Accommodating heterogeneity and nonlinearity in price effects for predicting brand sales and profits

Stefan Lang\textsuperscript{a,1}, Winfried J. Steiner\textsuperscript{b,\ast}, Anett Weber\textsuperscript{b,2}, Peter Wechselberger\textsuperscript{c,3}

\textsuperscript{a}Department of Statistics, Faculty of Economics and Statistics, University of Innsbruck, Universitätsstrasse 15, A-6020 Innsbruck, Austria
\textsuperscript{b}Department of Marketing, Clausthal University of Technology, Julius-Albert-Strasse 2, 38678 Clausthal-Zellerfeld, Germany
\textsuperscript{c}RSU Rating Service Unit GmbH & Co. KG, Karlstraße 35, 80333 München, Germany

\textbf{Abstract}

We propose a hierarchical Bayesian semiparametric approach to account simultaneously for heterogeneity and functional flexibility in store sales models. To estimate own- and cross-price response flexibly, a Bayesian version of P-splines is used. Heterogeneity across stores is accommodated by embedding the semiparametric model into a hierarchical Bayesian framework that yields store-specific own- and cross-price response curves. More specifically, we propose multiplicative store-specific random effects that scale the nonlinear price curves while their overall shape is preserved. Estimation is fully Bayesian and based on novel MCMC techniques. In an empirical study, we demonstrate a higher predictive performance of our new flexible heterogeneous model over competing models that capture heterogeneity or functional flexibility only (or neither of them) for nearly all brands analyzed. In particular, allowing for heterogeneity in addition to functional flexibility can improve the predictive performance of a store sales model considerably, while incorporating heterogeneity alone only moderately improved or even decreased predictive validity. Taking into account model uncertainty, we show that the proposed model leads to higher expected profits as well as to materially different pricing recommendations.

© 2015 Elsevier B.V. and Association of European Operational Research Societies (EURO) within the International Federation of Operational Research Societies (IFORS). All rights reserved.

\section{1. Motivation and literature review}

In recent years, two streams of research for estimating sales response models based on store-level data have evolved: on the one hand, researchers have proposed hierarchical Bayesian (HB) store sales models allowing for heterogeneity of marketing effects across stores (e.g., Andrews, Currim, Leeflang, & Lim, 2008; Blattberg & George, 1991; Boatwright, McCulloch, & Rossi, 1999; Hruschka, 2006b; Montgomery, 1997; Montgomery & Rossi, 1999). While some of these studies have shown that considering heterogeneity can improve model fit, the accuracy of sales forecasts, or expected profits (e.g., Hruschka, 2006b; Montgomery, 1997), recent research of Andrews et al. (2008) has demonstrated rather marginal improvements in fit and predictive performance from incorporating store heterogeneity. One possible reason for this latter finding is that the HB models mentioned above assume a strictly parametric functional form thereby limiting the scope for model calibration to an a priori fixed parametrization. Hence, although accounting for heterogeneity, a source of bias remains if the assumed parametric form differs from the true underlying function.

On the other hand, researchers have proposed nonparametric regression models in order to accommodate potential nonlinearities in store sales response (e.g., Brezger & Steiner, 2008; Haupt, Kagerer, & Steiner, 2014; van Heerde, Leeflang, & Wittink, 2001; Kalyanam & Shively, 1998; Steiner, Brezger, & Belitz, 2007). The empirical results of this second stream indicate that own- and cross-price effects may show complex nonlinearities which are difficult or not at all to capture by parametric models. The main potential weakness of this second group of nonparametric approaches, however, is that heterogeneity across stores has not been considered. Consequently, bias due to potential heterogeneity across stores here remains.

There is so far only one approach that has consolidated the two streams: Hruschka (2006a, 2007) proposed a hierarchical Bayesian multilayer perceptron (MLP) that allows for nonlinearity in price effects and yields store-specific coefficients. In an empirical study, his flexible heterogenous MLP turned out to be superior in terms of
### Table 1
Descriptive statistics for weekly brand prices, market shares, and unit sales.

<table>
<thead>
<tr>
<th>Brand</th>
<th>Retail price</th>
<th>Market share</th>
<th>Unit sales</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range ($)</td>
<td>Mean ($)</td>
<td>SD ($)</td>
</tr>
<tr>
<td><strong>Premium brands</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tropicana Pure</td>
<td>[1.60; 3.55]</td>
<td>2.95</td>
<td>.53</td>
</tr>
<tr>
<td>Florida Natural</td>
<td>[1.57; 3.16]</td>
<td>2.86</td>
<td>.33</td>
</tr>
<tr>
<td><strong>National brands</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citrus Hill</td>
<td>[1.09; 2.82]</td>
<td>2.31</td>
<td>.31</td>
</tr>
<tr>
<td>Minute Maid</td>
<td>[1.29; 2.92]</td>
<td>2.23</td>
<td>.40</td>
</tr>
<tr>
<td>Tropicana</td>
<td>[1.49; 2.75]</td>
<td>2.20</td>
<td>.35</td>
</tr>
<tr>
<td>Florida Gold</td>
<td>[.99; 2.83]</td>
<td>2.17</td>
<td>.39</td>
</tr>
<tr>
<td>Tree Fresh</td>
<td>[1.07; 2.48]</td>
<td>2.15</td>
<td>.27</td>
</tr>
<tr>
<td><strong>Store brand</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominick’s</td>
<td>[.99; 2.47]</td>
<td>1.75</td>
<td>.4</td>
</tr>
</tbody>
</table>

a The unit sales of all eight brands amount to 96.25 percent of the total sales volume in the refrigerated orange juice category (64 oz) during the time span considered.
b Reading example: For Tropicana Pure, the lowest observed price across all stores and weeks was 1.60 $, its lowest market share (unit sales) in a week pooled across stores was 3 percent (6388 units), and its mean price level averaged across all stores and weeks was 2.95 $.

The advantages of using nonparametric or seminonparametric techniques for estimating response functions were demonstrated previously by Hruschka in the context of market share modeling (Hruschka, 2002), brand choice modeling (Hruschka, Fettes, & Probst, 2004) and catalog allocation modeling (Baumgartner & Hruschka, 2005); too.

To model store sales response, we use weekly store-level scanner data for eight brands of orange juice offered by Dominick’s Finer Foods (DFF), a major supermarket chain in the Chicago metropolitan area. The data were provided by the James M. Kilts Center, GSB, University of Chicago, and include unit sales (qts), price (priceₘᵣₜ), and a deal code indicating the use of a display (displayₘᵣₜ) for each of the eight brands in each s = 1,...,81 stores of the chain over a time horizon of t = 1,...,89 weeks (resulting in about 7,000 data points per brand). The price data further reveal whether a 9- or 99-ending price (end₉ₚ, end₉₉ₚ) has been set. The brands can be grouped into three price-quality tiers: the premium brands which are made from freshly squeezed oranges, the national brands which are reconstituted from frozen orange juice concentrate, and the retailer’s own store brand.

Table 1 provides summary statistics pooled across stores for weekly prices, market shares and unit sales of the individual brands.

Since cross-item price effects are usually much weaker than own-item price effects (e.g., see Hanssens, Parsons, & Schultz, 2003), we capture them in our demand equation in a more parsimonious way at the tier level. Following Brezger and Steiner (2008), we define priceₙᵃᵗⁱᵒⁿᵃˡₙᵣₜ (priceₚʳᵉ․ₐᵣᵣₜ) as the lowest price of a competing national (premium) brand in store s and week t. For example, if a store sales model is estimated for the national brand Citrus Hill, priceₙᵃᵗⁱᵒⁿᵃˡₙᵣₜ captures the lowest price level of either of the other national brands in store s and week t.₄ price₉ₚ₉₉ₙᵣₜ denotes the observed price for Dominick’s, the only private label brand in our data.

We further use 11 characteristics (collected in the vector νₛ) of each store’s trading area to explain possible store-level variation in price response due to sociodemographic and competitive effects. A detailed description of these background covariates (among others relating to age, education, family size, income, distances to nearest

---

₄ Note that for the computation of priceₙᵃᵗⁱᵒⁿᵃˡₙᵣₜ, we consider Sunny Delight as another national brand in the refrigerated orange juice category. However, we did not estimate store sales models for Sunny Delight due to the lack of (substantial) price variation of this brand.
دریافت فوری
متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات