for (unobserved) heterogeneity can also increase expected chain profits (e.g., Montgomery 1997; Lang et al. 2015).

Third, it is reasonable to assume that past prices of brands or expected prices in future periods might affect a brand’s sales or price response in the current period. Therefore, the accommodation of price dynamics in sales response models has also been of a long tradition, be it via considering lagged or lead price covariates (e.g., Van Heerde et al. 2000, 2004), allowing for time-varying parameters (e.g., Foekens et al. 1999; Kopalle et al. 1999), or incorporating market-level reference price variables (e.g., Greenleaf 1995; Fibich et al. 2003; Aschersleben

¹ We use the term nonparametric function to model the relationship between a specific metric predictor (like price) and a metric criterion variable (like sales) using flexible regression techniques, while the term semiparametric approach will describe a model where one or more predictor effects are captured parametrically, while others nonparametrically.

and Steiner 2022). Almost all of these dynamic approaches have been directed at either explaining structural effects of sales promotions (for example the lack of postpromotion dips in store data), or showing that dynamic pricing can increase profits over static models. Importantly, dynamic pricing allows the computation of optimal price paths, i.e., optimal prices that vary over time due to the consideration of time-dependent effects on price or sales response.

From the previous explanations, it is clear that accommodating functional flexibility in price response, store heterogeneity in marketing effects, and price dynamics can improve both sales predictions and expected chain profits, which is why it seems promising to consider these features jointly in a store-level sales response model. However, not a single approach has yet accounted for all three dimensions simultaneously. Therefore, our first research question is whether a sales response model that combines the three components can provide higher expected chain profits compared to existing simpler (nested) models. In particular, our proposed approach will allow for the derivation of optimal price paths by modeling heterogeneous nonlinear price effects via nonparametric price functions (including the one-period lagged price effect), and we will assess whether this higher model complexity pays off compared to simpler models.

As indicated above, some researchers went beyond the descriptive modeling stage and used an estimated sales or price response model as basis to subsequently derive optimal prices or price paths. Most of these optimization models assume linear price effects, i.e., price effects estimated by strictly parametric response modeling, and almost all of these parametric approaches also included price dynamics. The optimization results for many of these parametric approaches provide evidence or at least suggest the existence of corner solutions, i.e., that optimized prices hit the upper boundary of observed prices, which would make the price optimization exercise seem less useful. At this point, it is important to note that none of the dynamic optimization approaches has modeled price response more flexibly, like we propose in our model. Our second research question therefore relates to this boundary pricing problem, making our proposed approach recommendable only if optimal prices do not, or not always or not very frequently hit the boundaries of observed prices (and as well do not induce boundary solutions for related sales quantities). More explicitly, we want to analyze how prone our proposed model is to boundary price hits. If the boundary pricing problem is not an issue, we are probably the first to enable the calculation of optimal price paths for each of the stores of a retail chain via nonparametric heterogeneous price effects modeling.

The rest of the paper is organized as follows: In Sect. 2, we provide a review of the relevant literature for our approach. In Sect. 3, we introduce our new dynamic semiparametric Hierarchical Bayesian sales response model together with nested model versions. We will use the latter for model comparison. We further propose a discrete dynamic programming approach for calculating optimal price paths. Section 4 starts with a short description of the store-level scanner data used in our empirical study and provides details about model estimation and validation procedures. Moreover, optimal prices, optimized profits, and expected losses when not

using the proposed model are analyzed. In Sect. 5, we conclude with a summary of findings and discuss limitations of our research.

2 Literature review

Table 1 provides a summary of relevant articles related to sales response modeling in the field of fast-moving consumer goods that have combined at least two of the dimensions discussed in the introduction (i.e., functional flexibility in price response, store heterogeneity in marketing effects, price dynamics, price optimization) in their empirical applications.

As mentioned before, nonlinear parametric models were state-of-the-art for a long time to capture price or more generally sales effects for fast-moving consumer goods (e.g., Hruschka 1997; Montgomery 1997; Foekens et al. 1999; Kopalle et al. 1999; Van Heerde et al. 2000, 2002; Hruschka 2006b; Andrews et al. 2008), and they still serve as benchmarks for the more sophisticated nonparametric approaches observed nowadays in many publications. Possible nonparametric operationalizations have ranged from kernel regression (e.g., Van Heerde et al. 2001) and neural nets (e.g., Hruschka 2006a, 2007) to splines such as stochastic cubic splines (Kalyanam and Shively 1998), B-splines (Hruschka 2000; Haupt and Kagerer 2012; Haupt et al. 2014) or Bayesian P-Splines (Steiner et al. 2007; Brezger and Steiner 2008; Weber and Steiner 2012; Lang et al. 2015; Weber et al. 2017). It is clear from the existing studies in this research area that more flexible estimation techniques are more powerful to uncover complex nonlinearities in sales response and, if these are at work, may lead to much better sales predictions and higher expected profits.

The consideration of (unobserved) store heterogeneity in sales response modeling has become even more established, see the column ‘Store heterogeneity’ in Table 1. Some researchers have only included store intercepts in order to account for heterogeneity in baseline sales across stores. The more advanced approaches to also capture store heterogeneity in marketing effects have all been embedded into Hierarchical Bayesian (HB) estimation frameworks,² where some of them have accommodated both heterogeneity and functional flexibility in price effects (Hruschka 2006a, 2007; Lang et al. 2015; Weber et al. 2017) and the latter three of them price optimization in addition. Interestingly, some researchers found that accounting for store heterogeneity in a parametric sales response model did not necessarily improve the predictive model performance (Andrews et al. 2008; Weber and Steiner 2012, 2021), whereas Weber et al. (2017) did report major advantages from addressing store heterogeneity as soon as functional flexibility in price response was taken into account. Leeflang et al. (2000) already very early called for more intense research with regard to the comparison of heterogeneous versus homogeneous store-level sales response models, and we pick up their suggestion for model comparison in our empirical study later on.

² While the single normal distribution is typically used in these HB models to represent store heterogeneity on the upper (population) level, Wedel and Zhang (2004) used a Dirichlet process prior to capture heterogeneity in price effects across stores more flexibly.

Table 1 Overview of empirical studies in the context of (store-level) sales response modeling considering at least two of the following dimensions: functional flexibility in price response, store heterogeneity in marketing effects, price dynamics, price optimization

Study	Price dynamics	Functional flexibility	Store heterogeneity	Price optimization
Greenleaf (1995)	Reference prices	-	-	Dynamic programming
Montgomery (1997)	-	-	Hierarchical Bayes (HB)	Sequ. quadr. programming
Foelkens et al. (1999)	Time-varying param.	-	Store intercepts	-
Kopalle et al. (1999)	Time-varying param.	-	Store intercepts	Dynamic programming
Van Heerde et al. (2001)	-	Kernel regression	Store intercepts	-
Fibich et al. (2003)	Reference prices	-	-	Analytical solution
Van Heerde et al. (2004)	Lead and lag prices	Local polynom. regression	-	-
Fok et al. (2006)	Error correction model	-	Hierarchical Bayes (HB)	-
Hruschka (2006a)	-	Neural nets	Hierarchical Bayes (HB)	-
Hruschka (2006b)	-	-	Hierarchical Bayes (HB)	Grid search
Hruschka (2007)	-	Neural nets	Store clusters, HB	Improving hit-and-run
Steiner et al. (2007)	-	Bayesian P-splines	Store intercepts	-
Brezger and Steiner (2008)	-	Bayesian P-splines	Store intercepts	-
Haupt and Kagerer (2012)	-	B-splines	Store intercepts	-
Horváth and Fok (2013)	Lag prices	-	Hierarchical Bayes (HB)	-
Haupt et al. (2014)	-	B-splines	Store intercepts	-
Lang et al. (2015)	-	Bayesian P-splines	Hierarchical Bayes (HB)	Grid search
Weber et al. (2017)	-	Bayesian P-splines	General heterogeneity SUR model, HB	Evolutionary algorithm
Ascherleben and Steiner (2022)	Reference prices	Bayesian P-splines	-	-
This study	Lag prices	Bayesian P-splines	Hierarchical Bayes (HB)	Dynamic programming

Price dynamics were only seldom considered in models with nonlinear price effects, Van Heerde et al. (2004) and more recently Aschersleben and Steiner (2022) are the exceptions with their empirical studies. These two approaches, however, did not account for store-specific marketing effects. Fok et al. (2006) and Horváth and Fok (2013) proposed Hierarchical Bayesian vector autoregression models to analyze brand-/category-specific differences in dynamic price effects on sales. But the focus of these two studies was more on the investigation of brand- and category-specific characteristics as moderators for differences in own- or cross-price effects between brands and product categories rather than on within-brand store heterogeneity in price response (especially as Fok et al. 2006 used data from only one single store). The two approaches that combined price dynamics and store heterogeneity (Foekens et al. 1999; Kopalle et al. 1999) only included store intercepts to consider differences in baseline sales across stores. Similar to Leeflang et al. (2000), we here see the need for a more intense research to assess the effect of price dynamics in store-level sales response models that as well account for store-specific effects and ideally also for functional flexibility in price response. Generally, we observe quite different strategies to accommodate price dynamics in sales response models: via lagged and lead prices (Van Heerde et al. 2004; also, e.g., Van Heerde et al. 2000), via time-varying parameter models (Foekens et al. 1999; Kopalle et al. 1999; also, e.g., Ataman et al. 2010), via vector-autoregressive specifications (Fok et al. 2006; Horváth and Fok 2013; also, e.g., Nijs et al. 2001), as well as via market-level reference prices (Greenleaf 1995; Fibich et al. 2003; Aschersleben and Steiner 2022).

Only the minority of the papers collected in Table 1 provided normative implications with respect to expected profits or to expected losses from not using the sales response model with the best predictive performance. Like the milestone article on micro-marketing pricing strategies by Montgomery (1997), all more recent optimization approaches used a Hierarchical Bayesian estimation framework to allow for store-specific (or at least clusterwise) optimal pricing (Hruschka 2006a, b, 2007; Weber et al. 2017; Lang et al. 2015). None of them, however, incorporated price dynamics for price optimization. On the other hand, the three remaining approaches that considered dynamic price effects on sales response (Greenleaf 1995; Kopalle et al. 1999; Fibich et al. 2003) did not allow for price optimization by modeling nonlinear and/or heterogeneous store-level pricing effects. But, the findings from the latter three studies still indicated that accommodating price dynamics can increase profits over static price optimization.

As another important issue, several of the price optimization studies seem to provide evidence of corner solutions for optimized prices, i.e., that optimized prices would have hit the upper boundary of the observed price range if this upper boundary had been imposed as a constraint in the model.³ Interestingly, this applies to all three papers that incorporated price dynamics for price optimization (Greenleaf 1995; Kopalle et al. 1999; Fibich et al. 2003), and these three dynamic optimization approaches have further in common that they used linear pricing effects (i.e., price effects estimated by a parametric sales response model). The corner solution problem seems as well evident in the paper of Montgomery (1997), who did not consider

³ We thank one referee for pointing us to this boundary price phenomenon.

price dynamics but also used a parametric sales response model (an exponential one) to determine optimal prices. Three optimization approaches were based on semiparametric sales response models, where nonlinear price effects were estimated using neural nets or splines in the descriptive model stage (Hruschka 2007; Lang et al. 2015; Weber et al. 2017). Indeed, the findings reported in Hruschka (2007) and Weber et al. (2017) suggest that flexible estimation methods might be less affected by the boundary pricing problem.⁴ Note that we found no evidence that optimized prices hit the lower bound of observed prices. And, neither paper further allows conclusions about boundary effects for predicted sales quantities based on optimized prices.

At this point, it is very important to mention that the findings from our explorative research on boundary price effects should be treated with caution since the corner solution problem is not explicitly addressed in any of the papers and information on boundary effects is sparse. In general, optimal prices depend on price elasticities and on variable costs (in our case the wholesale prices of a retailer). Still, the price elasticity in parametric models is much more “rigid”, as it depends on one or only few price parameter estimates which globally determine the shape of the response function over the entire observed price range. Nonparametric estimation techniques like splines, on the other hand, fit the data locally, entailing greater flexibility for uncovering complex price response patterns and allowing locally varying price elasticities. This locally fitting property can make flexible models less prone to corner solutions for optimal prices. Note that none of the dynamic optimization approaches has modeled price response more flexibly, like we propose in our dynamic model. And different from all previous studies in this field, we will put a special focus on the boundary pricing problem in our empirical study.

Based on the literature review, the central contribution of this paper is that our proposed model will allow for price optimization by modeling heterogeneous nonlinear pricing effects through nonparametric functions for own-, cross-, and lagged price response. In other words, we are probably the first to enable the computation of truly dynamic price paths⁵ for each store of a retail chain due to heterogeneous nonlinear pricing effects embedded in a semiparametric sales response model. None of the previously proposed sales response models has combined the three features functional flexibility, store heterogeneity, and price dynamics to derive optimal pricing strategies for fast-moving consumer goods.

⁴ Hruschka (2007) determined an eight-cluster solution for the stores of a retail chain with equal optimal brand prices for all stores within a cluster, and optimized prices largely varied across clusters for most brands considered. On the one hand, the large variation of optimal prices might indicate less boundary problems. It remained unclear, however, how often the optimal cluster prices hit the upper boundary of observed prices at the individual store level. Weber et al. (2017) provided a plot of optimal store-specific brand prices for one brand as example, suggesting a fairly low number of boundary effects for their flexible model.

⁵ Note that time-varying optimal prices can also result from static price optimization if variable costs are not constant over time.

3 Model specification and optimization approach

In the following, we introduce a semiparametric, heterogeneous, and dynamic store-level sales response model. We use Bayesian P-splines as nonparametric method to estimate immediate and lagged price effects flexibly, allowing us to uncover possibly exceptional pricing effects (like distinct threshold or saturation effects) directly from the data. An advantage of P-splines is that they can easily be constrained to provide monotonic shapes of price response (Brezger and Steiner 2008), which is reasonable from an economic point of view for fast-moving consumer goods. Following Lang et al. (2015), heterogeneity in price response across stores is captured via store-specific scaling factors, which can be as well easily embedded as additional parameters into the Gibbs sampling procedure of our Hierarchical Bayesian estimation framework. Since the shape of the price response is pooled over stores by this heterogeneity specification (i.e., functional flexibility and heterogeneity in price response are not decoupled), we further provide robustness checks by varying the number of knots and the degree of the underlying B-spline basis functions. Dynamics are accommodated via the one-period lagged own-item price since more time lags were not supported by the data for almost all brands considered. We will also check the assumptions required for a proper estimation of the models (multicollinearity, heteroskedasticity, autocorrelation). We apply a discrete dynamic programming algorithm to derive optimal price paths.

3.1 Semiparametric heterogeneous dynamic model

We use the following additive sales response model with smooth nonparametric and multiplicative random price effects.⁶ Accordingly, the (log) unit sales of a brand in a specific store and week are assumed to depend on own- and competitive price effects (where the latter are captured at the quality tier level), as well as a dynamic, one-week lagged own-price effect. We further include promotional activities for the brand (use of a display, odd pricing) and accommodate seasonality effects via a smooth monthly trend and unobserved store-specific effects. Models are estimated for each brand separately, thus brand indices are omitted.

$$\begin{aligned} \log(q_{st}) &= \eta_{st} + \varepsilon_{st} \\ &= f_0(m_t) + (1 + \alpha_{s1})f_1(p_{st}) + (1 + \alpha_{s2})f_2(p_{s;t-1}) \\ &\quad + \sum_i (1 + \alpha_{s,c_i})f_{c_i}(p_{st}^{c_i}) + \mathbf{v}'_{st}\boldsymbol{\gamma}_s + \beta_s + \varepsilon_{st}, \end{aligned} \quad (1)$$

where f_0 is an unknown smooth nonlinear time trend of the calendar month (m_t); f_1 is an unknown smooth nonlinear decreasing function of the brand's own price (p_{st}); f_2 is an unknown smooth nonlinear increasing function of the brand's one-week lagged own price ($p_{s;t-1}$); f_{c_i} are unknown smooth nonlinear increasing functions for cross-price effects ($p_{st}^{c_i}$) captured at the level of the price-quality tiers c_i ,

⁶ Effects of marketing instruments other than prices as well as store intercepts are captured parametrically, which is why the model is called a semiparametric model, also see footnote 1.

where $c_i \in \{(\text{prem}), (\text{nat}), (\text{priv})\}$ denotes premium, national, and private label brands (compare Table 6 in the Appendix); $(1 + \alpha_{s1}), (1 + \alpha_{s2}), (1 + \alpha_{s,c_i})$ are store-specific scaling factors capturing price heterogeneity in a multiplicative way; \mathbf{v}'_{st} contains store-specific random slopes for marketing activities represented by indicator variables (e.g., use of a display, odd price endings), and the vector $\boldsymbol{\gamma}_s$ captures the corresponding parametric effects; β_s is a store-specific random intercept for store s ; ε_{st} is a Gaussian error term with mean zero and variance σ^2 .

The unknown smooth nonlinear functions f are modeled via Bayesian P-splines (Lang and Brezger 2004), and the scaling factors $\alpha_{s,j}$ of the price response functions account for unobserved heterogeneity in (smooth) price response across the stores. Aschersleben and Steiner (2022) recently provided a larger paragraph on the advantages of using nonparametric regression techniques in general and Bayesian P-splines in particular (Aschersleben and Steiner 2022, p. 632). In a nutshell, nonparametric estimators “read out” the functional shape of a predictor effect directly from the data, while parametric models require the specification of a fixed functional form a priori. This property of nonparametric methods is especially beneficial in our context for uncovering existing steps or strong kinks in price response (e.g., caused by effects of psychological or reference prices, or the existence of different reservation prices across consumers), which are generally more difficult to capture with parametric functions. Bayesian P-splines additionally offer some very convenient estimation properties: (a) Overfitting as a possible undesirable result of the higher flexibility of nonparametric estimators can easily be counteracted through a roughness penalty term (hence the name *P*-splines). (b) The amount of smoothness of a predictor effect can be estimated jointly with all other parameters when using a Bayesian estimation framework (Lang and Brezger 2004), i.e., no additional smoothing parameter selection procedures are required. (c) Not least, as mentioned above, the extended Gibbs sampling procedure as developed by Brezger and Steiner (2008) allows to easily accommodate monotonicity constraints for price effects. Steiner et al. (2007) further provided a graphical illustration how the P-spline approach works (Steiner et al. 2007, p. 391). For P-spline estimation, we use by default 20 equidistant knots within the range of observed levels of each of the metric covariates (prices, calendar time). This is in line with Eilers and Marx (1996), who introduced P-splines into the statistical literature and recommended the use of 20 to 40 knots in order to guarantee enough flexibility for a smooth function to be estimated. As noted above, a roughness penalty is imposed on neighboring knots in order to prevent an overfitting of the splines at the same time, and we use the common setting of a second-order random walk for this penalty. This smoothing or regularization implies that the effective number of parameters to be estimated is much less than the number of knots. Note that because the underlying B-spline basis functions (roughly corresponding to the number of knots) overlap and the data set therefore cannot be divided, a P-spline is not estimated as a sequence of separate (Bayesian) regression functions but as a whole, while at the same time preserving its local fitting property. Lastly, P-splines are further determined by the degree of the B-spline basis functions, where we use B-splines of degree 3 by default corresponding to cubic splines. Technical details on the Bayesian inference of our models, especially for the smooth

Table 2 Comparison of different model variants

Simple	vs.	Complex
<i>Static</i> (STAT)	vs.	<i>Dynamic</i> (DYN)
<ul style="list-style-type: none"> • Lagged own price not included 		<ul style="list-style-type: none"> • Lagged own price included
<i>Parametric</i> (PAR)	vs.	<i>Flexible</i> (FLEX)
<ul style="list-style-type: none"> • Parametric modeling of (log) price effects 		<ul style="list-style-type: none"> • Nonparametric modeling of price effects via P-splines
<i>Homogeneous</i> (HOM)	vs.	<i>Heterogeneous</i> (HET)
<ul style="list-style-type: none"> • No store-specific scaling parameters / random effects for prices and other marketing activities 		<ul style="list-style-type: none"> • Store-specific scaling parameters / random effects for prices and other marketing activities
The following variables and effects as introduced in Eq. (1) are included in every model specification:		
<ul style="list-style-type: none"> • Baseline store sales (store-specific intercepts) • Seasonal effect (unknown smooth nonlinear monthly trend) 		

functions for price effects and their scaling, are available from the authors upon request.

By using single-brand equations, we implicitly take the perspective of brand managers, who should be interested in analyzing the consequences of suboptimal pricing strategies from using a wrong sales response model for their brand(s), assuming fixed price patterns for competing brands and the possibility to adjust the price of the own brand in the retailer's assortment. In case brand managers have only access to retail data on a more aggregate level, e.g., on the chain-level, they must cooperate with the retailer or outsource a project to obtain the desired analysis at the store-level (e.g., Hruschka 2006b, p. 94). In addition, product category managers of the retailer could use the brand perspective as well to analyze how price changes for one brand in the assortment affect category sales and profits. Independent of that, the study should provide evidence of the benefits of using a more complex model that includes heterogeneity, functional flexibility, and price dynamics for optimal pricing over simpler models that ignore some or all of these dimensions.

3.2 Benchmark models

As the focus of this article is on combining price *dynamics*, *flexible* nonlinear modeling, and store *heterogeneity*, the model as stated in Eq. (1) and referred to as DYN-FLEXHET in the following, is the most complex one, i.e., it combines all three dimensions of interest. Contrary to that, the other (benchmark) models only account for some of these dimensions or none of them such that the most simple model is a *static* model without price dynamics and with *parametric* and *homogeneous* effects only (STATPARHOM). Table 2 shows how the three model features can be combined to a total of $2^3 = 8$ different models, with the model in Eq. (1) as proposed model and the other seven as benchmark models. The benchmark models are nested in the DYNFLEXHET model in the sense that ignoring the lagged price effect leads to *static* models, and/or replacing the unknown smooth nonlinear functions for the price

effects by parametric effects leads to *parametric* models, and/or ignoring heterogeneity in marketing effects leads to *homogeneous* models.

As in Eq. (1), all benchmark models use log brand sales as dependent variable. The *parametric* models are specified as multiplicative models for capturing price effects, relating log brand sales to log-transformed prices. The multiplicative model is the most widely used parametric (store) sales response model, also because of its higher flexibility to modeling cross-price effects (convex, concave, or linearly) compared to other parametric models (see among others, Foekens et al. 1999; Kopalle et al. 1999; Van Heerde et al. 2001; Hruschka 2006a, b; Weber and Steiner 2012; Horváth and Fok 2013; Weber et al. 2017). Note that the seasonality effect is generally modeled via a flexible function across the competing models (i.e., even in the *parametric* models) and that *homogeneous* models still include store-specific random intercepts in order to accommodate differences in baseline sales of the brands across stores.

3.3 Discrete dynamic pricing

Our aim is to find optimal brand prices p^* for the stores in our data set so that a brand’s total chain profit V is maximized:

$$p^* = \underset{p}{\operatorname{argmax}} \{V(p) = (p - c)q(p)\}, \tag{2}$$

where c represents the variable cost (wholesale price) and q the sales response function.

Optimal prices are determined from a brand management perspective based on the assumptions of a fixed promotion calendar (e.g., as described in Silva-Risso et al. 1999; Chintagunta et al. 2003) and given competitive prices in each store and week as observed in the data. This is also known as fictitious play (Brown 1951, as cited by Hruschka 2006b, p. 99) and implies that communicative promotional activities (like display advertising) as well as future prices of substitute brands occur with frequencies as in the past and do not become part of the optimization. We further dispense with error correlations between brands like in the SUR (seemingly unrelated regression) model approach by Weber et al. (2017) to prevent additional model complexity. Note that there are no interactions between stores (e.g., Montgomery 1997), i.e., the search for optimal prices can be conducted for each store separately, even if heterogeneity across stores is considered. Thus, the objective function V can be written as an additive function of total profits v_{mi} across all brands (m) and stores (i):

$$V(p) = \sum_{m=1}^M \sum_{i=1}^N v_{mi}(p) = \sum_{m=1}^M \sum_{i=1}^N \sum_{t=1}^T \delta^t v_{mit}(p) \tag{3}$$

$$= \sum_{m=1}^M \sum_{i=1}^N \sum_{t=1}^T \delta^t (p_t - c_t) q_t(p_t, p_{t-1}), \tag{4}$$

where $\delta = 1/(1 + r)$ is a discount factor with discount rate r . As q_t does not only depend on the current own price p_t , but also on the lagged own price p_{t-1} (see Eq. (1)), this is a dynamic system, and we transfer Eq. (4) into a dynamic program by introducing the state and action variables s_t and x_t in the following to employ the optimization step. In our context, the brand’s one-period lagged own price represents the current state ($s_t = p_{t-1}$), and the action variable is defined as the brand’s own price to be set in period t ($x_t = p_t$). In the given setting, as indicated above, it is sufficient to optimize these brand- and store-specific profits $v = v_{mi}$ separately, leading to the following general formulation:

$$v(x_t, s_t) = \sum_{t=1}^T \delta^t v_t(x_t, s_t) = \sum_{t=1}^T \delta^t (x_t - c_t) q_t(x_t, s_t). \tag{5}$$

The state variable s_t is an element of the state space $S_t = S$ that contains all possible prices and is thus discrete and constant for all periods. The same holds for the action space $X_t(s_t) = X_t = X$ that is further independent of the current state, i.e., the action space of feasible levels for setting the price in the current period is independent from the price of the previous period (and thus constant in all periods). The next period’s state s_{t+1} equals the value of the action variable of the previous period so that the new lagged own price in period $t + 1$ is the selected own price for period t and the corresponding transition function g simplifies to

$$s_{t+1} = g(s_t, x_t) = x_t. \tag{6}$$

Then, setting the initial value $s_1 = p_0$, the complete dynamic program can be stated as follows:

$$\left. \begin{array}{l} \text{Maximize}_{x_t \in X} \quad v(s) = \sum_{t=1}^T \delta^t v_t(x_t, s_t) \\ \text{subject to} \quad \begin{array}{l} s_{t+1} = x_t, \quad t = 1, \dots, T, \\ s_1 = p_0, \\ s_t \in S, \quad t = 2, \dots, T, \\ x_t \in X, \quad t = 1, \dots, T. \end{array} \end{array} \right\} \text{(P)}$$

Following Bellman’s principle of optimality, we can write this problem as a Bellman equation

$$v_t^*(s_t) = \max_{x_t \in X} \{v_t(x_t, s_t) + \delta v_{t+1}^*(s_{t+1})\}, \quad t = 1, \dots, T \tag{7}$$

and solve it iteratively via forward or backward recursion (Miranda and Fackler 2002, Ch. 7). Note that optimization in the *static* model variants reduces to a week- and store-wise price search without the consideration of time interdependencies and therefore does not require a dynamic program.

4 Empirical study

In this section, we describe the data used in our empirical study and provide details about model estimation including a visual inspection of model fit. We further propose measures for evaluating the predictive model performance and discuss the predictive validity results for the competing models. The section ends with a detailed analysis of normative pricing and profit implications based on the competing models (optimal price paths, optimal profits, price image issues, expected losses), as well as with an illustration and evaluation of boundary price effects.

4.1 Data

For our empirical analysis, we use scanner data for refrigerated orange juice brands sold by a large supermarket chain in the Chicago metropolitan area (Dominick's Finer Foods). The data are freely available at the James M. Kilts Center of the University of Chicago for academic purposes. Our data refer to weekly unit sales of 64 oz. packages for $b \in \{1, \dots, B = 8\}$ brands that can be grouped into three price-quality tiers: the two *premium* brands "Florida Natural" and "Tropicana Pure", the five *national* brands "Citrus Hill", "Florida Gold", "Minute Maid", "Tree Fresh", and "Tropicana", and the *private* label brand "Dominick's" representing the supermarket's own store brand. The data were collected in $s \in \{1, \dots, S = 81\}$ stores of the supermarket chain over a time span of $t \in \{1, \dots, T_s\}$ weeks with $T_s \in [74;87]$, resulting in a total of 54,260 observations. Descriptive statistics for weekly prices and price changes, market shares, and unit sales are displayed in Table 6 in the Appendix. Importantly, the share of weeks with a different store-specific price compared to the previous week ranges between 38.6% and 56.1% across brands, indicating a high price variation of the brands in this product category and making the consideration of dynamic price effects on brand sales appear promising.

We treat cross-price effects at the tier level to reduce the model complexity on the one hand but also to account for the fact that cross-price effects are commonly much weaker than own-price effects on the other hand. We therefore capture cross-price effects by using the lowest price of the competing brands within each price-quality tier per store and week, following, e.g., Brezger and Steiner (2008). This leads to three cross-price variables representing the premium brand, national brand, and private label brand tiers, denoted as $p^{(\text{prem})}$, $p^{(\text{nat})}$, $p^{(\text{priv})}$ in the following. Note that the sets of premium and national brands considered for computing the lowest competitive price within a tier differ depending on the brand for which the sales response models are estimated. And, since the retailer's own brand is the only private label brand, sales equations for this brand only contain the two cross-price variables $p^{(\text{prem})}$ and $p^{(\text{nat})}$.

The data provide further information on the use of a display (D_{disp}) and whether a brand was offered at a 9-ending or a 99-ending price (D_9 , D_{99}) per week and store. We consider this information by using corresponding indicator variables (mutually exclusive for 9 and 99 price endings). Finally, we include a seasonality variable (m)

as additional covariate to capture seasonal effects on a (calendar) monthly level with a smooth nonlinear function.

4.2 Model estimation and validation

All models are estimated with the public domain software package BayesX (Brezger et al. 2005), and results are analyzed using the free and open source software R (R Core Team 2022). We use Gibbs sampling to generate a total of 52,000 draws from the posterior distribution, with a burn-in period of 2000 iterations and a thinning value of 50 to minimize autocorrelation of the draws, and finally save $D = 1000$ draws from the Markov chain. Model performance is assessed based on the individual parameter (Gibbs) draws instead of using the posterior means of estimated parameters (e.g., Montgomery 1997; Hruschka 2006b; Lang et al. 2015) to account for parameter uncertainty. This means that predictions $\hat{\eta}_{st}$ for log unit sales ($\log(q_{st})$) are calculated as the mean across 1000 draw-based predictions $\hat{\eta}_{st,d}$:

$$\hat{\eta}_{st} = \frac{1}{D} \sum_{d=1}^D \hat{\eta}_{st,d}. \quad (8)$$

Since predictions of a brand's unit sales (and not log unit sales) are relevant for managers, conditional mean predictions for unit sales can be obtained via $\hat{q}_{st} = \exp(\hat{\eta}_{st} + \hat{\sigma}^2/2)$, following Greene (2008, p. 100).

We compare the prediction accuracy of the different models along different error measures. One widely used measure is the Root Mean Squared Sales Prediction Error (RMSE, see, e.g., Van Heerde et al. 2001):

$$\text{RMSE} = \sqrt{\frac{1}{S} \sum_{s=1}^S \frac{1}{T_s} \sum_{t=1}^{T_s} \frac{1}{D} \sum_{d=1}^D (\hat{q}_{st,d} - q_{st})^2}. \quad (9)$$

In particular, we compute the Average Root Mean Squared Sales Prediction Error (ARMSE) in holdout samples based on a C -fold subsampling with $C = 9$ folds. For this, we randomly draw $(C - 1)/C = 8/9$ of the total sample of observations for a brand and use them for model estimation, calculate the RMSE for the remaining part (holdout), and finally average over the $C = 9$ individual RMSE values:

$$\text{ARMSE} = \frac{1}{C} \sum_{c=1}^C \sqrt{\frac{1}{S} \sum_{s=1}^S \frac{1}{T_s} \sum_{t=1}^{T_s^{(c)}} \frac{1}{D} \sum_{d=1}^D (\hat{q}_{st,d}^{(c)} - q_{st}^{(c)})^2}. \quad (10)$$

As our second measure to assess the quality of out-of-sample predictions, we introduce the Continuous Ranked Probability Score (CRPS), which has not been used before in the store sales modeling literature.

The general idea behind this scoring rule is to compare two distributions by the integral of the squared difference of their cumulative distribution functions (cdfs) (Matheson and Winkler 1976). The CRPS is, however, most often used in a nested

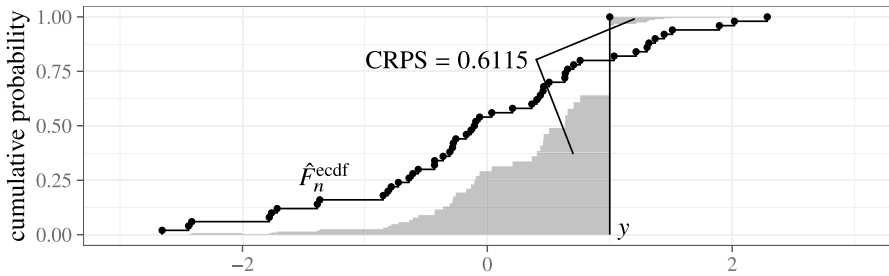


Fig. 1 Example figure for the CRPS scoring rule: the gray-shaded area represents the integrated squared difference between the empirical cumulative distribution function \hat{F}_n^{ecdf} of a sample of F and the single-point empirical cdf (ecdf) of observation y

version to compare probabilistic and deterministic outcomes. Then, it can be used to measure the squared error between a probabilistic forecast and a point measure, the former relating to the draw-based predictions for a brand’s unit sales and the latter relating to the corresponding observed sales of a brand’s unit sales in a certain store and week in our context. In this variant, the CRPS can be seen as a continuous version of the well-known Brier score that is defined as the mean squared error between a discrete outcome y and a corresponding probabilistic forecast based on the cdf F (see, e.g., Gneiting and Raftery 2007; Jordan 2016, pp. 37–38):

$$\text{CRPS}(F, y) = \int_{\mathbb{R}} (F(z) - \mathbb{1}_{z \geq y})^2 dz. \tag{11}$$

In empirical applications, the underlying true distribution F is commonly unknown. Using an empirical cdf \hat{F}_n^{ecdf} , based on n observations, as an approximation of F , the CRPS score can be computed by the following consistent modification of Eq. (11) (Jordan 2016, Sect. 6; Krüger et al. 2021):

$$\text{CRPS}(\hat{F}_n^{\text{ecdf}}, y) = \frac{2}{n^2} \sum_{i=1}^n (X_{(i)} - y) \left(n \mathbb{1}_{y < X_{(i)}} - i + \frac{1}{2} \right). \tag{12}$$

Figure 1 illustrates the concept behind the CRPS for a better understanding: as the score is calculated as the integral over squared differences between two distributions, with one of them being degenerated to a point measure, the CRPS represents the size of the gray-shaded area between the cumulative distribution functions. In other words, the distribution function for the predictive distribution is compared to the ideal distribution function of a point mass in a newly observed value by forming the squared distance and integrating over it. The more the two distribution functions overlap, i.e., the more they coincide, the lower is the distance between them.

In our case, the draws saved from the Markov chain after convergence are samples from an (unknown) posterior distribution F^{post} , and using this scoring rule we again account for parameter uncertainty. Note that we obtain one CRPS value as “counterpart” to every observation q_{st} , and averaging the CRPS values again across all weeks and stores provides us with a mean CRPS value (MCRPS) for each holdout:

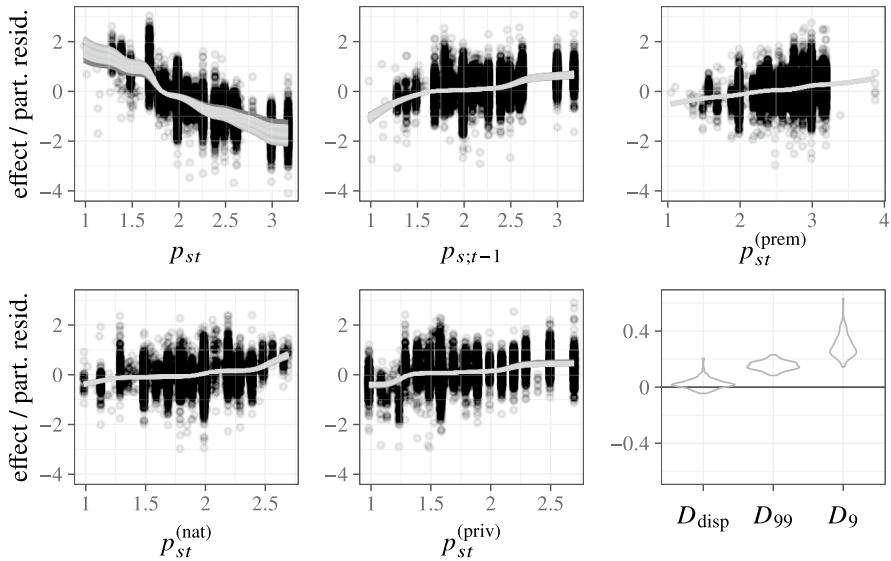


Fig. 2 Estimation results for the DYNFLEXHET model using the brand "Minute Maid" as example: estimated effects and partial residuals for own, lagged, and competitive price variables (accounting for heterogeneity via scaled spline slopes) as well as estimated effects for display and odd price endings (heterogeneous effects displayed by violin plots)

$$\text{MCRPS} = \frac{1}{S} \sum_{s=1}^S \frac{1}{T_s} \sum_{t=1}^{T_s} \text{CRPS}(F_{st}^{\text{post}}, q_{st}), \tag{13}$$

$$\Rightarrow \text{AMCRPS} = \frac{1}{C} \sum_{c=1}^C \frac{1}{S} \sum_{s=1}^S \frac{1}{T_s^{(c)}} \sum_{t=1}^{T_s^{(c)}} \text{CRPS}(F_{st}^{\text{post};(c)}, q_{st}^{(c)}). \tag{14}$$

Finally, averaging the individual MCRPS values over the C folds results in the AMCRPS value as our second predictive accuracy measure corresponding to the ARMSE. We use the R package `scoringRules` (Jordan et al. 2019) to compute the CRPS values according to Eq. (12).

4.3 Estimation results and predictive model performance

4.3.1 Estimation results

Estimation results for the most complex DYNFLEXHET model as described in Eq. (1) are illustrated for the brand “Minute Maid” as example in Fig. 2. Remember that this model accounts for *heterogeneous* effects across stores, *functional flexibility* in price effects, and *price dynamics* represented by the one-period lagged own-price effect. Depicted are the heterogeneous spline estimates together with the partial residuals as well as violin plots for the display and 9- or 99-ending price effects.

First, estimates and the corresponding effect sizes suggest face validity. The strongest effect is observed for the own-price effect, and it is also the one of the price effects showing a larger amount of heterogeneity across stores as is represented by the relatively large bandwidth of the scalings of the spline functions. More specifically, 21% of (a total of 3240) pairwise 80% credible intervals for the estimated scaling factors do not overlap between stores, confirming that own-price effects significantly differ between many stores. We will further discuss below in more detail that, as a rule, own-price effects for all eight brands are much more heterogeneous across stores than both cross-price effects and lagged price effects (see Fig. 4). The overall slope of the own-price effect reveals a threshold effect near 1.75\$ beyond which sales of “Minute Maid” strongly increase. The lagged price effect shows a steeper slope for small price levels than for medium and high price levels (where the lagged price effect is rather flat), and in addition a less distinct threshold effect at 2.50\$. This suggests that customers respond less to high prices of “Minute Maid” in the previous period, in terms of buying more of the brand in the current period, than to a very low previous price in terms of buying less of “Minute Maid” in the current period. In other words, less purchases seem to be postponed if the previous price is very high while a certain level of stockpiling appears to occur in case of a very low one, albeit these conclusions are difficult to validate with aggregate data. Some purchase deceleration however obviously exists for prices larger than 2.50\$. Compared to the own-price effects (including the lagged price effect), cross-price effects turn out rather flat with sales of “Minute Maid” being least (most) affected by the private label brand (premium brands). Interestingly, the sales effect of a 9-ending price (excluding a price ending in 99) is frequently (much) larger than the one for a 99-ending price (excluding other price endings in 9) and turns out significantly positive for 93.8% of the stores (D_{99} : 59.3%). The display effect is not significant for all stores but one.

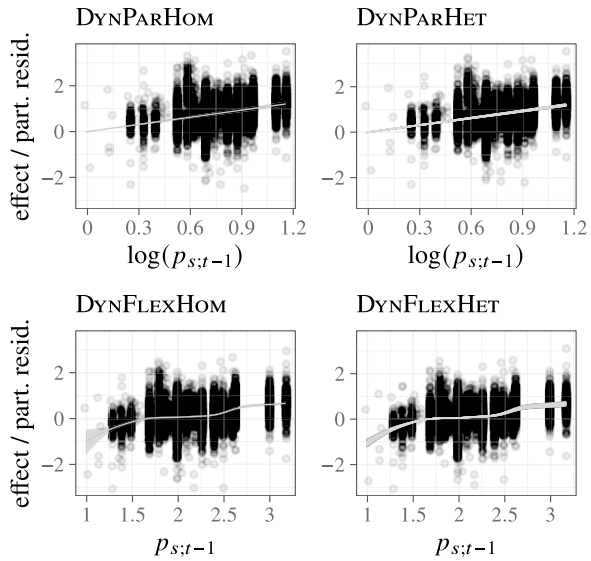
Figure 3 displays the lagged price effects in the four *dynamic* models for the brand “Minute Maid”, see the *homogeneous* and *heterogeneous parametric* models (top left and right panels) and the corresponding *flexible* model versions (bottom panels). The advantage of using a *flexible* approach instead of modeling the effect in a *parametric* way becomes obvious again, as was already visible in Fig. 2: non-linearities with piecewise steeper or flatter slopes can be modeled with Bayesian P-splines⁷ but not with parametric functions. Nevertheless, a closer look at the bottom left panel reveals wide confidence bands at the lowest price levels, where only a few observations are available.⁸

Note that there is virtually no difference in the estimated own-price effects between the respective *static* and *dynamic* model variants, as illustrated in Fig. 9 in the Appendix. This implies that the immediate own-price effect remains highly robust even if the models are extended to capture price dynamics. Plots of the estimated lagged price effects for all eight brands obtained by the DYNFLEXHET models

⁷ Note that we use the centered sampling method for spline estimation provided in BayesX such that the sum of the spline function over all observations equals zero.

⁸ We did not include the confidence bands in the lower right panel for the flexible heterogeneous dynamic model to prevent clutter. The corresponding confidence bands at the lower bound of the price range look highly similar.

Fig. 3 Estimated effect of the lagged price on the sales of the brand “Minute Maid” for the different *dynamic* model variants



are provided in Fig. 10 in the Appendix, showing very different shapes across brands and hence confirming the benefits of nonparametric estimation as well for the dynamic price effects. The complete estimation results for all brands are available from the authors upon request.

Figure 4 shows density plots of the estimated store-specific scaling or random effects parameters for the DYNPARHET and DYNFLEXHET models in order to get a deeper understanding about how much heterogeneity is inherent to our data and how

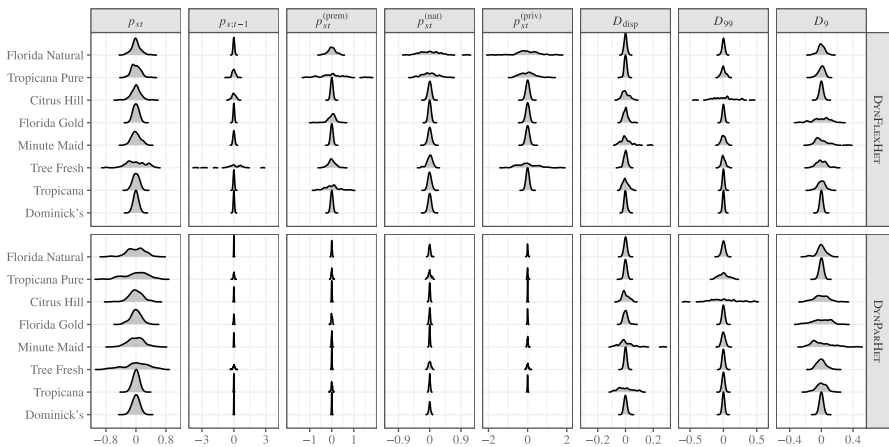


Fig. 4 Density plots of store-specific scaling or random effect parameter estimates (centered around their mean) for the DYNPARHET and DYNFLEXHET models

this heterogeneity is handled by these two model variants⁹. Some points from Fig. 4 are in particular noteworthy. First, as already illustrated in Fig. 1 for the brand “Minute Maid”, heterogeneity across stores is especially distinct for the own-price effect; this applies more or less to all eight brands independent whether the own-price effect is captured flexibly or parametrically. Second, heterogeneity between stores does not seem to be an issue for both the lagged price effect nor for cross-price effects if price effects are modeled parametrically (DYNPARHET). However, once price response is modeled flexibly both lagged price effects and cross-price effects do reveal a moderate amount of store heterogeneity for all brands. Lang et al. (2015) have previously provided a possible explanation for this phenomenon, albeit not in the context of price dynamics. And third, there is also some heterogeneity across stores in the effects of using odd prices and a display for all brands and, like for the own-price, the effects are rather stable across the two models (DYNPARHET, DYNFLEXHET).

4.3.2 Predictive performance

We evaluate the predictive performance of the competing models in terms of the Average Root Mean Squared Sales Prediction Error (ARMSE) and the Average Mean Continuous Ranked Probability Score (AMCRPS), which were introduced in Sect. 4.2. We use a 9-fold subsampling, where 8/9 of the data (randomly drawn without replacement) are used each time for model estimation and the remaining 1/9 of the data serves as holdout sample. The predictive validity results for the two performance measures are reported in Table 3 for the ARMSE measure and in Table 4 for the AMCRPS measure, together with relative improvements or deteriorations of the different models in predictive accuracy compared to the proposed DYNFLEXHET model (for each brand and aggregated at the median) and frequencies in how many cases each model provided the best performance (row ‘# best’).

From Table 3, we at first observe that different models perform best for different brands in terms of the ARMSE measure, i.e., there seems to be no clear favorite model under this error measure at first glance. For seven out of eight brands, a *flexible* model version provides the best predictive accuracy, as does a *dynamic* model version for six of the brands, while for three brands the best model is a *heterogeneous* one. On the other hand, when aggregated across brands, the proposed DYNFLEXHET model is only slightly outperformed by the DYNFLEXHOM model and provides the second-best predictions, see the median in relative changes in the ARMSE measure at the bottom of the table.

In Table 4, the picture is completely different and very clear. Using the CRPS measure to assess the predictive model performance, the most complex model (DYNFLEXHET) always provides the most accurate predictions, and concerning the three model dimensions *functional flexibility*, *store heterogeneity*, and *price dynamics* the more complex model variant outperforms its simpler counterpart (i.e.,

⁹ We here abstain from displaying the corresponding density plots for the *static heterogeneous* counterpart models (STATPARHET and STATFLEXHET) because there is virtually no difference between the *static* and *dynamic* model variants except that the lagged price is not included in the *static* models.

Table 3 Out-of-sample predictive performance of the competing models evaluated by the Average Root Mean Squared Sales Prediction Error (ARMSE) in holdout samples compared to the proposed DYNFLEX-HET model. Best models per brand are marked in bold

	STAT				DYN			
	PAR		FLEX		PAR		FLEX	
	HOM	HET	HOM	HET	HOM	HET	HOM	HET
Flor. Natrl.	26.5 (−0.2%)	26.8 (+1.1%)	25.4 (−4.5%)	28.5 (+7.5%)	23.8 (−10%)	23.8 (−10%)	24.6 (−7.2%)	26.6 −
Tropic. Pure	57.4 (+12%)	57.1 (+12%)	53.0 (+3.8%)	52.1 (+1.9%)	55.9 (+9.4%)	55.1 (+7.8%)	52.2 (+2.1%)	51.1 −
Citrus Hill	120.7 (+12%)	122.1 (+14%)	101.1 (−5.8%)	135.0 (+26%)	104.7 (−2.5%)	107.1 (−0.3%)	86.0 (−20%)	107.4 −
Flor. Gold	57.5 (+6.8%)	57.7 (+7.3%)	55.6 (+3.4%)	55.9 (+3.9%)	55.5 (+3.1%)	55.8 (+3.7%)	53.4 (−0.8%)	53.8 −
Min. Maid	53.8 (+1.8%)	53.3 (+0.9%)	52.5 (−0.6%)	52.0 (−1.6%)	54.2 (+2.6%)	53.6 (+1.4%)	53.2 (+0.7%)	52.8 −
Tree Fresh	83.6 (+4.8%)	83.1 (+4.2%)	64.4 (−19%)	83.7 (+5.0%)	81.1 (+1.7%)	80.6 (+1.1%)	64.5 (−19%)	79.7 −
Tropicana	110.4 (+17%)	112.3 (+19%)	105.4 (+11%)	108.1 (+14%)	100.9 (+6.7%)	101.9 (+7.7%)	93.7 (−1.0%)	94.6 −
Dominick’s	311.7 (+11%)	316.1 (+12%)	307.5 (+9.4%)	313.8 (+12%)	300.9 (+7.0%)	306.6 (+9.1%)	278.5 (−0.9%)	281.1 −
median	+8.9%	+9.5%	+1.4%	+6.2%	+2.8%	+2.5%	−1.0%	−
# best	−	−	1	1	−	1	4	1

The bottom lines show the median relative improvement / deterioration of each model (i.e., across brands) compared to the DYNFLEXHET model and the number of times a considered model outperformed the other models

heterogeneous models perform better than *homogeneous* ones, *flexible* models better than *parametric* ones, and *dynamic* models better than *static* ones).

To validate our conjecture that the ARMSE measure is more prone to extreme observations (e.g., very high sales at very low prices) than the AMCRPS measure, we computed the Average Root Median Squared Error (ARMedSE) as further performance measure, which is known to be much more robust against extreme observations compared to the ARMSE (Franses and Ghijssels 1999). According to the ARMedSE measure, the DYNFLEXHOM model performs best for all brands, and *dynamic* and *flexible* models almost always perform better than their *static* and *parametric* counterparts – just as when using the AMCRPS. Different from the AMCRPS results, *homogeneous* models are preferred to *heterogeneous* ones in

Table 4 Out-of-sample predictive performance of the competing models evaluated by the Average Mean Continuous Ranked Probability Score (AMCRPS) in holdout samples compared to the proposed DYN-FLEXHET model. Best models per brand are marked in bold

	STAT				DYN			
	PAR		FLEX		PAR		FLEX	
	HOM	HET	HOM	HET	HOM	HET	HOM	HET
Flor. Natrl.	10.1 (+24%)	9.8 (+20%)	9.5 (+17%)	8.6 (+6.0%)	9.2 (+14%)	8.8 (+8.1%)	9.0 (+11%)	8.1 -
Tropic. Pure	24.4 (+23%)	23.2 (+17%)	22.8 (+15%)	20.8 (+4.5%)	23.3 (+17%)	21.8 (+9.6%)	22.1 (+11%)	19.9 -
Citrus Hill	28.9 (+53%)	27.8 (+47%)	24.7 (+31%)	21.9 (+16%)	24.8 (+32%)	23.6 (+25%)	21.5 (+14%)	18.8 -
Flor. Gold	20.4 (+14%)	19.9 (+11%)	19.5 (+8.5%)	18.8 (+4.4%)	19.6 (+8.8%)	19.2 (+6.4%)	18.6 (+3.5%)	18.0 -
Min. Maid	21.8 (+17%)	20.6 (+11%)	20.9 (+12%)	19.3 (+3.9%)	20.8 (+12%)	19.5 (+4.8%)	20.1 (+8.1%)	18.6 -
Tree Fresh	19.2 (+42%)	18.5 (+37%)	15.9 (+17%)	13.7 (+0.8%)	18.8 (+39%)	18.1 (+33%)	15.8 (+17%)	13.5 -
Tropicana	55.0 (+25%)	53.3 (+21%)	52.0 (+18%)	49.8 (+13%)	50.9 (+15%)	49.4 (+12%)	46.7 (+5.8%)	44.2 -
Dominick's	150.9 (+18%)	148.3 (+16%)	145.0 (+13%)	140.2 (+9.4%)	144.4 (+13%)	141.4 (+10%)	134.6 (+5.0%)	128.2 -
median	+24%	+19%	+16%	+5.2%	+14%	+9.9%	+9.5%	-
# best	-	-	-	-	-	-	-	8

The bottom lines show the median relative improvement / deterioration of each model (i.e., across brands) compared to the DYNFLEXHET model and the number of times a considered model outperformed the other models

nearly all cases.¹⁰ Based on the overall consideration of the results for all three predictive validity measures (ARMSE, AMCRPS, ARMedSE) we can recommend the use of either the DYNFLEXHET or DYNFLEXHOM for sales predictions.

4.3.3 Further tests and robustness checks

We also conducted tests for all models (by brand) to assess the assumptions required for a proper estimation of them (multicollinearity, heteroskedasticity,

¹⁰ The ARMedSE results can be obtained from the authors upon request. We further conducted a small simulation study to test the sensitivity of the CRPS in comparison to the RMSE against outliers. For this, we first generated 100 observations from the standard normal and added in a second step one or two outliers to the observations. The RMSE increased considerably, while the CRPS remained highly robust every time. This means that the CRPS weighs central observations more strongly than more extreme observations compared to the RMSE.

autocorrelation), and briefly summarize results for the proposed DYNFLEXHET model in the following.¹¹ Multicollinearity was not a problem in any case: variance inflation factors (VIF) were never higher than 2.5 across brands and thus far from being critical, where as a rule the highest VIFs could be observed for the own-item price covariates (immediate and lagged prices). Heteroskedasticity was evaluated using the Brown-Forsythe test, which is robust against deviations from normality (Brown and Forsythe 1974). We applied the test to compare the variance of the residuals against the fitted values, and results were more mixed here. The Brown-Forsythe test was not significant for four brands, indicating that heteroskedasticity is not an issue there (e.g., with $p = 0.82$ for “Minute Maid”), while it turned out significant for the other four brands ($p < 0.01$). We further used a modified version of the Durbin-Watson test for panel data to assess autocorrelation of the residuals, which is as well robust against deviations from the normal distribution (Bhargava et al. 1982; Ali and Sharma 1993). Given the number of covariates and sample size per brand (between 6624 and 6827 observations), the lower critical value around 1.85 is slightly undercut for two brands ($\alpha = 0.05$), indicating positive autocorrelation only for these two brands. We found no evidence for negative autocorrelation across brands. Heteroskedasticity and autocorrelation lead to biased estimates of standard errors and can affect the validity of statistical tests, with the result of a possibly misleading statistical inference. However, as our focus lies on the predictive model performance and related profit implications, violations of these assumptions seem less severe.

We further performed robustness checks starting with modifications for the specification of the P-splines. By default, we used 20 knots and B-spline basis functions of degree 3 to estimate price effects flexibly. In order to assess the sensitivity of the predictive power (AMCRPS, ARMSE) of the DYNFLEXHET and DYNFLEXHOM models depending on the number of knots and/or the degree of the B-spline basis functions, we further estimated them with a lower degree of the B-splines (degree 1) and/or more knots (40 knots, corresponding to the suggested upper bound by Eilers and Marx 1996) for the immediate and lagged own-price effects. The CRPS scores turned out highly robust for all brands and, except for one brand, the RMSE measure as well, indicating that 20 knots are sufficient and B-splines of degree 1 would perform comparable. Details are provided in Tables 7 and 8 in the Appendix.

Finally, we considered only one time lag in the dynamic model versions to consider price dynamics. By definition, this implies that only short-term price-change or short-lived post-promotion effects can be analyzed, while longer persisting marketing effects cannot be accommodated. To test if the data support higher order autoregressive structures, we added two more lags for the own price to the DYNFLEXHET and DYNFLEXHOM models (with the default settings of 20 knots and degree 3 for the P-splines), capturing the higher-order lag dynamics (lag 2, lag 3) parametrically to keep the model complexity manageable. Again, the CRPS scores remained extremely robust in both models except for the store brand, where adding a second price lag somewhat improved the predictive performance. In terms of ARMSE,

¹¹ The complete results for all tests and models can be obtained from the authors upon request. Note that multicollinearity measures only differ across brands and not across models, because all models share the same covariates per brand.

adding more lags did not pay off for all brands in the DYNFLEXHOM model and for six brands (including the store brand) in the DYNFLEXHET model. Details about these dynamic model extensions are provided in Tables 9 and 10 in the Appendix.

4.4 Optimization

4.4.1 Settings and constraints

In order to find optimal (i.e., profit maximizing) store-specific prices for each brand m , we use the dynamic program (P) introduced in Sect. 3.3. As the action space for determining a brand’s optimal price in period t is independent of the current state (i.e., the brand’s price in the previous period), the set of generally possible prices for the brand is constant over time, hence in each period the same with $X_t(s_t) = X$. And since $s_{t+1} = g(x_t, s_t) = x_t$, the state space equals the action space ($S = X$). For each brand and store, we set individual state and action spaces as to contain a grid of possible prices in 0.01\$ steps between the minimum and maximum observed price levels, denoted as $[p_{m,i,min}^{obs}; p_{m,i,max}^{obs}]$, for the considered brand m in store i :

$$S = S_{mi} = \left\{ p_{m,i,min}^{obs}, p_{m,i,min}^{obs} + 0.01, \dots, p_{m,i,max}^{obs} \right\} = X_{mi} = X. \tag{15}$$

The restriction of optimized prices to the range of observed price levels is chosen to preserve the retailer’s price image perceived by customers. It further helps to obtain realistic predictions, especially when using flexible functions that fit the underlying data locally such that extrapolations could be still more problematic compared to parametric functions.

To start the recursive process and solve the dynamic optimization problem, the initial state $s_1 = p_0$ has to be fixed. As starting value, we decided to use the price in the middle of the observed price range of a considered brand and store. By conducting some sensitivity analyses, we checked that optimal price paths were fairly robust against variations of this starting value. We further set the discount rate r to 0 without loss of generality.

As a second constraint, we limit the predicted sales \hat{q} of a brand in a certain store to the largest number of sales $q_{m,i,max}^{obs}$ observed for this brand in this store in our data. Using this second constraint lets us stay conservative and prevents us from an overestimation of a brand’s unit sales beyond the upper bound of observed unit sales per week and store for (very) low prices. This guarantees realistic sales scenarios as observed in the data on the one hand, but may trim the greater flexibility of the spline estimates on the other hand. In other words, if the sales forecast for a brand in a store from one of our estimated models is higher compared to the maximum observed sales, we truncate the prediction to the maximum of observed sales:

$$\hat{q} = \min \left\{ \hat{q}, q_{m,i,max}^{obs} \right\}. \tag{16}$$

Note that predictions \hat{q} are again made at the draw level as already described in Eq. (8), i.e., we integrate the optimization criterion over the posterior distribution of the parameters by averaging the optimization criterion across draws, before optimizing it. To further speed up optimization, we use every tenth draw from the saved draws of the Markov chain after convergence.

4.4.2 Optimal price paths

We use forward recursion to solve the optimization problems and generate 5184 optimal price paths in total (81 stores times 8 brands times 8 sales response models). It seems reasonable to assume that optimal price paths might vary depending on the store-specific scaling factors $(1 + \alpha_{s1})$. For illustration, we continue with the brand “Minute Maid” as example and choose for this brand three out of the 81 stores with a low vs. medium vs. high scaling factor for the own-price effect estimated by the DYNFLEXHET model.

Figure 5 shows optimized price paths for the brand “Minute Maid” for the three selected stores and all *heterogeneous* model variants. Further depicted are variable costs (also provided in the data) and observed prices as benchmarks for optimized prices (panels in the top row). Observed prices show a very high variation during the first 25 weeks, switching back and forth mostly between only two different price levels with no more than three consecutive weeks at the high price level and no more than two consecutive weeks at the lower price level (note that the high price level is highest in the store with the high scaling parameter, 3.17\$, and lowest for the store with the low scaling parameter, 2.62\$.) This clear hi-lo price pattern dilutes after the first 25 weeks, but some price levels still show up more often than others (e.g., 1.99\$ in all three selected stores), and price cuts of varying depths continue to occur. Also note that the general price level of “Minute Maid” decreases with decreasing variable costs (wholesale prices), as expected.

The optimized prices resulting from the STATPARHET model (red lines in the middle row) “follow” the costs very closely, i.e., increased costs lead to increased prices and vice versa, and the correlation between observed costs and optimized prices (0.92) is considerably higher than the correlation between observed costs and observed prices (0.60). It is further reasonable that optimized prices end in 9 or 99 since odd pricing was shown to have positive effects on expected sales (see Sect. 4.3.1). This optimized price pattern in parts also holds for the STATFLEXHET model (middle row; blue lines), but with a more clear hi-lo pricing scheme and a smaller number of different price levels. Completely different optimal price paths are obtained for the *dynamic* models. For the store with the low scaling parameter for the own-price effect (bottom left panel), optimized prices resulting from the DYNPARHET model vary in a much smaller bandwidth. Small price cuts are observed primarily in the second half of the data time window, and the high price level is constant over time. Optimized prices obtained from the DYNFLEXHET model hardly show any variation (bottom row; blue lines). Deep price cuts are observed in only three weeks, in all other weeks the optimal price is set consistently high at the same unique price level as suggested by the DYNPARHET model (bottom row; red lines).

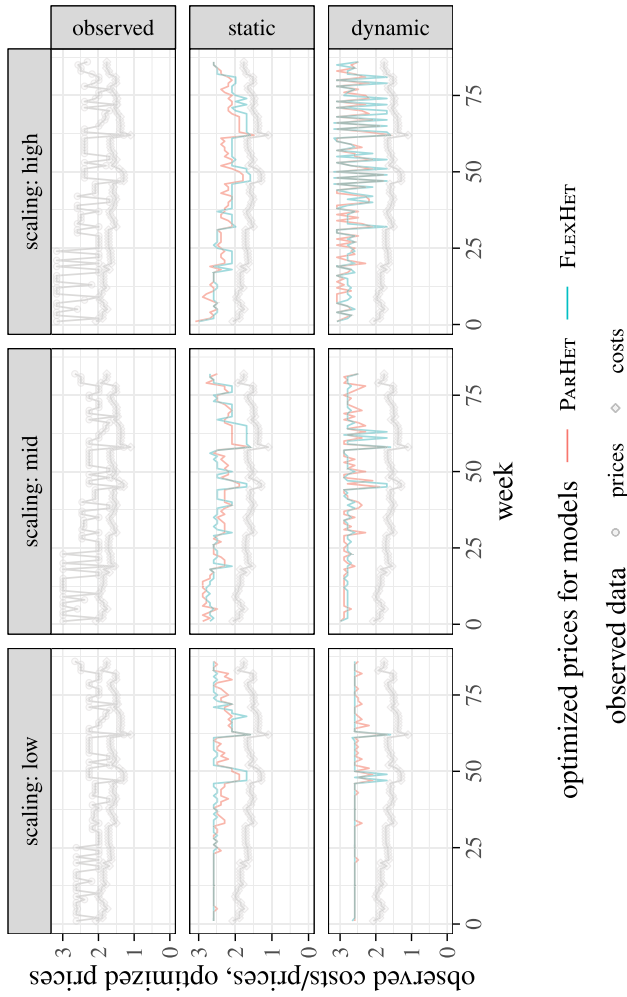


Fig. 5 Optimal price paths for the brand “Minute Maid” for three selected stores based on the DYNFLEXHET model, coming up with a low, medium, or large scaling factor $(1 + \alpha_1)$. Variable costs (wholesale prices) and observed store-specific prices are displayed in the top row, followed by the optimized price paths for the heterogeneous *static* (middle row) and *dynamic* (bottom row) model variants

Note that the optimal high price level resulting from the DYNPARHET and DYNFLEXHET models, also interpretable as suggested “regular” price, is a few cents below the maximum observed price, which served as constraint for the upper bound of optimized prices in the optimization model (see Sect. 4.4.1). A similar optimal pricing pattern but with more frequent price cuts from a higher optimal regular price level is found for the store with a medium scaling parameter (compared to the store with the low scaling parameter). Finally, the store with the high scaling parameter for own-price effects shows a much more distinct hi-lo pricing pattern with nearly weekly price changes and very deep price cuts in the second half of the considered time window. Interestingly, the deepest price cuts below a price of 1.75\$ are observed at variable costs lower than 1.50\$, while the highest suggested price levels exceed 3.00\$ in this store. Note that (only) six of these optimal high price levels determined by the DYNFLEXHET model hit the upper boundary of the observed price range for “Minute Maid” in this store (and none in the other two stores displayed).

We find similar optimal pricing patterns for most of the other brands. For some of them, optimized prices vary only between two or three different levels (especially for the *flexible* models). Across models and brands, we find that the number of different optimal price levels along the price paths is considerably smaller than the number of observed price levels in the data. Since the number of optimal price levels does not seem to vary that much, it seems interesting to compare optimized prices between different types of models in more detail, i.e., to see how optimized prices compare for the *static*, *parametric*, or *homogeneous* model variants versus the ones of the corresponding *dynamic*, *flexible*, or *heterogeneous* models. We contrast all competing models in a pairwise manner in Fig. 11 in the Appendix and leave a detailed analysis to the interested reader.

The corresponding results are not generalizable across brands. In a nutshell, we find high consistency in optimized prices between the different models only for one brand, i.e., where competing models share the same optimized prices very frequently. The opposite applies to “Minute Maid” (our brand for illustration), where the competing models often suggest different optimal price levels. Larger shares between 42% and 69% for identical optimized prices are observed only between *homogeneous* models and their corresponding *heterogeneous* counterparts for this brand. Note that, for many brands, *homogeneous* and *heterogeneous* counterpart models more often share the same optimal price levels in a particular store and week than can be observed for models that differ in the other dimensions (*parametric* vs. *nonparametric* price effects, *static* vs. *dynamic* models). However, even this is not generalizable for the data at hand, as counterexamples exist. Importantly, optimized and observed prices per store and week differ in at least 88% of cases for all brands and models.

To analyze whether or how the existing price image of a brand (and consequently of the retailer) is influenced when moving from the observed to the optimized prices or price paths, we compute the difference between optimized and observed price levels $(p^* - p^{\text{obs}})/p^{\text{obs}}$ per store and week relative to the observed price and collect these ratios via box plots in Fig. 6 (*homogeneous* models are omitted here for the sake of clarity). The values on the x-axis in Fig. 6 correspond to these percentage changes, where relative increases versus decreases are arranged symmetrically. That is, an optimized

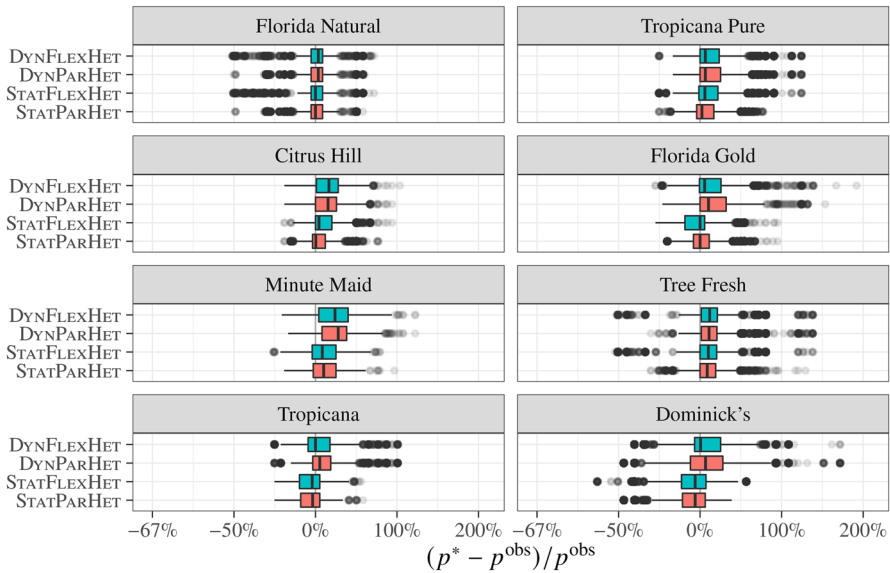


Fig. 6 Box plots of differences between optimized and observed price levels relative to the observed price $(p^* - p^{obs})/p^{obs}$ per store and week for the heterogeneous models

price half as high compared to the observed price (−50%) appears equally distant from the zero point as a price twice as high (+100%). The figure shows that price changes of more than −33% or +50% occur very seldom. Furthermore, relative price changes are rather symmetrically distributed with slight tendencies to relative price increases (i.e., larger optimized than observed prices), preferably for the *dynamic* model versions. Overall, the price image of the brands would not be affected too much from a move to optimized prices p^* from observed prices.

4.4.3 Profit implications

Up to now, we have searched for optimal (dynamic) prices or price paths to maximize *total* individual brand-store profits for the retail chain. In the following, we evaluate related profit implications along three dimensions:

$$\text{Observed profits: } V_{mi}^{obs} = \sum_t v_{mit}^{obs}, \tag{17}$$

$$\text{Predicted profits: } \hat{V}_{mi}^{obs} = \sum_t \hat{v}_{mit}(p^{obs}), \tag{18}$$

$$\text{Optimized profits: } \hat{V}_{mi}^* = \sum_t \hat{v}_{mit}(p^*). \tag{19}$$

Equation (17) represents the actual total profit of a brand m in store i based on its observed pricing strategy as given in the data, with v_{mit}^{obs} denoting the corresponding observed profits in week t . In Eq. (18), we compute expected total profits of a brand m in store i again based on its observed pricing but now using our estimated sales response models, where $\hat{v}_t(p^{\text{obs}}) = (p_t^{\text{obs}} - c_t) \cdot \hat{q}_t(p_t^{\text{obs}}, p_{t-1}^{\text{obs}})$. The comparison between actual and model-based predicted profits provides us with an assessment of a possible prediction bias for our models, which would translate into the prediction of total expected optimal profits, too. The latter are calculated by Eq. (19) in the same manner as in Eq. (18) but using optimized prices p_t^* instead of observed prices and the resulting predicted sales $q_t^*(p_t^*, p_{t-1}^*)$ (or $q_t^*(p_t^*)$ for the *static* model variants).

Table 5 summarizes the results for observed, predicted, and optimized (expected) profits for each model and brand, along with summary statistics across brands at the bottom of the table. For each brand, the first row shows predicted mean profits $\hat{V}_{mi}^{\text{obs}}$ averaged across stores by model type (cf. Eq. (18)) together with the relative increase or decrease compared to the observed mean profits across stores $((\hat{V}_{mi}^{\text{obs}} - V_{mi}^{\text{obs}})/V_{mi}^{\text{obs}})$. In a similar way, the second row reports optimized mean profits (cf. Eq. (19)) together with the related percentage changes compared to the predicted mean profits based on observed prices. As an example, predicted mean profits for “Minute Maid” based on the DYNFLEXHET model amount to 1611, a value that is only 2% lower compared to the actual mean profit resulting from the observed pricing strategy (1639, not reported in the table). Contrary to that, the optimized mean profit based on the DYNFLEXHET model (i.e., using optimized prices based on this model instead of observed prices) are much higher (2976), corresponding to an increase in expected profits of about 86% over the status quo pricing strategy for “Minute Maid”.¹²

The first important finding from Table 5 is that observed and predicted profits (based on observed prices) are very close across brands and models. Most of the predicted profits do not differ more than 3% in both directions as a rule, indicating that the two constraints imposed on the price optimization procedure are reasonable and prevent our models from an over- or underestimation of total profits. The only exception is “Citrus Hill”, where all *parametric* models consistently underestimate the observed profit by about 10% (note that this is not the case for all *flexible* models). Opposed to that, optimized profits are substantially higher than predicted profits (based on observed prices), with expected increases ranging between 19% and 86%. Optimized profits are highest under the DYNFLEXHET model for five brands.¹³ For seven out of eight brands, the highest expected optimal profits are provided by *flexible heterogeneous* models (DYNFLEXHET, or STATFLEXHET). For six brands, expected optimal profits are highest based on models including *price dynamics*. And the profit-maximizing model by brand is always a *heterogeneous* one. Aggregated across brands, the

¹² The reported percentage changes are averaged across stores as well, i.e., they cannot be simply recomputed from the numbers reported in the table.

¹³ Note that models with a better predictive performance are not necessarily expected to provide higher profits.

Table 5 Comparison of optimal profits \hat{V}_{mi}^* with actual and predicted profits (based on observed prices), i.e., V_{mi}^{obs} and \hat{V}_{mi}^{obs} , reported as averages across stores

Brand		STAT				DYN			
		PAR		FLEX		PAR		FLEX	
		HOM	HET	HOM	HET	HOM	HET	HOM	HET
Flor. Natrl.	$\overline{\hat{V}_m^{obs}}$	1675 (-0.4%)	1671 (-0.6%)	1705 (+1%)	1697 (+1%)	1681 (±0%)	1677 (-0.2%)	1701 (+1%)	1693 (+0.9%)
	$\overline{\hat{V}_m^*}$	2655 (+59%)	2664 (+60%)	2822 (+67%)	2842 (+69%)	2515 (+50%)	2525 (+51%)	2594 (+53%)	2576 (+53%)
Tropic. Pure	$\overline{\hat{V}_m^{obs}}$	3769 (-0.3%)	3768 (-0.4%)	3754 (-0.6%)	3744 (-0.8%)	3757 (-0.6%)	3754 (-0.8%)	3748 (-0.8%)	3738 (-1%)
	$\overline{\hat{V}_m^*}$	4620 (+22%)	4686 (+24%)	4926 (+31%)	5012 (+33%)	4853 (+29%)	4982 (+32%)	5159 (+37%)	5322 (+41%)
Citrus Hill	$\overline{\hat{V}_m^{obs}}$	1550 (-12%)	1543 (-13%)	1820 (+2%)	1853 (+2%)	1600 (-9%)	1586 (-10%)	1776 (-0.3%)	1815 (-0.3%)
	$\overline{\hat{V}_m^*}$	2132 (+38%)	2323 (+50%)	2603 (+46%)	2881 (+55%)	2646 (+65%)	2800 (+73%)	2735 (+55%)	3024 (+67%)
Flor. Gold	$\overline{\hat{V}_m^{obs}}$	1416 (-2%)	1409 (-2%)	1400 (-3%)	1409 (-4%)	1399 (-3%)	1393 (-3%)	1397 (-3%)	1388 (-4%)
	$\overline{\hat{V}_m^*}$	1864 (+32%)	1908 (+36%)	1775 (+26%)	1882 (+35%)	2057 (+46%)	2098 (+49%)	1882 (+34%)	1924 (+38%)
Min. Maid	$\overline{\hat{V}_m^{obs}}$	1667 (+2%)	1658 (+1%)	1647 (+0.6%)	1639 (+0.2%)	1640 (+0.3%)	1633 (-0.2%)	1617 (-1%)	1611 (-2%)
	$\overline{\hat{V}_m^*}$	2443 (+48%)	2524 (+54%)	2396 (+48%)	2486 (+54%)	2740 (+69%)	2815 (+73%)	2876 (+80%)	2976 (+86%)
Tree Fresh	$\overline{\hat{V}_m^{obs}}$	1975 (+2%)	1970 (+2%)	1934 (-0.3%)	1930 (-0.7%)	1974 (+2%)	1969 (+2%)	1935 (-0.2%)	1932 (-0.6%)
	$\overline{\hat{V}_m^*}$	2335 (+19%)	2351 (+20%)	2741 (+41%)	2882 (+48%)	2491 (+26%)	2489 (+27%)	2786 (+43%)	2873 (+48%)
Tropicana	$\overline{\hat{V}_m^{obs}}$	4151 (+0.6%)	4144 (+0.3%)	4158 (+0.6%)	4139 (±0%)	4130 (-0.1%)	4121 (-0.4%)	4126 (-0.3%)	4108 (-0.8%)
	$\overline{\hat{V}_m^*}$	6160 (+50%)	6217 (+52%)	6221 (+50%)	6305 (+53%)	5931 (+44%)	5998 (+46%)	6351 (+54%)	6434 (+57%)
Dominick's	$\overline{\hat{V}_m^{obs}}$	8927 (+2%)	8919 (+2%)	8828 (+0.5%)	8830 (+0.4%)	8888 (+1%)	8884 (+1%)	8758 (-0.4%)	8757 (-0.6%)
	$\overline{\hat{V}_m^*}$	11681 (+31%)	11694 (+31%)	12404 (+41%)	12452 (+41%)	11863 (+34%)	11888 (+34%)	13186 (+51%)	13231 (+51%)
median	$\overline{\hat{V}_m^{obs}}$	+0.1%	-0.1%	+0.6%	+0.1%	±0.0%	-0.3%	-0.4%	-0.7%

Table 5 (continued)

Brand	STAT	DYN							
		PAR		FLEX					
		HOM	HET	HOM	HET				
median	\widehat{v}_m^*	+35%	+43%	+43%	+51%	+45%	+48%	+52%	+52%
# best		–	–	–	2	–	1	–	5

The relative increases or decreases in $\widehat{v}_m^{\text{obs}}$ are related to the actual (observed) profit, and the relative changes of \widehat{v}_m^* to the predicted profits (based on observed prices), averaged across stores as well. Bold values indicate the model leading to the highest optimized profits by brand

median relative improvements in optimized profits over predicted profits (based on observed prices) are highest for the DYNFLEXHET and DYNFLEXHOM models (+52%), see the corresponding summary statistics at the bottom of the table. Note that the mean relative improvement (not displayed in the table) is highest for the DYNFLEXHET model (+55%) followed by the DYNFLEXHOM model (+51%), reflecting the better individual performance of the DYNFLEXHET model at the brand level (row “# best”). Further, as a rule, expected increases in profits based on optimal prices differ less between models by brand than between brands by model.

Relating to the classification of models given in Table 2, our results on expected profits can be summarized as follows. With the exception of one brand where optimal profits are consistently higher for *static* models (“Florida Natural”), *dynamic* model versions almost always lead to (much) higher optimal profits than their *static* counterparts. Also, optimized profits based on *heterogeneous* models usually turn out (sometimes considerably) larger compared to their *homogeneous* counterparts. Finally, *flexible* modeling (greatly) pays off for most brands compared to *parametric* modeling, and *always* provides (much) larger optimal profits once price dynamics have been accommodated. Our findings with regard to profit implications can also be confirmed on the disaggregate store level for six out of eight brands, where the pairwise comparisons between STAT/DYN, PAR/FLEX, and HOM/HET models lead to better values for the more complex variants in more than 50% (and often up to over 90%) of the cases.

4.4.4 Expected losses

In the previous subsection, we analyzed profit implications for each model based on the respective optimized prices determined for this model. But which model should be ultimately used for setting prices? As mentioned in the introduction, Van Heerde et al. (2002) have recommended that managers should rely on the model with the most accurate predictions for related decision-making. Our discussion on the predictive performance of the competing models in Sect. 4.3.2 has however produced an unambiguous picture insofar that no unique best model existed across the three

different predictive validity measures used. Nevertheless, more complexity always provided a better forecasting accuracy according to the AMCRPS measure, suggesting that the DYNFLEXHET model was the model of choice here, while the ARMedSE measure unequivocally tended to the DYNFLEXHOM model instead. Regarding the ARMSE measure, at least one of these two models (DYNFLEXHET, DYNFLEXHOM) also performed best for five brands as well as not dramatically worse for the other three brands. Therefore, without loss of generality, we choose the DYNFLEXHET model to compute expected losses next.

In particular, we take the optimal prices for a brand determined by the DYNFLEXHET model and calculate the total (store) profit for this brand (i.e., summing up over all weeks) under the DYNFLEXHET model. Expected losses in total (store) profit can then be computed when using optimal prices determined by a *different* model but inserted in the DYNFLEXHET model for expected profit calculations (cf., e.g., Lang et al. 2015). Note that expected store profits from the DYNFLEXHET model based on optimal prices of a different model can turn out higher than expected profits from the DYNFLEXHET model based on its true (i.e., own) optimal prices in particular weeks but not when summed up over all weeks. Therefore, expected losses in *total* (store) profits are negative by definition compared to the assumed “true” model.

Figure 7 shows box plots for expected losses by stores for the different *heterogeneous* model alternatives (*homogeneous* models are omitted in the figure as they provided patterns highly similar to those of their *heterogeneous* counterparts). One important finding here is that expected losses relative to the DYNFLEXHET model are generally not larger than 10% for the DYNPARHET model across all brands (although even a loss of “only” 10% can translate into a huge loss expressed in monetary units). The least complex heterogeneous model (STATPARHET) either leads to the highest expected losses for six brands (with a maximum median expected loss of 26% and largest individual store losses of up to -40% across brands) or expected losses similarly high as for its flexible counterpart (STATFLEXHET). The DYNFLEXHOM model on the other hand provides the most inconsistent results here: it shows (much) higher expected losses than the DYNPARHET model for five brands but performs excellently for the other three brands, where it comes up with only very small or almost negligible expected losses. Overall, the DYNPARHET seems to be the most robust model variant to reduce expected losses relative to the proposed DYNFLEXHET model. Therefore, at least for the data at hand, the use of a *heterogeneous dynamic* model for pricing decisions is strongly suggested. The main (preliminary) conclusion resulting from our optimization exercise is that nonlinear heterogeneous dynamic pricing provides better pricing decisions and higher expected chain profits, and can prevent losses due to suboptimal pricing. That’s why we recommend to prefer the proposed DYNFLEXHET model.

Finally, it is important to note that optimizing profits or evaluating expected losses based on an econometric model like we do is prone to the Lucas critique, at first glance. Transferred to our context, Lucas (1976) states that since the structure of an econometric model involves optimal decision rules of brand managers and since these rules change systematically with the structure of the time series data relevant to the retailer’s policy, any change in the retailer’s policy will alter the structure of the econometric model (Lucas 1976, p. 41). However, the focus of this paper does

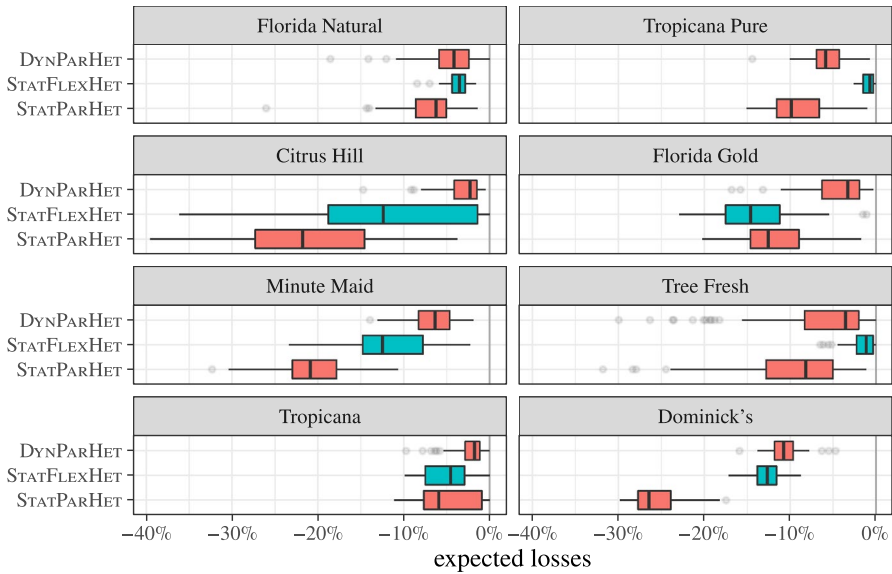


Fig. 7 Box plots of expected losses by store (aggregated over weeks) for the heterogeneous model variants using the DYNFLEXHET model as “true” model

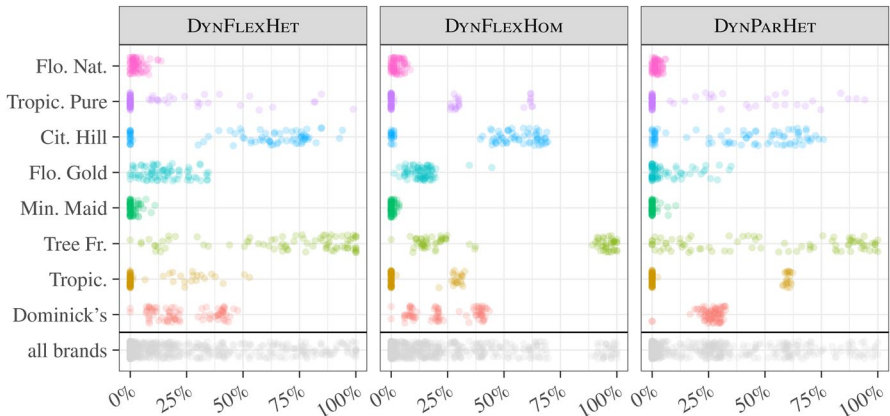


Fig. 8 Distributions of store-specific shares of optimized prices hitting the upper bound of the observed price range for the *dynamic* model variants DYNPARHET, DYNFLEXHOM, and DYNFLEXHET

not lie on the evaluation of the retailer’s pricing policy but on the comparison of the statistical and normative capabilities of different variants of econometric models and in particular on assessing the benefits of including functional flexibility in price response, store heterogeneity, and/or price dynamics in a sales response model.

4.4.5 Boundary pricing effects

All optimization results discussed above would be of less value if optimized prices always or frequently hit the boundaries of the observed store-specific price ranges. In the following, we illustrate the boundary pricing issue for the following three models: the DYNFLEXHET, which provided the highest expected profits for most brands (and as well on average across brands) and the most accurate sales predictions based on the CRPS score; the DYNFLEXHOM, which as well showed an excellent forecasting accuracy based on all three predictive validity statistics (CRPS, RMSE, RMedSE); and the DYNPARHET, which consistently (i.e., across brands) led to the least expected losses compared to the DYNFLEXHET. We included the latter model also because there seemed to be evidence from our literature search that dynamic parametric models were especially prone to upper bound corner solutions, compare Sect. 2. Figure 8 displays the distributions of store-specific shares of optimized prices hitting the upper bound of the observed price range, both by brand and across all brands. The plots indicate that boundary pricing is not a general problem by brand or model (in the sense that the boundary is always or mostly hit for certain brands or models) and that the occurrence of corner solutions is not critical for most brands. For our example brand “Minute Maid”, optimized prices almost never hit the upper bound in most stores and in the worst case in 12% of the weeks in a store. For two brands (“Citrus Hill”, “Tree Fresh”), however, the upper price bound is hit (much) more often in a larger number of stores. On the other hand, we observe highly different shares of boundary hits across the stores for these two brands (especially for the *heterogeneous* models), including stores with zero share (no upper boundary hits) as well. Across brands and the three *dynamic* models shown, 50% (75%) of the stores show upper boundary price hits in less than 10% (35%) of the weeks (not displayed in the figure). As it is further obvious, the upper price bound patterns differ more by brands than by models, which consistently applies for the other models not shown, too.

Our findings contradict the assumption that *parametric* models or models with linear price effects might be generally prone to boundary pricing effects. However, we consistently find more stores with higher shares of upper boundary price hits for the *dynamic* models compared to their *static* counterparts. As a rule, we also see a (much) greater bandwidth of stores with different shares of upper boundary price hits for *heterogeneous* models compared to their *homogeneous* counterparts across brands, as one might have expected. No such systematic can be derived for *flexible* versus *parametric* models.

In contrast to upper boundary price hits, lower boundary price hits as well as upper boundary sales hits were hardly observed. The share of lower bound price hits averaged across stores is not larger than 1% by both brand and type of model. The number of upper boundary sales hits never exceeds 5 times in a store across brands and models, corresponding to a maximum share of 6%. The detailed results on boundary effects by brand for all models can be obtained from the authors upon request.

5 Conclusions and limitations

In this paper, we proposed a Hierarchical Bayesian semiparametric store sales model with price dynamics for optimal pricing of fast-moving consumer goods offered in physical stores of a retailer and maximizing related expected brand profits. Using Bayesian P-splines as a nonparametric technique to estimate price effects allowed us to dispense with the specification of a specific functional form for own- and cross-price response a priori and to identify possible “irregular” pricing effects (e.g., kinks and steps in price response) that are difficult to capture with parametric functions. Heterogeneity of price effects across stores was accommodated via scaling factors for the P-splines, which serve as random effect parameters to scale the price functions up- or downwards for individual stores while preserving their overall shape. Accommodating heterogeneity has become state-of-the-art if not a must in econometric marketing models, far beyond the present context of store-level sales response models. Price dynamics were accounted for via one-week lagged own prices, which allowed us to implicitly address stockpiling or customer-holdover effects, although such *dynamic* effects are more difficult to explore based on aggregate sales data. For this reason, we further provided robustness checks for the *dynamic* variants of the *flexible* nonlinear models by including more time lags for the own-item price, indicating that higher order autoregressive structures were (mostly) not supported by the data. Optimal price paths for brands were determined by a discrete dynamic programming algorithm. We further imposed additional constraints for the price optimization step on price ranges and upper bounds for brand sales to preserve a realistic scenario for profit implications. To the best of our knowledge, addressing the three dimensions *functional flexibility*, *store heterogeneity*, and *price dynamics* simultaneously in one approach (both for sales response modeling and subsequent price optimization) has not been proposed previously. Therefore, this approach allows us for the first time to disentangle the effects of these three dimensions on sales, pricing and profits. In addition, we introduced the Continuous Ranked Probability Score as a new measure to assess the predictive model performance, a measure that has not been used before in the sales response modeling field.

We applied our proposed approach in an empirical study to data for refrigerated orange juice brands offered in physical stores of a large retail chain and compared it to several benchmark models (which ignore *heterogeneity*, *functional flexibility*, and/or *price dynamics*) in terms of forecasting accuracy, optimal pricing or price paths, optimized profits, and expected losses. Based on three different predictive performance measures, we found that accommodating both functional flexibility in price response and *price dynamics* provided the best sales predictions. Moreover, sales predictions from all models were almost fairly well dimensioned with regard to observed brand sales, i.e., the constraints imposed on the price optimization step worked well and protected our models from over- or underestimation biases in sales predictions and as a result in expected profit calculations, too. In addition, optimized expected profits were highest under the proposed *flexible*, *heterogeneous*, and *dynamic* sales response model for five brands and by either again a *flexible* or a

dynamic model for the other three brands. Importantly, optimized expected profits were highest for six brands based on models including *price dynamics*, suggesting that it is very important for managers to accommodate *price dynamics* for pricing strategies as well, at least for our data at hand. Not least, the best models by brand were all heterogeneous ones, and the proposed *flexible, heterogeneous, and dynamic* sales response model provided the highest expected profits on average across brands. As our first research question was whether a sales response model that combines *nonlinear pricing, store heterogeneity* in price (and other marketing) effects, and *price dynamics* can provide higher expected chain profits, the answer to this question is yes. The benefits from accommodating *price dynamics* for optimal pricing decisions were also clearly visible in our analyses on expected losses. Furthermore, and this is a very important point, we could show that upper boundary pricing effects were usually moderate across brands and never critical. In addition, the store-specific shares of price hits at the upper bound turned out highly heterogeneous for some brands. And lower boundary price hits as well as upper boundary sales hits occurred only in very few cases. As such, our second research question can also be positively evaluated in the sense that boundary pricing effects are not a critical issue for the proposed model, at least not for the data at hand. This finding together with the other ones summarized above let us recommend the use of the new model for price optimization.

Some limitations of our study should also be addressed. First, we took the perspective of a brand manager, who should be interested in analyzing the consequences of suboptimal pricing strategies from using a “wrong” sales response model for her/his brand, assuming fixed pricing patterns for substitute brands and the possibility to adjust the price of the own brand in the retailer’s assortment. This brand perspective is also an option for product category managers of retailers to analyze how price changes for one brand in their assortment would affect category sales and profits. Nevertheless, a next step could then be an extension of our model for optimal category pricing, using a seemingly unrelated regression (SUR) approach for model estimation as proposed, e.g., by Weber et al. (2017). Second, other options for accommodating *price dynamics* could be considered, for example by allowing for time-varying price effects instead of using simple lagged prices (leaning on reparametrization approaches such as proposed by Foekens et al. 1999 or Kopalle et al. 1999). Also, lead price effects could be added to account for anticipatory responses of consumers as a result of price promotion announcements (following, e.g., Van Heerde et al. 2000, 2004). Third, our approach could be further extended to allow for synergy effects between marketing instruments. For example, Ataman et al. (2010) proposed a dynamic model that relates brand sales to the long-term marketing strategy represented by advertising, product, and place in addition to price. Finally, we did not treat the issue of price endogeneity, which, if ignored, could lead to biased effect estimates. A good instrument to account for endogeneity in sales response models is the lagged price. However lagged prices are not exogenous in our *dynamic* model(s) and therefore could not be taken as instrument there. Wholesale prices, in our case

used as costs in our profit functions, represent another candidate for an instrumental variable. As objected by Rossi (2014), however, retail prices often show a much higher variation than wholesale prices. This also holds for many stores in our data, as was visible from Fig. 2 (top row) for the brand “Minute Maid” as example. Furthermore, two-step estimation as best known instrumental variable technique is restricted to linear models and cannot be applied to our nonlinear models (see, e.g., Hruschka 2017). Moreover, Ebbes et al. (2011) showed that models which account for endogeneity via instrumental variables do not necessarily provide better predictions. In the absence of adequate instruments one could use instrument-free techniques to accommodate endogeneity. Hruschka (2017) has recently provided an overview of the various options here, emphasizing, however, that currently only the copula-based method by Park and Gupta (2012) could be extended to nonlinear models like the ones proposed in this article (cf. Hruschka 2017, p. 28). To the best of our knowledge, such an extension of the copula-based method has not yet been developed for the flexible scaling models we consider. We therefore leave the issue of endogeneity as well as the other limitations mentioned above, for future research.

Appendix

Additional figures

See Figs. 9, 10 and 11.

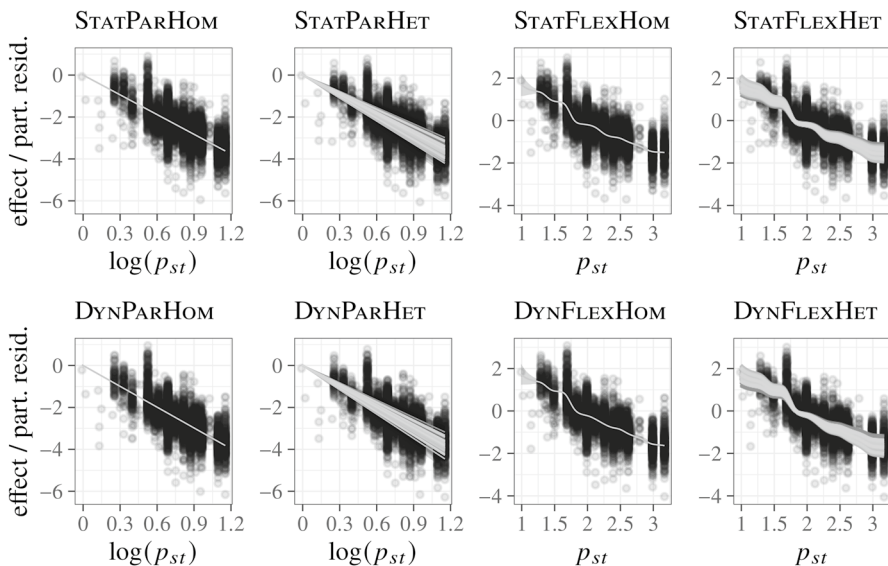


Fig. 9 Comparison of estimated own-price effects between *static* and *dynamic* model variants for the brand “Minute Maid”

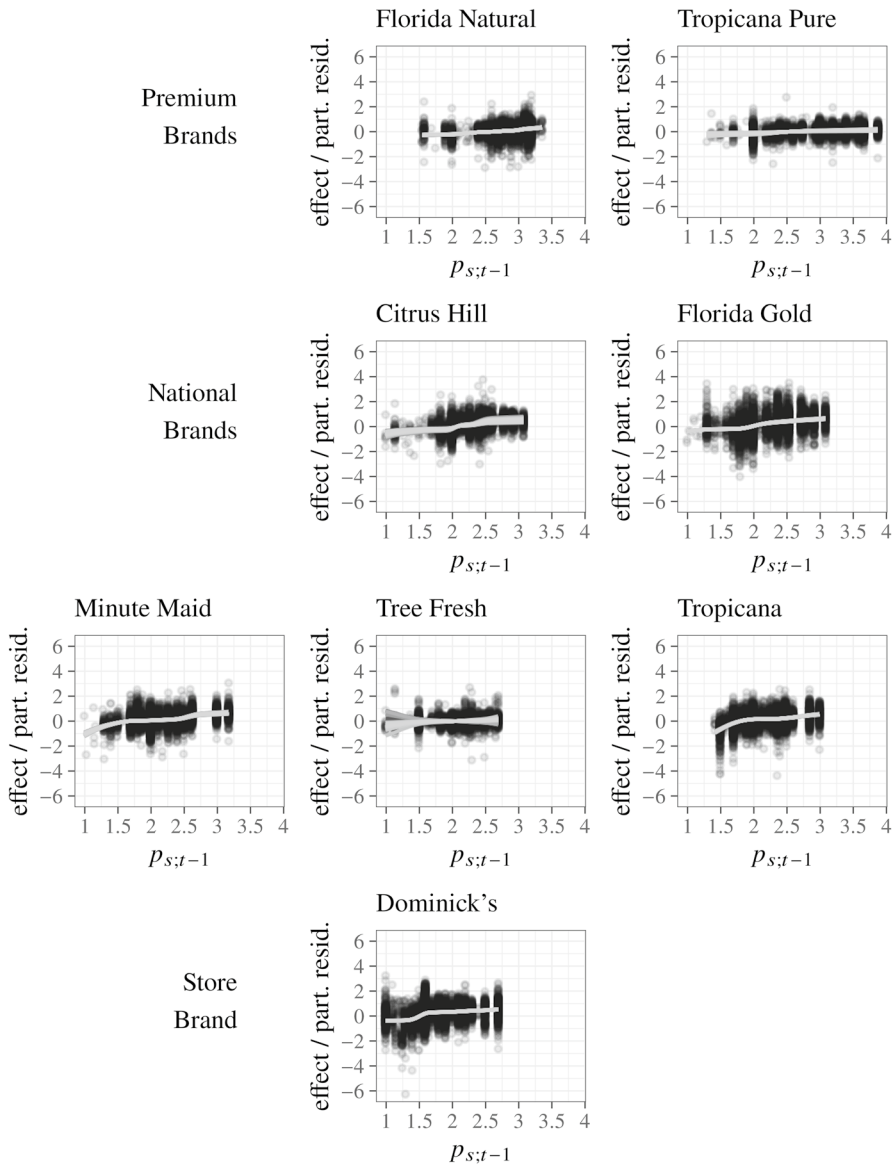


Fig. 10 Estimated lagged price effects resulting from the DYNFLEXHET model for all eight brands in the refrigerated orange juice category

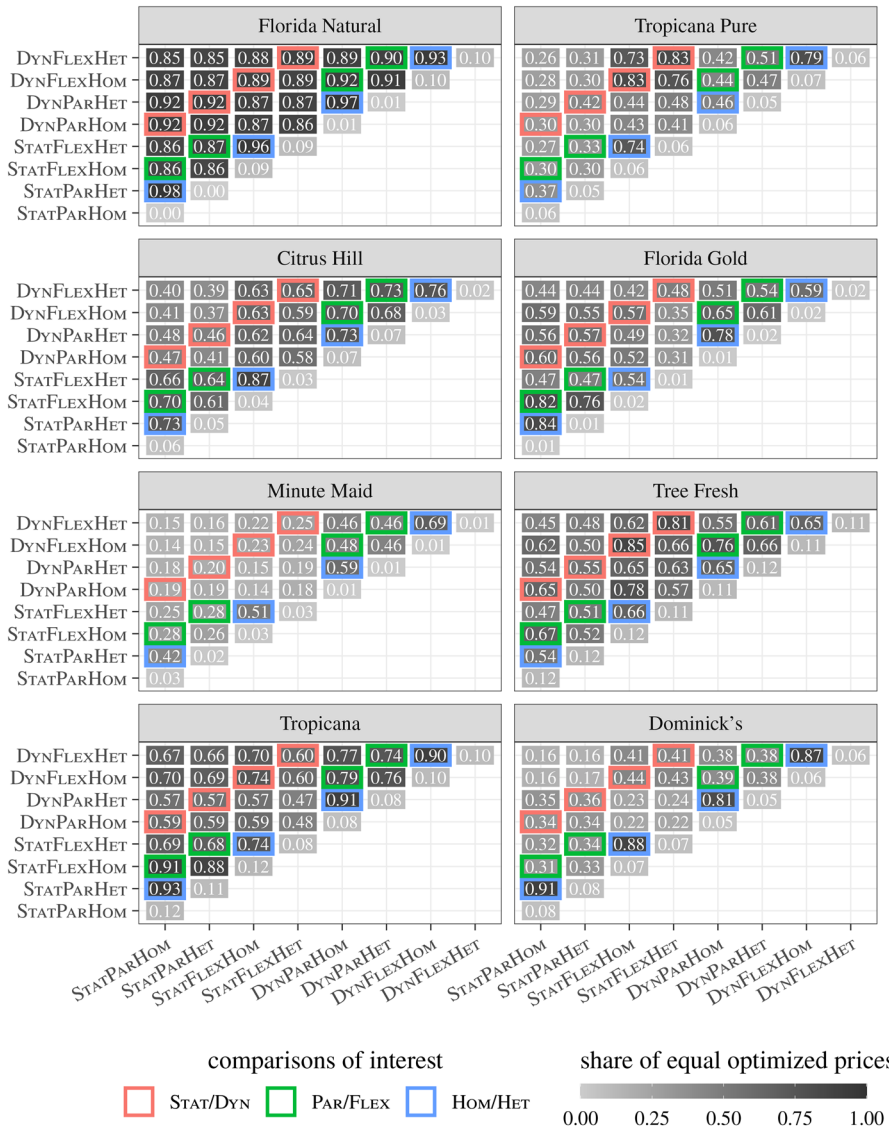


Fig. 11 For each brand, the numbers in the upper triangular part of the panels refer to the relative frequencies with which each two different models share the same optimal price levels in a particular store and week. Additionally, the diagonal elements indicate the relative frequencies where the optimized price and the observed price per store and week coincide for a particular model. Highlighted by colored borders are the pairwise comparisons between models that are more similar to each other, i.e., that differ in only one of the three model dimensions (*functional flexibility, store heterogeneity, price dynamics*). For example, the comparison between the DYNFLEXHOM model and the STATFLEXHOM model is framed in red as both models are *flexible* and *homogeneous* and just differ in the *dynamic* component

Additional tables

See Tables 6, 7, 8, 9 and 10.

Table 6 Descriptive statistics for weekly prices, market shares, and unit sales per store

Brand	Retail price (\$) / Price changes (%)*				M. share (%)		Unit sales	
	Range	Mean	SD	$p_t \neq p_{t-1}$	Mean	SD	Mean	SD
<i>Premium brands</i>								
Florida Natural	[1.54, 3.35]	2.85	0.33	38.6	4.7	6.5	27.2	46.2
Tropicana Pure	[1.29, 3.87]	2.96	0.57	46.7	12.3	13.6	74.8	98.2
<i>National brands</i>								
Citrus Hill	[0.99, 3.07]	2.31	0.35	42.8	8.0	12.7	53.5	157.4
Florida Gold	[0.99, 3.08]	2.18	0.39	43.3	5.2	7.9	33.4	64.0
Minute Maid	[0.99, 3.17]	2.22	0.42	55.2	10.2	13.8	52.0	76.8
Tree Fresh	[0.99, 2.69]	2.15	0.31	42.3	7.7	8.6	48.8	92.5
Tropicana	[1.41, 2.99]	2.20	0.38	56.1	18.4	21.1	112.8	159.0
<i>Private brand</i>								
Dominick's	[0.99, 2.69]	1.76	0.42	47.3	34.0	25.5	308.4	538.8

*The share of price changes represents the share of weeks with a different store-specific price compared to the previous week ($p_t \neq p_{t-1}$)

Table 7 Out-of-sample predictive performance of the DYNFLEXHOM and DYNFLEXHET models for different numbers of knots and different degrees of the B-spline basis functions, evaluated by the Average Root Mean Squared Error (ARMSE) in holdout samples (relative improvements/deteriorations in ARMSE values refer to the default specification of the P-splines with 20 knots and degree 3)

Brand	DYNFLEXHOM				DYNFLEXHET			
	20/1	20/3	40/1	40/3	20/1	20/3	40/1	40/3
Flor. Natrl.	24.9 (-0.2%)	25.0 -	24.9 (-0.3%)	25.0 (-0.1%)	27.0 (-0.7%)	27.2 -	27.0 (-0.4%)	26.9 (-0.8%)
Tropic. Pure	52.8 (-0.1%)	52.9 -	52.9 (+0.1%)	53.0 (+0.2%)	51.5 (-0.3%)	51.6 -	51.7 (+0.1%)	51.7 (+0.2%)
Citrus Hill	84.6 (-3.1%)	87.3 -	84.0 (-3.8%)	86.3 (-1.1%)	107.2 (-4.2%)	112.0 -	107.2 (-4.3%)	110.4 (-1.4%)
Flor. Gold	54.0 (-0.2%)	54.1 -	54.0 (-0.1%)	54.1 (±0.0%)	54.5 (-0.1%)	54.6 -	54.5 (-0.1%)	54.5 (-0.1%)
Min. Maid	53.8 (-0.5%)	54.1 -	53.7 (-0.7%)	54.1 (-0.1%)	53.7 (-0.2%)	53.8 -	53.5 (-0.5%)	54.2 (+0.8%)
Tree Fresh	65.8 (+0.1%)	65.7 -	65.8 (±0.0%)	65.8 (+0.1%)	76.5 (-2.4%)	78.4 -	76.9 (-1.9%)	78.5 (+0.1%)
Tropicana	91.5 (-0.3%)	91.8 -	91.7 (-0.1%)	91.8 (±0.0%)	92.7 (-0.3%)	93.0 -	93.0 (-0.1%)	93.1 (±0.0%)
Dominick's	282.9 (-0.3%)	283.6 -	283.5 (±0.0%)	284.5 (+0.3%)	285.1 (-0.4%)	286.2 -	286.3 (±0.0%)	286.9 (+0.3%)

Table 8 Out-of-sample predictive performance of the DYNFLEXHOM and DYNFLEXHET models for different numbers of knots and different degrees of the B-spline basis functions, evaluated by the Average Mean Continuous Ranked Probability Score (AMCRPS) in holdout samples (relative improvements/deteriorations in AMCRPS values refer to the default specification of the P-splines with 20 knots and degree 3)

Brand	DYNFLEXHOM			DYNFLEXHET				
	20/1	20/3	40/1	40/3	20/1	20/3	40/1	40/3
Flor. Natrl.	8.9 (-0.7%)	9.0	8.9 (-0.7%)	8.9 (-0.4%)	7.9 (-0.9%)	8.0	7.9 (-0.8%)	8.0 (-0.5%)
Tropic. Pure	22.3 (-0.1%)	22.4	22.3 (-0.2%)	22.3 (-0.2%)	20.1	20.1	20.1	20.1
Citrus Hill	21.5 (-1.4%)	21.8	21.5 (-1.4%)	21.7 (-0.5%)	18.8 (-1.7%)	19.1	18.9 (-1.2%)	19.0 (-0.5%)
Flor. Gold	18.9 (-0.2%)	19.0	18.9 (-0.1%)	18.9 (-0.1%)	18.3	18.4	18.3	18.3
Min. Maid	20.3 (-0.6%)	20.4	20.3 (-0.7%)	20.4 (-0.4%)	18.9 (-0.4%)	19.0	18.8 (-1.0%)	18.8 (-0.7%)
Tree Fresh	16.1 (+0.1%)	16.0	16.0 (-0.2%)	16.0 (±0.0%)	13.6 (-0.3%)	13.7	13.7 (+0.2%)	13.8 (+0.5%)
Tropicana	45.3 (-0.5%)	45.5	45.3 (-0.4%)	45.4 (-0.2%)	43.3 (-0.4%)	43.4	43.4 (±0.0%)	43.5 (+0.1%)
Dominick's	137.7 (-0.6%)	138.4	138.0 (-0.3%)	138.5 (±0.0%)	131.6 (-0.5%)	132.2	132.0 (-0.1%)	132.3 (+0.1%)

Table 9 Out-of-sample predictive performance of the DYNFLEXHOM and DYNFLEXHET models for different numbers of time-lags for the brand's own price, evaluated by the Average Root Mean Squared Error (ARMSE) in holdout samples (relative improvements/deteriorations refer to the default lag-1 specification)

Brand	DYNFLEXHOM			DYNFLEXHET		
	1	2	3	1	2	3
Flor. Natrl.	25.0	24.9	25.0	27.2	26.4	26.2
	–	(-0.4%)	(+0.1%)	–	(-2.9%)	(-3.4%)
Tropic. Pure	52.9	53.1	52.5	51.6	51.9	51.3
	–	(+0.5%)	(-0.8%)	–	(+0.6%)	(-0.6%)
Citrus Hill	87.3	85.6	85.3	112.0	105.0	105.7
	–	(-1.9%)	(-2.2%)	–	(-6.2%)	(-5.6%)
Flor. Gold	54.1	54.0	54.0	54.6	54.5	54.4
	–	(-0.1%)	(-0.1%)	–	(-0.1%)	(-0.2%)
Min. Maid	54.1	53.9	54.0	53.8	53.7	53.8
	–	(-0.4%)	(-0.2%)	–	(-0.3%)	(+0.1%)
Tree Fresh	65.7	65.7	65.8	78.4	80.5	83.9
	–	(±0.0%)	(+0.1%)	–	(+2.7%)	(+7.0%)
Tropicana	91.8	92.2	91.9	93.0	93.5	93.1
	–	(+0.4%)	(+0.1%)	–	(+0.5%)	(+0.1%)
Dominick's	283.6	284.2	285.2	286.2	287.8	288.8
	–	(+0.2%)	(+0.5%)	–	(+0.6%)	(+0.9%)

Table 10 Out-of-sample predictive performance of the DYNFLEXHOM and DYNFLEXHET models for different numbers of time lags for the brand's own price, evaluated by the Average Mean Continuous Ranked Probability Score (AMCRPS) in holdout samples (relative improvements/deteriorations refer to the default lag-1 specification)

Brand	DYNFLEXHOM			DYNFLEXHET		
	1	2	3	1	2	3
Flor. Natrl.	9.0	8.9	8.9	8.0	7.9	8.0
	–	(-0.7%)	(-0.7%)	–	(-1.0%)	(-0.7%)
Tropic. Pure	22.4	22.4	22.1	20.1	20.2	19.8
	–	(+0.2%)	(-1.1%)	–	(+0.1%)	(-1.7%)
Citrus Hill	21.8	21.7	21.8	19.1	18.9	18.9
	–	(-0.5%)	(-0.4%)	–	(-1.0%)	(-0.8%)
Flor. Gold	19.0	19.0	18.9	18.4	18.3	18.3
	–	(±0.0%)	(-0.4%)	–	(-0.2%)	(-0.4%)
Min. Maid	20.4	20.4	20.4	19.0	18.9	18.9
	–	(-0.2%)	(-0.3%)	–	(-0.2%)	(-0.7%)
Tree Fresh	16.0	16.0	16.0	13.7	13.7	13.7
	–	(-0.1%)	(-0.4%)	–	(+0.2%)	(-0.1%)
Tropicana	45.5	45.4	45.3	43.4	43.3	43.3
	–	(-0.2%)	(-0.3%)	–	(-0.2%)	(-0.4%)
Dominick's	138.4	133.6	133.6	132.2	127.3	126.9
	–	(-3.5%)	(-3.5%)	–	(-3.7%)	(-4.0%)

Acknowledgements All models were estimated with the public domain software package BayesX (Brezger et al. 2005). Further calculations and plots were conducted with the free and open source software R (R Core Team 2022) and the *tidyverse* collection of R packages (Wickham et al. 2019), amongst others. The authors would like to thank Mareike Karnebogen and Alexander Beckmann for their helpful comments and discussions about dynamic programming.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The data analyzed in this study are freely available at the James M. Kilts Center of the University of Chicago for academic purposes: <https://www.chicagobooth.edu/research/kilts/datasets/dominicks>.

Declarations

Conflict of interest The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

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

- Abe M (1991) A moving ellipsoid method for nonparametric regression and its application to logit diagnostics with scanner data. *J Mark Res* 28(3):339–346. <https://doi.org/10.1177/002224379102800308>
- Abe M (1995) A nonparametric density estimation method for brand choice using scanner data. *Mark Sci* 14(3):300–325. <https://doi.org/10.1287/mksc.14.3.300>
- Abe M, Boztuğ Y, Hildebrandt L (2004) Investigating the competitive assumption of multinomial logit models of brand choice by nonparametric modeling. *Comp Stat* 19(4):635–657. <https://doi.org/10.1007/BF02753916>
- Ali MM, Sharma SC (1993) Robustness to nonnormality of the durbin-watson test for autocorrelation. *J Econ* 57(1–3):117–136. [https://doi.org/10.1016/0304-4076\(93\)90061-9](https://doi.org/10.1016/0304-4076(93)90061-9)
- Andrews RL, Currim IS, Leeftang PSH, Lim J (2008) Estimating the SCAN*PRO model of store sales: HB, FM or just OLS? *Int J Res Mark* 25(1):22–33. <https://doi.org/10.1016/j.ijresmar.2007.10.001>
- Aschersleben P, Steiner WJ (2022) A semiparametric approach to estimating reference price effects in sales response models. *J Bus Econ* 92(4):591–643. <https://doi.org/10.1007/s11573-022-01083-y>
- Ataman MB, Van Heerde HJ, Mela CF (2010) The long-term effect of marketing strategy on brand sales. *J Mark Res* 47(5):866–882. <https://doi.org/10.1509/jmkr.47.5.866>
- Bhargava A, Franzini L, Narendranathan W (1982) Serial correlation and the fixed effects model. *Rev Econ Stud* 49(4):533. <https://doi.org/10.2307/2297285>
- Boztuğ Y, Hildebrandt L, Raman K (2014) Detecting price thresholds in choice models using a semi-parametric approach. *OR Spectr* 36(1):187–207. <https://doi.org/10.1007/s00291-012-0313-4>
- Brezger A, Steiner WJ (2008) Monotonic regression based on Bayesian P-splines: An application to estimating price response functions from store-level scanner data. *J Bus Econ Stat* 26(1):90–104. <https://doi.org/10.1198/073500107000000223>
- Brezger A, Kneib T, Lang S (2005) BayesX: Analyzing Bayesian structured additive regression models. *J Stat Softw* 14(11):1–22. <https://doi.org/10.18637/jss.v014.i11>, <http://bayesx.org>
- Brown GW (1951) Iterative solutions of games by fictitious play. In: Koopmans TC (ed) *Activity analysis of production and allocation*. Wiley, New York, pp 374–376
- Brown MB, Forsythe AB (1974) Robust tests for the equality of variances. *J Am Stat Assoc* 69(346):364–367. <https://doi.org/10.1080/01621459.1974.10482955>

- Chintagunta PK, Dubé JP, Singh V (2003) Balancing profitability and customer welfare in a supermarket chain. *Quant Mark Econ* 1(1):111–147. <https://doi.org/10.1023/A:1023534028314>
- Christen M, Gupta S, Porter JC, Staelin R, Wittink DR (1997) Using market-level data to understand promotion effects in a nonlinear model. *J Mark Res* 34(3):322–334. <https://doi.org/10.1177/002224379703400302>
- Ebbes P, Papies D, van Heerde HJ (2011) The sense and non-sense of holdout sample validation in the presence of endogeneity. *Mark Sci* 30(6):1115–1122. <https://doi.org/10.1287/mksc.1110.0666>
- Eilers PHC, Marx BD (1996) Flexible smoothing with *B*-splines and penalties. *Stat Sci* 11(2):89–102. <https://doi.org/10.1214/ss/1038425655>
- Fibich G, Gaviou A, Lowengart O (2003) Explicit solutions of optimization models and differential games with nonsmooth (asymmetric) reference-price effects. *Oper Res* 51(5):721–734. <https://doi.org/10.1287/opre.51.5.721.16758>
- Foekens EW, Leeflang PSH, Wittink DR (1999) Varying parameter models to accommodate dynamic promotion effects. *J Econ* 89(1–2):249–268. [https://doi.org/10.1016/S0304-4076\(98\)00063-3](https://doi.org/10.1016/S0304-4076(98)00063-3)
- Fok D, Horváth C, Paap R, Franses PH (2006) A hierarchical Bayes error correction model to explain dynamic effects of price changes. *J Mark Res* 43(3):443–461. <https://doi.org/10.1509/jmkr.43.3.443>
- Franses PH, Ghijsels H (1999) Additive outliers, GARCH and forecasting volatility. *Int J Forecast* 15(1):1–9. [https://doi.org/10.1016/S0169-2070\(98\)00053-3](https://doi.org/10.1016/S0169-2070(98)00053-3)
- Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc* 102(477):359–378. <https://doi.org/10.1198/016214506000001437>
- Greene WH (2008) *Econometric analysis*, 6th edn. Pearson Prentice Hall, Upper Saddle River, NJ
- Greenleaf EA (1995) The impact of reference price effects on the profitability of price promotions. *Mark Sci* 14(1):82–104. <https://doi.org/10.1287/mksc.14.1.82>
- Hanssens DM, Parsons LJ, Schultz RL (2002) *Market Response Models: Econometric and Time Series Analysis*, vol 12, 2nd Edn. Kluwer Academic Publishers, Boston
- Härdle W (1990) *Applied nonparametric regression*, vol 19. Cambridge University Press, Cambridge
- Haupt H, Kagerer K (2012) Beyond mean estimates of price and promotional effects in scanner-panel sales-response regression. *J Retail Consum Serv* 19(5):470–483. <https://doi.org/10.1016/j.jretconser.2012.06.002>
- Haupt H, Kagerer K, Steiner WJ (2014) Smooth quantile-based modeling of brand sales, price and promotional effects from retail scanner panels. *J Appl Econ* 29(6):1007–1028. <https://doi.org/10.1002/jae.2347>
- Horváth C, Fok D (2013) Moderating factors of immediate, gross, and net cross-brand effects of price promotions. *Mark Sci* 32(1):127–152. <https://doi.org/10.1287/mksc.1120.0748>
- Hruschka H (1997) Schätzung und normative Analyse ausgewählter Preis-Absatz-Funktionen. *Z Betriebswirtsch* 67(8):845–864
- Hruschka H (2000) Specification, estimation, and empirical corroboration of Gutenberg's kinked demand curve. In: Albach H, Brockhoff KKL, Eymann E, et al (Eds) *Theory of the Firm*. Springer, Berlin and Heidelberg, p 153–168. https://doi.org/10.1007/978-3-642-59661-2_8
- Hruschka H (2006a) Relevance of functional flexibility for heterogeneous sales response models: A comparison of parametric and semi-nonparametric models. *Eur J Oper Res* 174(2):1009–1020. <https://doi.org/10.1016/j.ejor.2005.05.003>
- Hruschka H (2006b) Statistical and managerial relevance of aggregation level and heterogeneity in sales response models. *Mark ZFP – J Res Manag* 28(12):94–102. <https://doi.org/10.15358/0344-1369-2006-JRM-2-94>
- Hruschka H (2007) Clusterwise pricing in stores of a retail chain. *OR Spectr* 29(4):579–595. <https://doi.org/10.1007/s00291-006-0075-y>
- Hruschka H (2017) Functional flexibility, latent heterogeneity and endogeneity in aggregate market response models. *Mark ZFP – J Res Manag* 39(3):17–31. <https://doi.org/10.15358/0344-1369-2017-3-17>
- Jordan A (2016) *Facets of forecast evaluation*. PhD dissertation, Karlsruher Institut für Technologie, Karlsruhe. <https://doi.org/10.5445/IR/1000063629>
- Jordan A, Krüger F, Lerch S (2019) Evaluating probabilistic forecasts with scoringrules. *J Stat Softw*. <https://doi.org/10.18637/jss.v090.i12>
- Kalyanam K, Shively TS (1998) Estimating irregular pricing effects: A stochastic spline regression approach. *J Mark Res* 35(1):16–29. <https://doi.org/10.1177/002224379803500104>
- Kopalle PK, Mela CF, Marsh L (1999) The dynamic effect of discounting on sales: empirical analysis and normative pricing implications. *Mark Sci* 18(3):317–332. <https://doi.org/10.1287/mksc.18.3.317>

- Krishnamurthi L, Raj SP, Selvam R (1990) Statistical and managerial issues in cross-sectional aggregation, working paper, Northwestern University
- Krüger F, Lerch S, Thorarinsdottir T, Gneiting T (2021) Predictive inference based on Markov chain Monte Carlo output. *Int Stat Rev* 89(2):274–301. <https://doi.org/10.1111/insr.12405>
- Lang S, Brezger A (2004) Bayesian P-splines. *J Comput Graph Stat* 13(1):183–212. <https://doi.org/10.1198/1061860043010>
- Lang S, Steiner WJ, Weber A, Wechselberger P (2015) Accommodating heterogeneity and nonlinearity in price effects for predicting brand sales and profits. *Eur J Oper Res* 246(1):232–241. <https://doi.org/10.1016/j.ejor.2015.02.047>
- Leefflang PSH, Wittink DR, Wedel M, Naert PA (2000) Building models for marketing decisions, vol 9. Springer, Boston, MA
- Lucas RE (1976) Econometric policy evaluation: a critique. *Carnegie-Rochester Conf Ser Public Policy* 1:19–46. [https://doi.org/10.1016/S0167-2231\(76\)80003-6](https://doi.org/10.1016/S0167-2231(76)80003-6)
- Matheson JE, Winkler RL (1976) Scoring rules for continuous probability distributions. *Manag Sci* 22(10):1087–1096. <https://doi.org/10.1287/mnsc.22.10.1087>
- Miranda MJ, Fackler PL (2002) Applied computational economics and finance. MIT Press, Cambridge, MA
- Montgomery AL (1997) Creating micro-marketing pricing strategies using supermarket scanner data. *Mark Sci* 16(4):315–337. <https://doi.org/10.1287/mksc.16.4.315>
- Nijs VR, Dekimpe MG, Steenkamps JBEM, Hanssens DM (2001) The category-demand effects of price promotions. *Mark Sci* 20(1):1–22. <https://doi.org/10.1287/mksc.20.1.1.10197>
- Park S, Gupta S (2012) Handling endogenous regressors by joint estimation using copulas. *Mark Sci* 31(4):567–586. <https://doi.org/10.1287/mksc.1120.0718>
- R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>
- Rossi PE (2014) Bayesian non- and semi-parametric methods and applications. Princeton University Press, Princeton
- Silva-Risso JM, Bucklin RE, Morrison DG (1999) A decision support system for planning manufacturers' sales promotion calendars. *Mark Sci* 18(3):274–300. <https://doi.org/10.1287/mksc.18.3.274>
- Steiner WJ, Brezger A, Belitz C (2007) Flexible estimation of price response functions using retail scanner data. *J Retail Consum Serv* 14(6):383–393. <https://doi.org/10.1016/j.jretconser.2007.02.008>
- Van Heerde HJ (1999) Models for sales promotion effects based on store-level scanner data. Labyrinth Publication, Capelle a/d IJssel
- Van Heerde HJ, Leefflang PSH, Wittink DR (2000) The estimation of pre- and postpromotion dips with store-level scanner data. *J Mark Res* 37(3):383–395. <https://doi.org/10.1509/jmkr.37.3.383.18782>
- Van Heerde HJ, Leefflang PSH, Wittink DR (2001) Semiparametric analysis to estimate the deal effect curve. *J Mark Res* 38(2):197–215. <https://doi.org/10.1509/jmkr.38.2.197.18842>
- Van Heerde HJ, Leefflang PSH, Wittink DR (2002) How promotions work: SCAN*PRO-based evolutionary model building. *Schmalenbach Bus Rev* 54(3):198–220. <https://doi.org/10.1007/BF03396653>
- Van Heerde HJ, Leefflang PSH, Wittink DR (2004) Decomposing the sales promotion bump with store data. *Mark Sci* 23(3):317–334. <https://doi.org/10.1287/mksc.1040.0061>
- Weber A, Steiner WJ (2012) Zur Berücksichtigung von Heterogenität versus funktionaler Flexibilität in Absatzreaktionsmodellen: Eine empirische Studie auf Basis von Handelsdaten. *Z Betriebswirtsch* 82(12):1337–1365. <https://doi.org/10.1007/s11573-012-0638-0>
- Weber A, Steiner WJ (2021) Modeling price response from retail sales: An empirical comparison of models with different representations of heterogeneity. *Eur J Oper Res* 294(3):843–859. <https://doi.org/10.1016/j.ejor.2020.07.055>
- Weber A, Steiner WJ, Lang S (2017) A comparison of semiparametric and heterogeneous store sales models for optimal category pricing. *OR Spectr* 39(2):403–445. <https://doi.org/10.1007/s00291-016-0459-6>
- Wedel M, Zhang J (2004) Analyzing brand competition across subcategories. *J Mark Res* 41(4):448–456. <https://doi.org/10.1509/jmkr.41.4.448.47017>
- Wickham H, Averick M, Bryan J et al (2019) Welcome to the tidyverse. *J Open Sourc Softw* 4(43):1686. <https://doi.org/10.21105/joss.01686>