European Journal of Operational Research 277 (2019) 684–698

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Addressing endogeneity in aggregate logit models with time-varying parameters for optimal retail-pricing

Daniel Guhl

Institute of Marketing, School of Business and Economics, Humboldt University Berlin, Spandauer Str. 1, Berlin 10178, Germany

ARTICLE INFO

Article history:
Received 23 April 2018
Accepted 26 February 2019
Available online 6 March 2019

Keywords:
Retailing
Discrete choice
Endogeneity
Time-varying parameters
Heterogeneity

ABSTRACT

It is well known that price endogeneity is a severe problem in demand models for market-level data (e.g., aggregate logit models) because it leads to biased estimates and therefore incorrect managerial implications. If the price parameter varies over time, as is usually the case, the relevance of the issue increases because standard methods to correct endogeneity biases (e.g., generalized method of moments) fail. This paper presents a control function approach as a remedy. A comprehensive simulation study demonstrates this method's suitability, such that addressing endogeneity with the control function approach is the best choice. Moreover, addressing the endogeneity problem incorrectly may be even more harmful than simply ignoring it. To further illustrate the control function approach, we analyze the demand for canned tuna using aggregate retailer-level data. Here, all utility parameters vary over time and price endogeneity is indeed an issue. Effectively addressing price endogeneity correct has positive economic consequences: a normative model analysis reveals that implementing the control function approach yields a 3 % increase in retailer profits.

1. Introduction

Logit models for aggregate data are increasingly popular in management science, operations research and marketing (e.g., Besanko, Gupta, & Jain, 1998; Casado & Ferrer, 2013; Dubé et al., 2002; Hruschka, 2017; Kök & Fisher, 2007; Park & Gupta, 2009). Today, they are among the most important empirical tools to model demand and investigate markets with differentiated products (Ackerberg, Benkard, Berry, & Pakes, 2007; Morrow & Skerlos, 2011; Reiss & Wolak, 2007). This popularity has (at least) three causes: (1) logit models are grounded in the economic theory of utility maximization. This theoretical foundation allows for estimating structural parameters that are economically meaningful (Dubé et al., 2002) and delineating complex competition patterns parsimoniously (Ackerberg et al., 2007). (2) The models can handle both endogeneity of individual variables and consumer heterogeneity. Many studies show that endogeneity (especially price endogeneity) is a severe problem in nonexperimental data. Endogeneity arises when explanatory variables are correlated with the error term of the model (Greene, 2011). As endogeneity might cause biased estimates and lead to false implications (Chintagunta, 2001; Chintagunta, Dubé, & Goh, 2005; Hruschka, 2010; Villas-Boas & Winer, 1999), it is essential that many research studies effectively address this problem (Berry, 1994). Additionally, it is essential to take consumer heterogeneity in choice models into account (Kalouptsidis & Psaraki, 2010); understanding differences in the preferences and needs of individual consumers is the cornerstone of marketing (Allenby & Rossi, 1999), and the use of homogeneous models can also lead to biased results (Chintagunta, Jain, & Vilcassim, 1991). Therefore, modeling heterogeneity is the defacto standard in modern empirical marketing research and practice. (3) Compared to disaggregate models, the data requirements are “lower” and using aggregate (e.g., retailer-level) data has a significant advantage, because aggregate data are more accessible and manageable for retailers (Bodapati & Gupta, 2004), and they are less criticized for questionable representativeness (Gupta, Chintagunta, Kaul, & Wittink, 1996).

Despite their advantageous features, logit models for aggregate data usually lack dynamic flexibility and assume that consumer preferences do not change over time (Van Heerde & Neslin, 2008). There are many reasons to expect the opposite, however. Examples include seasonality and market trends, product innovations, changes in psychographic variables (e.g., brand awareness, attitudes, etc.), as well as loyalty and variety seeking of consumers (Leeflang et al., 2009). Given available datasets typically span many periods (sometimes years), the assumption of entirely “stable” utility structures seems questionable. For example, Park and Gupta (2012a) study quarterly data from 1981 to 1998, Jiang, Manchanda, and Rossi (2009) analyze

E-mail address: daniel.guhl@hu-berlin.de

https://doi.org/10.1016/j.ejor.2019.02.058
0377-2217/

data for almost 400 weeks, and (Chintagunta, 2002) use 196 weeks.

Discrete choice models for disaggregate data frequently consider dynamic effects via purchase-event feedback or reference price variables. Purchase-event feedback variables (Guadagni & Little, 1983; Keane, 1997) and reference prices (Casado & Ferrer, 2013; Winer, 1986) are based on consumers' individual purchase (i.e., choice and price) histories. However, such variables cannot be used if only aggregate data are available because it does not contain individual information. Sometimes, through reparameterization, parameters of the choice model are modeled as functions of time-variable covariates (Jedidi, Mela, & Gupta, 1999). This popular approach is easy to apply to aggregate data (Foekens, Leeflang, & Wittink, 1999; Kopalle, Mela, & Marsh, 1999). However, as Van Heerde, Mela, and Manchanda (2004, p. 167) emphasize, this technique only models the temporal variation of the dynamic variables (as a conditional expectation) and not their temporal evolution. Finally, there is the possibility to split the data and, e.g., perform before-and-after analyses (Sriram & Kadiyali, 2009). Note that absolute parameter values in logit models are not comparable across datasets because of scale differences (Mela, Gupta, & Lehmann, 1997). This limits the applicability of the procedure to comparable measures, e.g., marginal effects or elasticities, and data splitting also reduces the statistical efficiency.

If applicable, these approaches are restrictive, allow for dynamics only in a restricted form, and require a priori additional information and assumptions. Therefore, we propose an alternative approach using state space models (SSM) and the Kalman filter (KF) (Kalman, 1960; Pennings & van Dalen, 2017), which enable flexible modeling and estimation of dynamics in choice models. The parameters of the utility functions vary stochastically over time and are specified as random walk processes (Lachaab, Ansari, Jedidi, & Trabelsi, 2006). In current papers dealing with "marketing dynamics," this approach becomes increasingly popular due to its many desirable features, e.g., Van Heerde et al. (2004) model the dynamic effect of innovation on market structure and apart from flexible parameter dynamics (that nest random walks as special cases) also consider store and brand heterogeneity. However, they do not account for price endogeneity, and as it is common in many marketing applications of SSM and the KF (see Leeflang et al., 2009 for an overview), do not apply a nonlinear logit model (i.e., structural model).

Sriram, Chintagunta, and Neelammegham (2006) and Sriram and Kalwani (2007) were the first to use SSM for handling dynamics in logit models for aggregate data. In contrast, we assume not only consumers' brand preferences but also consumers' marketing-mix sensitivities to vary over time, which is crucial in many situations (e.g., Lachaab et al., 2006). Marketing-mix sensitivities change, e.g., due to seasonality (Chevalier, Kashyap, & Rossi, 2003; Meza & Sudhir, 2006). Additionally, frequent price-promotion activities can further influence consumers' price sensitivity Jedidi et al. (1999) and Erdem, Keane, and Sun (2008) showed that TV advertising also impacts price sensitivity. However, if endogenous variables are time-varying, eliminating endogeneity becomes a methodological challenge. In several publications, Kim demonstrated that standard procedures are not suitable when endogenous variables have a time-varying effect (Kim, 2006; 2010; Kim & Kim, 2011). Here, the approach of Sriram et al. (2006) and Sriram and Kalwani (2007) based on the generalized method of moments (GMM) is not applicable. Kim proposes a control function (CF) approach, which still uses the endogenous variable in its original form but adds additional variables (the CF) to the model to solve the endogeneity issue. Another advantage of the CF approach is that it enables direct testing for endogeneity. Therefore, we adopt Kim's CF approach and apply it to heterogeneous logit models for aggregate data.

This paper addresses three research questions. (1) Is it possible to estimate time-varying parameters in logit models for aggregate data using a likelihood-based approach while taking price endogeneity and consumer heterogeneity into account? (2) What are the consequences if endogeneity is handled incorrectly? What happens if endogeneity is ignored? (3) How does the specific modeling approach affect the optimal retail-pricing problem? We answer these questions using a simulation study and illustrate our findings in an empirical application. We aim to improve the understanding of logit models for aggregate data. In light of those models' popularity and relevance, it is important to know how they have to be specified when parameters are time-varying. In addition to this econometric motivation, answering the research questions is highly relevant for marketing research and practice, where dynamics are essential (Leeflang et al., 2009).

The remainder of this paper is organized as follows: Section 2 presents and discusses the model and the CF approach. The estimation of constant and time-varying parameters as well as an endogeneity-test are described in Section 3. In Section 4, a simulation study examines the performance of the CF approach, which is further illustrated in the empirical and normative study in Section 5. We conclude in Section 6 with a summary of the results and suggestions for future research.

2. Model

The model consists of two parts: the demand model and the price model. The demand model is based on the heterogeneous ("mixed") logit model (MXL) for aggregate data, which originates from Berry (1994) and Berry, Levinsohn, and Pakes (1995) and is discussed in detail by Nevo (2000). The original version of the demand model is extended by time-varying parameters. The price model serves the purpose of solving the (potential) endogeneity problem in the demand model via the CF approach ((Kim, 2010)).

2.1. Demand model

We start in Eq. (1) with the (indirect) utility function. Suppose there are $i = 1, \ldots, I$ consumers who choose from $j = 1, \ldots, J$ brands at the time (e.g., weeks) $t = 1, \ldots, T$ the brand providing the highest utility u_{ijt}. To incorporate category purchases, the "no-purchase"-option is modeled additionally with $j = 0$ (Chintagunta, 2001).

$$u_{ijt} = \begin{cases} \mathbf{x}'_{jt}\boldsymbol{\alpha}_{it} + \beta_{it}\ln(p_{jt}) + \varepsilon_{jt} + \nu_{ijt}, & j = 1, \ldots, J \\ \nu_{ijt}, & j = 0, \end{cases} \tag{1}$$

where $\mathbf{x}_{jt}$ is a $K \times 1$ vector with exogenous variables (e.g., promotion activities) and dummy variables for the brand-intercepts. The log-price[1] variable ($\ln(p_{jt})$) is potentially endogenous and is considered separately for didactic reasons. Thus, in sum, there are $K + 1$ variables in the utility function.

Furthermore, ε_{jt} and ν_{ijt} are two error terms representing different phenomena. $\varepsilon_{jt} \sim N(0, \sigma^2_{\varepsilon j})$ is normally distributed (with brand-specific variances) and captures demand shocks due to (by the researcher) "unmeasured brand characteristics" (UMC), since it seems very unlikely that the model contains all variables affecting demand and ignoring such UMC might bias the results (Chintagunta et al., 2005; Park & Gupta, 2009). Additionally, if the UMC are correlated with the price-variable (i.e., $\mathbb{E}[\varepsilon_{jt} \,|\, \ln(\mathrm{p}_{jt})] \neq 0$), they will cause an endogeneity problem and the price effect will be biased (Petrin & Train, 2010). The second error term, ν_{ijt}, is extreme value type-I distributed and represents consumer-, brand-,

[1] Log-prices improve the fit in the empirical study, but the model derivation and description also apply to a linear price effect.

and time-specific randomness (pertaining to the researcher rather than the decision-maker).

The parameters of the utility function (1) are consumer- and time-specific and are thus indexed with i and t. Heterogeneity is modeled as being continuous via the normal distribution. The normality assumption is quite convenient because it implies additive-separable dynamics (i.e., changes in the mean values) and heterogeneity (i.e., deviations of individual values from the mean value):

$$\boldsymbol{\alpha}_{it} = \bar{\boldsymbol{\alpha}}_t + \boldsymbol{\upsilon}_{\boldsymbol{\alpha}_i} \text{ and} \tag{2}$$

$$\beta_{it} = \bar{\beta}_t + \upsilon_{\beta_i}, \text{ where } \left[\boldsymbol{\upsilon}'_{\boldsymbol{\alpha}_i} \;\; \upsilon_{\beta_i}\right]' = \boldsymbol{\upsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}). \tag{3}$$

The mean values $\bar{\boldsymbol{\alpha}}_t$ and $\bar{\beta}_t$ follow a random walk process[2]:

$$\bar{\boldsymbol{\alpha}}_t = \bar{\boldsymbol{\alpha}}_{t-1} + \boldsymbol{\eta}_{\boldsymbol{\alpha}_t} \text{ and} \tag{4}$$

$$\bar{\beta}_t = \bar{\beta}_{t-1} + \eta_{\beta_t}, \text{ where } \left[\boldsymbol{\eta}'_{\boldsymbol{\alpha}_t} \;\; \eta_{\beta_t}\right]' = \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}). \tag{5}$$

In contrast to a constant parameter model, the time-dimension (unobservable dynamics) and the consumer-dimension (unobservable heterogeneity) both vary. Moreover, our particular specification is highly flexible, and it also allows for testing alternative specifications (e.g., with/without dynamics and/or with/without heterogeneity) within nested models.

To facilitate the understanding of the model and its estimation, it is convenient as well as customary in the literature to structure the utility function's components according to whether they are consumer-specific or not (Nevo, 2000). Substituting Eqs. (2) and (3) into the utility function leads to:

$$u_{ijt} = \mathbf{x}'_{jt}(\bar{\boldsymbol{\alpha}}_t + \boldsymbol{\upsilon}_{\boldsymbol{\alpha}_i}) + (\bar{\beta}_t + \upsilon_{\beta_i})\ln(p_{jt}) + \varepsilon_{jt} + \nu_{ijt}, \tag{6}$$

$$\Leftrightarrow u_{ijt} = \underbrace{\mathbf{x}'_{jt}\bar{\boldsymbol{\alpha}}_t + \bar{\beta}_t \ln(p_{jt}) + \varepsilon_{jt}}_{\delta_{jt}} + \underbrace{\mathbf{x}'_{jt}\boldsymbol{\upsilon}_{\boldsymbol{\alpha}i} + \upsilon_{\beta i}\ln(p_{jt})}_{\mu_{ijt}} + \nu_{ijt}, \tag{7}$$

$$\Leftrightarrow u_{ijt} = \delta_{jt} + \mu_{ijt} + \nu_{ijt}. \tag{8}$$

δ_{jt} is the mean utility across consumers, whereas μ_{ijt} and ν_{ijt} contain all consumer-specific (heteroscedastic) deviations, which the researcher does not observe. Note that δ_{jt} includes all variation in the dimensions j (brand) and t (time). Therefore, conditional on δ_{jt}, biases due to UMC, endogeneity, and dynamics in the utility function are irrelevant.

The extreme value distribution of ν_{ijt} allows the closed-form derivation of the logit choice probability pr_{ijt} that the utility-maximizing consumer i selects brand j at time t (Train, 2009):

$$pr_{ijt} = \frac{\exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{k=1}^{J}\exp(\delta_{kt} + \mu_{ikt})}. \tag{9}$$

Aggregate market shares s_{jt} result from integrating over the consumers' unobserved heterogeneous deviations, and since the heterogeneity in μ_{ijt} results from $\boldsymbol{\upsilon}_i$, we have:

$$s_{jt} = \int_{-\infty}^{\infty} pr_{ijt}(\boldsymbol{\upsilon}_i)\phi(\boldsymbol{\upsilon}_i|\mathbf{0}, \boldsymbol{\Psi})\mathrm{d}\boldsymbol{\upsilon}_i, \tag{10}$$

where ϕ is the density function of a (multivariate) normal distribution. Because there is no closed-form solution for the integral of this market share equation, Monte Carlo methods are required to compute s_{jt}. If $\boldsymbol{\upsilon}_r$ denotes the rth draw of $\boldsymbol{\upsilon}$ from N($\mathbf{0}$, $\boldsymbol{\Psi}$), and $\mathscr{R}$ is the number of random draws, it follows:

$$s_{jt} \approx \frac{1}{\mathscr{R}}\sum_{r=1}^{\mathscr{R}} \frac{\exp\left(\delta_{jt} + \mu_{rjt}(\boldsymbol{\upsilon}_r)\right)}{1 + \sum_{k=1}^{J}\exp\left(\delta_{kt} + \mu_{rkt}(\boldsymbol{\upsilon}_r)\right)}. \tag{11}$$

We use random draws based on Halton sequences because they achieve a higher precision for the calculation of the market share integrals than pseudorandom numbers (Train, 2009). Multiplying the market share (11) with the market volume leads to the demand for brand j at time t: $q_{jt} = M_t \cdot s_{jt}$.

2.2. Price model

Modeling the price function raises the question of how the observed price variation is caused. The literature distinguishes between two different approaches: structural and descriptive pricing models (Chintagunta et al., 2005). Structural models assume that observed prices result from an optimization process, e.g., profit maximization of manufacturers and retailers (Draganska & Klapper, 2007; Yang, Chen, & Allenby, 2003). Such models have the following advantages: (1) in addition to demand, prices are also modeled structurally, (2) if demand and price models are estimated simultaneously, the statistical efficiency is higher, and (3) the endogeneity problem can be solved by explicitly modeling why $\mathbb{E}[\varepsilon_{jt}|\ln(p_{jt})] \neq 0$ holds. However, a disadvantage is that all results are biased if the assumptions with respect to the price model are incorrect (Dubé & Chintagunta, 2003). Furthermore, the estimation – especially with likelihood-based methods and in combination with the demand model – is quite elaborate (Yang et al., 2003, p. 258).

In contrast, descriptive ("reduced-form") price models are less restrictive because they do not require specific assumptions about an optimization process. We prefer the latter approach, which is why the mentioned potential bias due to a mis-specified pricing mechanism is not a problem here. Nevertheless, reduced-form models using instrumental variables allow to control for endogeneity (Train, 2009):

$$\ln(p_{jt}) = \mathbf{z}'_{jt}\boldsymbol{\gamma} + \zeta_{jt}. \tag{12}$$

We specify the log-price as a linear function of the exogenous instrumental variables ($\mathbf{z}_{jt}$) and need for each endogenous variable at least one instrument. For explaining as much exogenous variation in $\ln(p_{jt})$ as possible, $\mathbf{z}_{jt}$ also contains all exogenous variables of the demand model ($\mathbf{x}_{jt}$), as well as possibly nonlinear transformations and interactions of the variables in $\mathbf{z}_{jt}$ (Jiang et al., 2009). The particular choice of valid instruments depends on the specific application (see Section 5.2 for more details). Typical instruments are cost shifters (Chintagunta, 2001), lagged prices (Villas-Boas & Winer, 1999), or prices of other (geographic) markets (Nevo, 2000). All instrumental variables are assumed to be correlated with the price variable, but not with ε_{jt}.

The error term ζ_{jt}, with $\zeta_{jt} \sim N(0, \sigma^2_{\zeta j})$, absorbs any time- and brand-specific deviations of the price variation, from those of the exogenous instruments. As in the demand model, we allow for brand-specific variances. If all the instruments meet the exogeneity condition, possible correlations of UMC and prices result from price variations caused by ζ_{jt}. Thus, the basic idea of Eq. (12) is to separate the price variation into the part that can be explained by exogenous variables and represents the unbiased price effect ($\mathbf{z}'_{jt}\boldsymbol{\gamma}$) and the part that causes the endogeneity problem (ζ_{jt}).

To keep things simple, we do not consider time-varying parameters in the price model (Eq. (12)) and do not find strong dynamics in preliminary analyses in our empirical application. Nevertheless, this model extension is relevant if the effects of instruments change considerably over time (e.g., if the retail scenario changes in an emerging economy). We refer to Kim and Kim (2011) for details.

2.3. Control function approach

As mentioned above, a price-endogeneity problem exists if $\mathbb{E}[\varepsilon_{jt}|\ln(p_{jt})] \neq 0$. There are several possible reasons why UMC

[2] Alternative specifications for the parameter dynamics are possible, e.g., as a function of observed exogenous time-varying variables or in the form of higher-order autoregressive processes (c.f. Van Heerde et al., 2004 and Lachaab et al., 2006). However, these specifications are not relevant to the research questions in this paper and are therefore not considered.

and prices correlate (Train, 2009, p. 315). If a rather high-priced brand has a better shelf space at the retailer than lower-priced brands, ε_{jt} and $\ln(p_{jt})$ are positively correlated and price effects are underestimated. On the other hand, if (unobserved) promotional activities, such as displays, are combined with price reductions, ε_{jt} and $\ln(p_{jt})$ are negatively correlated and price effects are overestimated. It is important to note that ignoring the UMC term will not solve the endogeneity problem. Furthermore, using more data will not solve the endogeneity problem because this will not reduce (or even eliminate) the correlation between ε_{jt} and $\ln(p_{jt})$ (Park & Gupta, 2009; Petrin & Train, 2010).

A simple yet flexible way to model the correlation between the UMC and the prices is a correlation between the error terms (Villas-Boas & Winer, 1999). The distribution for ζ_{jt} and ε_{jt} results in the following multivariate normal distribution (Kim & Kim, 2011):

$$\begin{bmatrix}\zeta_{jt}\\ \varepsilon_{jt}\end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma^2_{\zeta j} & \rho_j\sigma_{\zeta j}\sigma_{\varepsilon j}\\ \rho_j\sigma_{\varepsilon j}\sigma_{\zeta j} & \sigma^2_{\varepsilon j}\end{bmatrix}\right). \quad (13)$$

The covariance matrix in Eq. (13) is denoted as $\mathbf{\Omega}_j$ and all parameters are brand-specific. This is especially important for the correlations ρ_j, because (the degree of) the endogeneity problem can differ across brands (Park & Gupta, 2009). A Cholesky factorization ($\mathbf{\Omega}_j = \mathbf{U}_j' \cdot \mathbf{U}_j$) allows us to decompose the correlated error terms ζ_{jt} and ε_{jt} into the uncorrelated, and standard normal distributed error terms ζ^*_{jt} and ε^*_{jt} (Train, 2009, p. 208):

$$\begin{bmatrix}\zeta_{jt}\\ \varepsilon_{jt}\end{bmatrix} = \begin{bmatrix}\sigma_{\zeta j} & 0\\ \rho_j\sigma_{\varepsilon j} & \sqrt{1-\rho_j^2}\,\sigma_{\varepsilon j}\end{bmatrix} \cdot \begin{bmatrix}\zeta^*_{jt}\\ \varepsilon^*_{jt}\end{bmatrix}, \text{ with } \begin{bmatrix}\zeta^*_{jt}\\ \varepsilon^*_{jt}\end{bmatrix} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_2). \quad (14)$$

The utility function can now be set up to satisfy the condition $\mathbb{E}[\varepsilon_{jt}|\ln(p_{jt})] = 0$. We limit further explanations to δ_{jt} because the mean utility absorbs the endogeneity problem. Using Eq. (14) leads to:

$$\delta_{jt} = \mathbf{x}'_{jt}\bar{\boldsymbol{\alpha}}_t + \bar{\beta}_t \ln(\mathrm{p}_{jt}) + \rho_j\sigma_{\varepsilon j}\zeta^*_{jt} + \sqrt{1-\rho_j^2}\,\sigma_{\varepsilon j}\varepsilon^*_{jt}. \quad (15)$$

The part of ε_{jt} correlating with the price is therefore separated. From Eq. (15) it is clear that there is no endogeneity problem for $\rho_j = 0$ (i.e., no correlation of the error terms) and/or $\sigma_{\varepsilon j} = 0$ (i.e., no UMC). The price is still correlated with ζ^*_{jt}, but we can compute an estimated value, e.g., by regressing $\ln(p_{jt})$ on $\mathbf{z}_{jt}$. ζ^*_{jt} can be calculated from the price Eq. (12):

$$\zeta^*_{jt} = \sigma^{-1}_{\zeta j}\left(\ln(p_{jt}) - \mathbf{z}'_{jt}\boldsymbol{\gamma}\right). \quad (16)$$

This residual term is the so-called control function (CF) (Hausman, 1978; Heckman, 1978; Petrin & Train, 2010), and inserting it into Eq. (15) leads to:

$$\begin{aligned}\delta_{jt} = \;& \mathbf{x}'_{jt}\bar{\boldsymbol{\alpha}}_t + \bar{\beta}_t \ln(p_{jt}) + \rho_j\sigma_{\varepsilon j}\sigma^{-1}_{\zeta j}\left(\ln(\mathrm{p}_{jt}) - \mathbf{z}'_{jt}\boldsymbol{\gamma}\right)\\ & + \sqrt{1-\rho_j^2}\,\sigma_{\varepsilon j}\varepsilon^*_{jt}.\end{aligned} \quad (17)$$

Therefore, the idea of the CF approach is that – conditioned on the price model's residuals – the original price and the error term of the utility function ε_{jt} do not correlate. Solving the endogeneity problem using CF has two advantages. The CF approach (1) can also be used in models with time-varying parameters and (2) enables testing for endogeneity. In the next subsection, we briefly describe both advantages. See also Kim (2010) and Kim and Kim (2011) for more details.

In case suitable instruments are not available, instrument-free methods for dealing with the endogeneity problems have been recently proposed (see Hruschka, 2017 for an overview). In particular, the copula approach of Park and Gupta (2012b) seems promising in our context as it also includes an additional regressor in the model to correct for the endogeneity bias. This correction term uses the nonnormality of the endogenous variable for identification instead of additional information from exogenous instruments as the CF approach. Implementing and testing such an instrument-free method for models with time-varying parameters is a potentially fruitful area for future research. However, we must keep in mind that if the model is estimated in two stages, the problem of "generated regressors" arises, which would lead to biased estimates for the time-varying parameter; hence, caution is in order (cf. Kim & Kim, 2011 for a similar argument regarding a two-step estimation procedure for the CF approach).[3]

2.3.1. The problem of conventional procedures

To solve the endogeneity problem in the demand model, usually, an adjustment of the endogenous variable is carried out using conventional procedures such as 2SLS or the GMM (Greene, 2011). However, these methods are only applicable if two conditions hold: (1) the model must be linear (in its parameters) and (2) the parameter of the endogenous variable must be constant over time. Since the model of Berry et al. (1995) fulfills both conditions, the standard procedures apply. Indeed, being able to handle price endogeneity with standard methods is the main argument for the approach (outlined in Section 3) that linearizes the nonlinear logit model to estimate aggregate discrete choice models (Berry, 1994).

However, linearity is a sufficient condition only if the parameter of the endogenous variable is not time-varying. In the case of models with time-varying parameters, a further condition, namely, $\mathbb{E}[\varepsilon_t|\bar{\beta}_t] = 0$, must be met to avoid an endogeneity bias. The exogeneity condition, therefore, applies to both, the variables and the parameters.

A simple example illustrates that the 2SLS estimator is not valid in the case of time-varying parameters. Suppose the predicted price $\widehat{\ln(p_{jt})} = \mathbf{z}'_{jt}\boldsymbol{\gamma}$ is plugged in Eq. (15). This leads to:

$$\delta_{jt} = \mathbf{x}'_{jt}\bar{\boldsymbol{\alpha}}_t + \bar{\beta}_t\widehat{\ln(p_{jt})} + \varepsilon^\dagger_{jt} \text{ and} \quad (18)$$

$$\varepsilon^\dagger_{jt} = \varepsilon_{jt} + \bar{\beta}_t(\ln(p_{jt}) - \widehat{\ln(p_{jt})}). \quad (19)$$

Conditioned on the exogenous variation of the price ($\widehat{\ln(p_{jt})}$) the time-varying parameter $\bar{\beta}_t$ is correlated with the error term $\varepsilon^\dagger_{jt}$ creating another endogeneity problem. A closer examination of Eq. (17) reveals that this issue is irrelevant for the described CF approach: neither $\bar{\beta}_t$ nor $\ln(p_{jt})$ correlate with the error term in Eq. (17), which is why the CF approach is a valid solution to the endogeneity problem in logit models with time-varying parameters.

2.3.2. Testing for endogeneity

The presented CF approach allows a simple endogeneity test directly via the estimated correlation parameters ρ_j. Under the null hypothesis of a nonexistent endogeneity problem, H_0: $\rho_j = 0$ holds. This hypothesis can be checked with a simple t-test for each brand, or with a likelihood-ratio for all brands at the same time (Greene, 2011). Such endogeneity tests can be understood as extensions of the test of Hausman (1978) for the case of aggregate logit models with heterogeneity and time-varying parameters. Note that the "regular" Hausman test (Greene, 2011) does not apply to models with time-varying parameters because it is explicitly based on the models' estimated parameters, and here, we have T values.

2.4. State space model

To estimate the time-varying parameters, the previously derived demand model (including CF) is formulated compactly as

[3] We thank a reviewer for bringing the idea of instrument-free methods to our attention.

SSM (Harvey, 1989). The advantage is that the KF recursion can be directly applied to the SSM to determine the time-varying parameters (Kalman, 1960). An SSM consists of two coupled (linear) matrix equations: the observation equation that “measures” a dynamic relationship and the state equation that contains the dynamic relationship itself. For the demand model, the SSM $\forall\, t$ is:

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{a}_t + \mathbf{d}_t + \boldsymbol{\varepsilon}_t, \quad \text{with } \boldsymbol{\varepsilon}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_t) \text{ and} \tag{20}$$

$$\mathbf{a}_t = \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, \quad \text{with } \boldsymbol{\eta}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}_t). \tag{21}$$

In this paper, the J mean utility functions Eqs. (15) or (17) form the observation Eq. (20) and hence $\mathbf{y}_t = \left[\delta_{1t} \cdots \delta_{Jt}\right]'$. The matrix $\mathbf{H}_t$ has the dimension $J \times (K+1)$ and contains all variables with a time-varying effect. This includes the dummy variables for the brand intercepts as well as the marketing-mix variables. The $J \times 1$ vector $\mathbf{d}_t$ incorporates all exogenous influences on the mean utility that have no time-varying effect, which are in our case the J effects of the CF: $\mathbf{d}_t = [\rho_1 \sigma_{\varepsilon 1} \zeta^*_{1t} \cdots \rho_J \sigma_{\varepsilon J} \zeta^*_{Jt}]'$. The $J \times 1$ error term vector is $\boldsymbol{\varepsilon}_t = [\varepsilon_{1t} \cdots \varepsilon_{Jt}]'$ with the covariance matrix $\mathbf{R}_t = \mathbf{I}_J \cdot [(1-\rho_1^2)\,\sigma^2_{\varepsilon 1} \cdots (1-\rho_J^2)\,\sigma^2_{\varepsilon J}]'$. The state Eq. (21) includes dynamics in the parameters according to Eqs. (4) and (5). Therefore, all time-varying parameters are combined in the $(K+1) \times 1$ vector $\mathbf{a}_t = [\bar{\boldsymbol{\alpha}}_t' \; \bar{\beta}_t]'$. The assumptions according to Eq. (5) apply to the error term $\boldsymbol{\eta}_t$ and its covariance matrix $\mathbf{Q}_t$.

3. Estimation

This section describes the estimation of the model presented previously. We employ a frequentist approach and distinguish two sets of parameters: time-varying parameters in the demand model, and constant parameters in the demand and the price model. We use likelihood-based estimation for the constant parameters and the construction of the likelihood function is explained in Section 3.1. Based on values for the estimated constant parameters, the time-varying parameters can then be estimated using the KF or Kalman Smoother, which Section 3.2 presents.

3.1. Constant parameters

All constant parameters ($\boldsymbol{\rho}$, $\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}$, vec($\boldsymbol{\Psi}$), $\bar{\boldsymbol{\alpha}}_0$, $\bar{\beta}_0$, vec($\mathbf{Q}$), $\boldsymbol{\gamma}$, and $\boldsymbol{\sigma}_{\boldsymbol{\zeta}}$) are estimated simultaneously to avoid the problem of generated regressors (Pagan, 1984). If the CF were estimated separately first and then were used as regressors in the demand model subsequently, the standard errors at the second stage would be invalid due to uncertainty being ignored. Moreover, in dynamic models, this procedure would bias the estimated time-varying parameters (see Kim, 2010 and Kim & Kim, 2011).

Because simulation techniques are necessary to calculate the market shares in Eq. (10), the estimation is performed with maximum simulated likelihood (MSL) (Train, 2009). Here, the distributional assumptions regarding the error terms of the demand and price model as well as the time-varying parameters, are used to determine the joint distribution of the unknown parameters given the data. The simulated (log-)likelihood function is derived in four steps: (1) the market share inversion, (2) the calculation of the price model's (log-)likelihood function, (3) the calculation of the demand model's (log-)likelihood function, and (4) the calculation of the Jacobian. These four steps are discussed in detail below.

3.1.1. Market share inversion

The first step contains the market share inversion to obtain δ_{jt} (Berry, 1994). These mean utility values then serve in the demand model as dependent variable together with the UMC as an error term. For given parameters in $\boldsymbol{\Psi}$ and random draws, μ_{rjt} can be computed. Berry et al. (1995) suggest a contraction mapping to solve for δ_{jt} and prove that (for MXL models) a unique solution exists, where predicted and observed market shares equate. We compute the following equation iteratively (Train, 2009):

$$\delta_{jt}^{(h+1)} = \delta_{jt}^{(h)} + \ln\left(\frac{\mathrm{s}_{jt}}{s_{jt}\left(\boldsymbol{\delta}_t^{(h)}, \boldsymbol{\mu}_t\right)}\right). \tag{22}$$

The observed and predicted market shares are s_{jt} and s_{jt}, respectively, where the latter is a function of the estimated average utility values as well as the heterogeneous deviations (see Eq. (11)). Eq. (22) is computed until the difference between $\delta_{jt}^{(h+1)}$ and $\delta_{jt}^{(h)}$ is sufficiently small. We follow (Dubé, Fox, & Su, 2012) and choose a very strict criterion: $\|\delta_{jt}^{(h+1)} - \delta_{jt}^{(h)}\| < 1 \cdot 10^{-14}$, $\forall j, t$.

In the case of a model without heterogeneity, δ_{jt} can be obtained via a simple transformation of the observed market shares. Here, the relation $\delta_{jt} = \ln\left(\mathrm{s}_{jt}/\mathrm{s}_{0t}\right)$ holds (Berry, 1994). The resulting homogeneous average values are also useful as initial values for the heterogeneous model in Eq. (22).

3.1.2. Log-likelihood function of the price model

To obtain an estimate for the CF (ζ_{jt}), we regress $\ln(\mathrm{p}_{jt})$ on z_{jt} (see Eq. (16)). Since the price model (12) is linear in ζ_{jt}, the log-likelihood function follows immediately from the assumption of a normally distributed error term (Greene, 2011, p. 550):

$$\mathscr{L}^{\mathrm{P}} = \ln\left(L\left(\boldsymbol{\theta}^{\mathrm{P}} | \mathbf{p}, \mathbf{z}\right)\right) = \ln\left(\prod_{t=1}^{T}\prod_{j=1}^{J} f\left(\mathrm{p}_{jt}, \mathrm{z}_{jt} | \boldsymbol{\theta}^{\mathrm{P}}\right)\right) \tag{23}$$

$$= -0.5 \cdot \sum_{t=1}^{T}\sum_{j=1}^{J}\left[\ln(2\pi\sigma^2_{\zeta j}) + \zeta^2_{jt}/\sigma^2_{\zeta j}\right]. \tag{24}$$

The vector $\boldsymbol{\theta}^{\mathrm{P}} = [\boldsymbol{\gamma}' \; \boldsymbol{\sigma}_{\zeta}']'$ contains all constant parameters of the price model.

3.1.3. Log-likelihood function of the demand model

It is helpful to derive the likelihood function of the demand model via the so-called prediction error decomposition (Schweppe, 1965). The main idea is to utilize the distribution of the prediction errors containing the UMC, instead of directly using the distribution of the UMC. For this purpose, the following KF-equations have to be run through all $t = 1, \ldots, T$ recursively (Harvey, 1989):

$$\boldsymbol{a}_{t|t-1} = \boldsymbol{a}_{t-1|t-1}, \qquad \boldsymbol{P}_{t|t-1} = \boldsymbol{P}_{t-1|t-1} + \mathbf{Q}_t, \tag{25}$$

$$\boldsymbol{e}_t = \boldsymbol{y}_t - \mathbf{H}_t \boldsymbol{a}_{t|t-1} - \mathbf{d}_t, \qquad \boldsymbol{\Sigma}_t = \mathbf{H}_t \boldsymbol{P}_{t|t-1}\mathbf{H}_t' + \mathbf{R}_t, \tag{26}$$

$$\boldsymbol{a}_{t|t} = \boldsymbol{a}_{t|t-1} + \boldsymbol{P}_{t|t-1}\mathbf{H}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{e}_t, \quad \boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\mathbf{H}_t'\boldsymbol{\Sigma}_t^{-1}\mathbf{H}_t\boldsymbol{P}_{t|t-1}. \tag{27}$$

The starting values ($\boldsymbol{a}_{0|0} = [\bar{\boldsymbol{\alpha}}_0' \; \bar{\beta}_0]'$) are also estimated as constant parameters and in this case $\boldsymbol{P}_{0|0}$ is a matrix of zeros (Harvey, 1989, p. 137). The remaining vectors and matrices follow the definitions in Section 2.4 and remember that $\boldsymbol{y}_t$ contains the mean utility $\boldsymbol{\delta}_t$ from the previous market share inversion.

The log-likelihood function of the demand model can easily be constructed with the prediction error $\boldsymbol{e}_t$ from the KF recursion and its covariance matrix $\boldsymbol{\Sigma}_t$:

$$\mathscr{L}^{\mathrm{D}} = \ln\left(L\left(\boldsymbol{\theta}^{\mathrm{D}} | \boldsymbol{\delta}, \mathbf{p}, \mathbf{x}, \boldsymbol{\zeta}^*\right)\right) = \ln\left(\prod_{t=1}^{T} f\left(\boldsymbol{\delta}_t, \mathbf{p}_t, \mathbf{x}_t, \boldsymbol{\zeta}_t^* | \boldsymbol{\theta}^{\mathrm{D}}\right)\right) \tag{28}$$

$$= -0.5 \cdot \sum_{t=1}^{T}\left[\ln(2\pi^J |\boldsymbol{\Sigma}_t|) + \boldsymbol{e}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{e}_t\right]. \tag{29}$$

The vector $\boldsymbol{\theta}^{\mathrm{D}} = [\boldsymbol{\rho}' \; \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}' \; \mathrm{vec}(\boldsymbol{\Psi})' \; \boldsymbol{a}_{0|0}' \; \mathrm{vec}(\mathbf{Q})']'$ contains all constant parameters of the demand model.

3.1.4. Calculation of the Jacobian

To derive the log-likelihood function of the demand model, a nonlinear transformation of the market shares is required. It follows from the change-of-variable theorem for probability density functions (see, e.g., Greene, 2011) that the determinant of the Jacobian of the nonlinear transformation $|\mathbf{J}_{\boldsymbol{e}\to\mathbf{s}}|$ has to be computed and added to the log-likelihood function (Jiang et al., 2009, p. 137). $\mathbf{J}_{\boldsymbol{e}\to\mathbf{s}}$ is the matrix of the first derivatives of the prediction errors with respect to the market shares. This is a block diagonal matrix with T blocks of the dimension $J \times J$. Ignoring this term leads to a misspecification of the entire log-likelihood function and hence to biased estimates. However, as shown by Jiang et al. (2009), the Jacobian only depends on the heterogeneity parameters (i.e., it can be ignored in MNL models). Because of $|\mathbf{J}_{\boldsymbol{e}\to\mathbf{s}}| = |\mathbf{J}_{\mathbf{s}\to\boldsymbol{e}}|^{-1}$, the calculation of the Jacobian determinant is simple (Jiang et al., 2009, p. 138):

$$|\mathbf{J}_{\mathbf{s}_t\to\boldsymbol{e}_t}| = \left|\begin{bmatrix} \partial s_{1t}/\partial e_{1t} & \cdots & \partial s_{1t}/\partial e_{Jt} \\ \vdots & \ddots & \vdots \\ \partial s_{Jt}/\partial e_{1t} & \cdots & \partial s_{Jt}/\partial e_{Jt} \end{bmatrix}\right|. \tag{30}$$

Because the prediction error is linear in $\boldsymbol{\delta}_t$, for the derivative terms in Eq. (30) we have:

$$\frac{\partial s_{jt}}{\partial e_{kt}} = \begin{cases} \int pr_{ijt}(1 - pr_{ijt})\phi(\boldsymbol{\upsilon}_i|\mathbf{0}, \boldsymbol{\Psi})\mathrm{d}\boldsymbol{\upsilon}_i & j = k, \\ -\int pr_{ijt}pr_{ikt}\phi(\boldsymbol{\upsilon}_i|\mathbf{0}, \boldsymbol{\Psi})\mathrm{d}\boldsymbol{\upsilon}_i & j \neq k. \end{cases} \tag{31}$$

The computation of the integrals in Eq. (31) also requires Monte Carlo integration. Conveniently, simulated values (pr_{rjt}) are already available from the market share inversion.

3.1.5. Log-likelihood function of the full model

The four steps described above enables us to derive the joint log-likelihood function, i.e., the product of the log-likelihood functions of the demand and price model as well as the (log) Jacobian determinant:

$$\mathscr{L}(\boldsymbol{\theta}^{\mathrm{D}}, \boldsymbol{\theta}^{\mathrm{P}}|\mathbf{s}, \mathbf{x}, \mathbf{p}, \mathbf{z}) = \mathscr{L}^{\mathrm{D}} + \mathscr{L}^{\mathrm{P}} - \ln(|\mathbf{J}_{\mathbf{s}_t\to\mathbf{e}_t}|). \tag{32}$$

The maximization of this joint log-likelihood function yields the MSL estimates for all constant parameters in the model (summarized in $\boldsymbol{\theta}$):

$$\boldsymbol{\theta}_{MSL} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ \mathscr{L}(\boldsymbol{\theta}^{\mathrm{D}}, \boldsymbol{\theta}^{\mathrm{P}}|\mathbf{s}, \mathbf{x}, \mathbf{p}, \mathbf{z}). \tag{33}$$

To maximize Eq. (32) we employ the gradient-based Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Greene, 2011, p. 1099), and for each iteration, all four steps are necessary.

3.2. Time-varying parameters

To derive the likelihood function of the demand model, the KF recursion has already been presented. There, calculating the prediction errors was its sole purpose. The estimation of the time-variable parameters was a by-product, and only the information available up to and including period t is used. This is appropriate for predictive applications. After estimating the constant parameters, the KF can be applied for each period; as soon as new data are available, the time-variable parameters can be updated.

However, if we want to include the full information in the data (i.e., up to and including period T) for estimating all parameter values for $t = 1, \ldots, T$ (e.g., because we explore parameter dynamics ex-post), it is necessary to apply the Kalman Smoother (KS). The KS is also a recursive algorithm that uses the results of the KF backwards for $t = T-1, T-2, \ldots, 1$ (Petris, Petrone, and Campagnoli, 2009, p. 61):

$$\boldsymbol{a}_{t|T} = \boldsymbol{a}_{t|t} + \boldsymbol{G}_t(\boldsymbol{a}_{t+1|T} - \boldsymbol{a}_{t+1|t}) \text{ and} \tag{34}$$

Table 1
Model specifications and research questions.

Model	Endogeneity	Research question
TVP-MXL(CF)	✓	Are the true parameters identified with the correct model?
TVP-MXL(IV)	↯	What are the consequences when endogeneity is addressed incorrectly?
TVP-MXL	–	What are the consequences when endogeneity is ignored?

$$\boldsymbol{P}_{t|T} = \boldsymbol{P}_{t|t} + \boldsymbol{G}_t(\boldsymbol{P}_{t+1|T} - \boldsymbol{P}_{t+1|t})\boldsymbol{G}_t'. \tag{35}$$

The $(K+1)\times(K+1)$-matrix $\boldsymbol{G}_t$ with $\boldsymbol{G}_t = \boldsymbol{P}_{t|t}\boldsymbol{P}_{T+1|t}^{-1}$, determines the "degree" of the correction (i.e., how smooth the paths of the time-varying parameters are).

This algorithm's purpose is to use the additional available information to smooth the parameter. Once the constant parameters are estimated, the KF's results from its final iteration are inserted into the KS to obtain the smoothed estimates of the time-varying parameters (Petris et al., 2009, p. 144).

4. Simulation study

In this section, we present the setup and results of a Monte Carlo simulation experiment (Kiviet, 2011), which serves three purposes. (1) In our simulation, the true model, error terms, and parameters are known. This allows us to evaluate how well different model specifications are able to recover the true parameters and how large possible biases are (Shugan, 2002). Additionally, the statistical efficiency of different approaches is revealed. (2) Monte Carlo simulations also help to better understand and interpret empirical results (Gelman & Hill, 2007, p. 157), especially if they do not have "textbook character." (3) There is no standard software available for the models in this paper; hence, we wrote our own code. The simulation is also an important tool for testing our programs (Cook, Gelman, & Rubin, 2006).

Starting with a data-generating process (DGP) based on the logit model with heterogeneity, endogeneity, and dynamics, datasets are simulated and then used to estimate three different model specifications. Model 1 (TVP-MXL(CF)) is the model of the DGP and corrects for endogeneity via the CF approach. Model 2 (TVP-MXL(IV)) takes heterogeneity, endogeneity, and dynamics into account, but incorporates ad hoc endogeneity-adjusted prices instead of CF (see Eqs. (18) and (19)). In model 3 (TVP-MXL) endogeneity is ignored completely. Therefore, models 2 and 3 are misspecified and serve the purpose of answering three questions shown in Table 1.

4.1. Data-generating process

The DGP for the simulation is based on the discrete choice model for aggregate data from Section 2. We model both demand and prices for brands in a typical fast-moving consumer goods category (e.g., canned tuna) at a retailer. The setup is similar to the one in Jiang et al. (2009) and the empirical study in Section 5.

There are three brands ($J = 3$) as well as an outside good, and the demand is only influenced by one (endogenous) marketing variable, the (log-)price. As an instrument, the (log-)wholesale price is available, which can be interpreted as a cost for the retailer. This logarithmized variable is simulated from a uniform distribution with values in the interval $[-1.0, -0.2]$, which implies wholesale prices between 0.37 and 0.82, with an average value of 0.56 (sd = 0.13). To facilitate the identification of time-varying parameters, we opt for a long horizon of $T = 300$ (weeks) (Sriram & Kalwani, 2007).

The simulation study is based on the following parameterization. In the price model, we assume $[\gamma_1\ \gamma_2\ \gamma_3\ \gamma_{\text{costs}}]' = [0.2\ 0.1\ 0.1\ 0.6]'$ for the three brand intercepts and the cost parameter. Because of the log-prices and log-costs, the cost parameter can be interpreted as pass-through[4] elasticity with a realistic magnitude of 0.6 (see, e.g., Nijs, Misra, Anderson, Hansen, & Krishnamurthi, 2010). For the elements of the covariance matrix $\boldsymbol{\Omega}$ we have $\sigma_\varepsilon = 0.5,\ \sigma_\zeta = 0.2,$ and $\rho = 0.4$.[5] These values imply that the costs explain approx. 60% of the price variation and thus there is no problem of weak instruments. The correlation of 0.4 is similar to the value assumed by Jiang et al. (2009) who interpret it as a moderate degree of endogeneity. The positive sign means that positive price shocks are related to positive demand shocks; without endogeneity correction, this lowers price effects in absolute terms.[6]

For the parameter values in the utility function at $t = 0$, we define $[\bar{\alpha}_{10}\ \bar{\alpha}_{20}\ \bar{\alpha}_{30}\ \bar{\beta}_0]' = [-5\ \ -6\ \ -6\ \ -5]'$. Because only differences in utility matter, brand $j = 1$ has the highest value. Since the outside good has a utility of zero, the brand intercepts' negative (starting) values imply that the outside good's market share is relatively high ($>90\%$). The temporal variation of the parameters is based on the standard deviations $\sigma_{\eta_\alpha} = 0.05$ and $\sigma_{\eta_\beta} = 0.1$. These values generally lead to plausible parameter values for all periods.[7] For the sake of simplicity, heterogeneity is only assumed in the price effect with $\sigma_{\upsilon_\beta} = \sqrt{2}$.

We employ $\mathscr{R} = 250$ random draws from N(0, 1) based on Halton sequences for simulating the heterogeneity integrals. $\mathscr{S} = 100$ different datasets (replications) are generated by drawing random numbers from the specified distributions of the DGP. Therefore, each replication is based not only on different realizations of the error terms but also on different paths of the time-varying parameters (Kim and Kim, 2011, p. 493). For each replication, constant and time-varying parameters are estimated using the previously described MSL approach and the KS, respectively.

4.2. Results

4.2.1. Constant parameters

Table 2 summarizes for each constant parameter (in each model) the mean value and standard deviation, calculated over all replications. We use true parameter values (2nd column) to test for biased estimates.

Price model: The results in the price model are very good for all models (i.e., no bias and rather small SD-values). Although the price model is irrelevant for the TVP-MXL model, it was estimated to study the price model in isolation. The results show that the choice of the demand model does not influence the results (except for rounding errors), which is not surprising since we use a reduced-form price model (i.e., we do not model prices as the outcome of an optimization process of manufacturers and/or retailers). This step, therefore, ensures that the simulation works without errors.

TVP-MXL(CF) model: Using the correct model also leads to very satisfactory estimation results. The mean values are close to the true values and the empirical standard deviations are small. A significant (but very small) bias is only present for α_{10} and $\sigma_{\eta\alpha}$. Of particular interest is σ_{υ_β}. The simulation demonstrates that it is possible to estimate heterogeneity from aggregate data. The precision of the estimate is comparable to the likelihood-based estimator of Park and Gupta (2009).

TVP-MXL(IV) model: Using the dynamic model with incorrect endogeneity correction leads to biased results in all distribution parameters. It is also striking that the estimates in the demand model are less accurate compared to the TVP-MXL(CF) model; in some cases, the SD-values are up to five times larger (e.g., σ_{υ_β}) Furthermore, σ_ε is overestimated. This is plausible because the price component that is correlated with the UMC is included in the error term. We will discuss consequences for the estimates of the time-varying parameter below. The price parameter heterogeneity is underestimated on average. However, large standard deviations indicate very imprecise results.

TVP-MXL model: Using the dynamic model without endogeneity correction leads to the results regarding heterogeneity and dynamics ($\sigma_{\eta_\alpha},\ \sigma_{\eta_\beta},$ and σ_{υ_β}) that are very similar to the TVP-MXL(CF) model. However, the starting values of the utility parameters ($\{\alpha_{j0}\}_{j=1}^J$ and β_0) are biased. β_0 is (in absolute terms) underestimated, which is due to the positive sign of ρ. Interestingly, the standard deviations of the estimates are almost the same as in the TVP-MXL(CF) model. We can conclude that ignoring price endogeneity leads to typical biases in the demand model. However, negative consequences for the estimation of heterogeneity and dynamics do not seem to exist.

4.2.2. Time-varying parameters

Next, we discuss the results for the time-varying parameters (see Table 3). To evaluate the quality of the estimations, we calculate for each replication mean error (ME), the root mean square error (RMSE), the correlation (Cor), and the maximum absolute deviation (MXAD) per parameter over all periods and then average the results over all replications.

TVP-MXL(CF) model: The correct model enables a precise estimation of the time-varying parameters. The true and estimated parameter trajectories do not differ much, the correlations are very high, and the maximum absolute deviations have reasonable magnitudes, given typical values of the time-varying parameters and their temporal variation. For example, the majority of the price parameters lie between −8 and −2 and here a maximum absolute deviation of 0.71 can be interpreted as small (see Fig. 1 for examples).

TVP-MXL(IV) model: The incorrect endogeneity treatment leads to slightly biased parameter trajectories on average, which can be seen from the mean error values. The brand intercepts (α_{jt}) are overestimated on average, and the time-varying price parameter (β_t) is too high. In contrast to the TVP-MXL(CF) model, the estimated paths are less correlated with the true parameter trajectories. Also, the maximum absolute deviations are larger. Although the average biases are not particularly high, the results suggest that the estimates of the parameter trajectories are inferior to the TVP-MXL(CF) model.

TVP-MXL model: The TVP-MXL leads to opposite results. Here, the utility parameters are clearly biased, which also corresponds to the biased starting values shown in Table 2. The price parameter is underestimated, and the endogeneity problem is evident. On the other hand, the correlations are of similar size as in the TVP-MXL(CF) model, which implies that the estimated parameter paths over time are recovered well. Compared to the TVP-MXL(IV) model, the maximum absolute deviations are even slightly lower despite the endogeneity bias.

In summary, applying the correct model allows a very good estimation of the time-varying parameters. Neglecting price endogeneity or accounting for it incorrectly leads to biased estimates (in particular the price parameter). The biases, however, are of a

[4] "Pass-through" is defined as the fraction of a wholesale price change that the retailer passes through to the consumer.

[5] Please note that we refrain here in the sampling experiment from using brand-specific values for the elements in the covariance matrix to keep things simple.

[6] Train (2009, p. 316) and Park and Gupta (2009, p. 540) discuss the direction of endogeneity biases. All simulations were repeated with a negative correlation ($\rho = -0.4$) without substantially affecting the results.

[7] In particular, the price effect should always be negative. This cannot be guaranteed for random walk processes, but the chosen parameterization aids this condition.

Table 2
Results of the Monte Carlo study (constant parameters).

Parameter	True	TVP-MXL(CF)		TVP-MXL(IV)		TVP-MXL	
		Mean	SD	Mean	SD	Mean	SD
γ_1	0.200	0.201	0.024	0.201	0.023	0.201	0.024
γ_2	0.100	0.101	0.021	0.102	0.021	0.101	0.021
γ_3	0.100	0.100	0.022	0.101	0.021	0.100	0.022
γ_{cost}	0.600	0.604	0.032	0.604	0.032	0.602	0.032
σ_ζ	0.200	0.199	0.005	0.199	0.005	0.199	0.005
$\bar{\alpha}_{10}$	−5.000	−5.039*	0.164	−5.032	0.245	−4.922*	0.164
$\bar{\alpha}_{20}$	−6.000	−6.018	0.145	−5.995	0.257	−5.855*	0.147
$\bar{\alpha}_{30}$	−6.000	−5.988	0.169	−5.960	0.272	−5.820*	0.170
$\bar{\beta}_0$	−5.000	−5.031	0.357	−5.084	0.694	−4.360*	0.355
σ_ε	0.500	0.502	0.063	0.981*	0.166	0.475*	0.012
ρ	0.400	0.396	0.056				
σ_{η_α}	0.050	0.047*	0.008	0.043*	0.016	0.047*	0.008
σ_{η_β}	0.100	0.097	0.021	0.073*	0.061	0.097	0.022
σ_{υ_β}	1.414	1.388	0.173	0.890*	0.761	1.436	0.179

The results are based on $\mathscr{S} = 100$ replications and $\boldsymbol{\theta}_s$ summarizes the results of the sth replication for all time-constant parameters of the model. In addition, $\boldsymbol{\theta}^{\text{true}}$ is the vector of true parameter values. Mean: $\bar{\boldsymbol{\theta}} = \mathscr{S}^{-1} \cdot \sum_{s=1}^{\mathscr{S}} \boldsymbol{\theta}_s$ and standard deviation (SD): $\boldsymbol{\sigma_\theta} = (\mathscr{S}-1)^{-1} \cdot \sum_{s=1}^{\mathscr{S}} (\boldsymbol{\theta}_s - \bar{\boldsymbol{\theta}})^2 \big)^{1/2}$. (*) indicates that the t-statistic is greater than the critical value on the 5% level ($|t| > 1.984$), with $\boldsymbol{t} = \mathscr{S}^{1/2} \cdot (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\text{true}}) \cdot \boldsymbol{\sigma_\theta}^{-1}$.

Table 3
Results of the Monte Carlo study (time-varying parameters).

	TVP-MXL(CF)				TVP-MXL(IV)				TVP-MXL			
	ME	RMSE	Cor	MXAD	ME	RMSE	Cor	MXAD	ME	RMSE	Cor	MXAD
$\bar{\alpha}_{1t}$	< 0.01	0.11	0.91	0.32	0.03	0.17	0.81	0.45	0.11	0.11	0.91	0.39
$\bar{\alpha}_{2t}$	< 0.01	0.11	0.90	0.33	0.04	0.18	0.75	0.49	0.17	0.11	0.90	0.46
$\bar{\alpha}_{3t}$	< 0.01	0.12	0.92	0.34	0.04	0.18	0.81	0.49	0.18	0.12	0.92	0.48
$\bar{\beta}_t$	< 0.01	0.22	0.92	0.71	−0.04	0.45	0.70	1.29	0.69	0.26	0.92	1.26

The results are based on $\mathscr{S} = 100$ replications each with $T = 300$ periods and $\boldsymbol{\theta}_{t,s}$ summarizes the results of the sth replication and tth period for all time-varying parameters of the model. In addition, $\boldsymbol{\theta}_{t,s}^{\text{true}}$ is the vector of true parameter values. Fit measured are computed for each replication across time periods and then averaged over replications. Mean error (ME) $= \mathscr{S}^{-1} \cdot \sum_{s=1}^{\mathscr{S}} T^{-1} \cdot \sum_{t=1}^{T} (\boldsymbol{\theta}_{t,s} - \boldsymbol{\theta}_{t,s}^{\text{true}})$; root mean squared error (RMSE) $= \mathscr{S}^{-1} \cdot \sum_{s=1}^{\mathscr{S}} \big(T^{-1} \cdot \sum_{t=1}^{T} (\boldsymbol{\theta}_{t,s} - \boldsymbol{\theta}_{t,s}^{true})^2\big)^{1/2}$; correlation (Cor) $= \mathscr{S}^{-1} \cdot \sum_{s=1}^{\mathscr{S}} \big(\sum_{t=1}^{T} (\boldsymbol{\theta}_{t,s} - \bar{\boldsymbol{\theta}}_s) \cdot (\boldsymbol{\theta}_{t,s}^{\text{true}} - \bar{\boldsymbol{\theta}}_s^{\text{true}})\big) / \big((\sum_{t=1}^{T} (\boldsymbol{\theta}_{t,s} - \bar{\boldsymbol{\theta}}_s)^2 \cdot \sum_{t=1}^{T} (\boldsymbol{\theta}_{t,s}^{\text{true}} - \bar{\boldsymbol{\theta}}_s^{true})^2)^{1/2}\big)$; max absolute deviation (MXAD) $= \mathscr{S}^{-1} \cdot \sum_{s=1}^{\mathscr{S}} \max_t \big(\text{abs}(\boldsymbol{\theta}_{t,s} - \boldsymbol{\theta}_{t,s}^{\text{true}})\big)$.

different nature. The results suggest that ignoring endogeneity only "shifts" the parameters in parallel while representing the shape of the paths over time well. In the TVP-MXL(IV) model, on the other hand, the evolutions of utility parameters are also biased, but the average values are similar to the true (average) values.

To further examine these findings, we visually inspect the paths of the price parameters for all replications. We find similar results and can roughly distinguish four scenarios (A, B, C, and D). Fig. 1 depicts four replications, each representing one particular scenario.

From the four scenarios, the findings discussed in Table 3 become immediately apparent: (1) the TVP-MXL(CF) model is suited to accurately estimate the true price parameter trajectories, and (2) the TVP-MXL model causes a bias in the form of a parallel shift. The TVP-MXL(IV) model leads to different results for the four scenarios. At best (scenario A), the differences in the results between the TVP-MXL(IV) and the TVP-MXL(CF) model are small. The TVP-MXL(CF) model is nevertheless obviously superior. In scenario B, the temporal variation of the price parameter in the TVP-MXL(IV) model is too small, but there is some similarity to the true parameter trajectory. This kind of bias is presumably less problematic than the one of the TVP-MXL model. However, it might also be the case (scenario C) that there is a very high temporal variation in the estimated price parameters of the TVP-MXL(IV) model, thus resulting in a bias that exceeds the bias of the TVP-MXL model. In scenario D, the TVP-MXL(IV) model is not able to identify any dynamics in the price parameter at all. Here, the TVP-MXL(IV) model fails to fulfill its main purpose and is not useful if the estimation of time-varying parameters is of interest. This happens in 30 of the 100 replications.

We conclude from the simulation study that handling endogeneity in case of time-varying parameters requires the proposed CF approach (TVP-MXL(CF) model). Otherwise, the results can be worse from an incorrect endogeneity treatment (TVP-MXL(IV) model) than for ignoring endogeneity completely (TVP-MXL model). Furthermore, using the TVP-MXL(IV) model also leads a loss of (statistical) efficiency.

5. Empirical study

In this section, we illustrate the application of the model with a real dataset. After a description of the data, the estimation results are presented and discussed. Finally, the implications of correcting for endogeneity on optimal retail pricing are analyzed.

5.1. Data

We use aggregate data from the Dominick's Finer Foods (DFF) database of the University of Chicago (Booth School of Business).[8] DFF is one of the major supermarket chains in the greater Chicago area with more than 80 stores. Weekly values are used for the variables unit sales, price, promotion, and cost (wholesale prices) of the product category canned tuna between September 1989 and May 1997 (almost 400 weeks). Due to missing values, only 337 weeks are available for the estimation. The data are aggregated at the brand level across the stores, and the four top-selling brands, namely, Bumble Bee (BB), Chicken of the Sea (CS), the store brand Heritage House (HH), and StarKist (SK), which are responsible for

[8] https://www.chicagobooth.edu/research/kilts/datasets/dominicks.

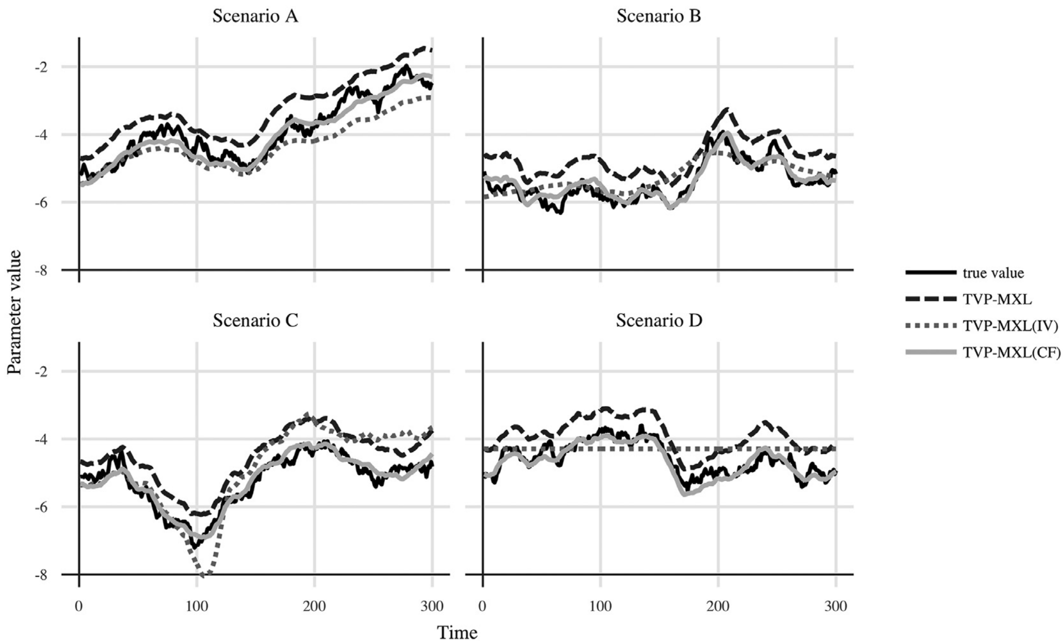

Fig. 1. Time-varying price parameters of selected replications.

Table 4
Descriptive statistics of the canned tuna category.

Variable	Bumble Bee		Chicken of the Sea		Heritage House		StarKist	
Unit sales	1.470	(3.323)*	1.443	(3.905)	0.924	(1.328)	2.085	(3.759)
Market share	0.223	(0.186)	0.229	(0.186)	0.196	(0.137)	0.352	(0.205)
Price	0.779	(0.112)	0.801	(0.085)	0.691	(0.068)	0.804	(0.102)
Cost	0.544	(0.056)	0.573	(0.045)	0.497	(0.046)	0.561	(0.051)
Promotion	0.234	(0.348)	0.344	(0.405)	0.244	(0.364)	0.308	(0.389)

(*) Mean and standard deviation in parentheses. The variables are scaled as follows: unit sales in $10^4 \times$ 6-oz cans, market shares in % as well as prices and costs in US $ per 6-oz can. The promotion variable refers to the percentage of the sales sold under promotion. The market shares are "inside-good market shares" and are therefore conditioned on buying canned tuna. The outside-good itself has a market share of approx. 97 %.

most of the category sales, are used in the analysis. Demand (and prices) refer to tuna with low quality ("chunky light") in cans 6-oz ($\approx$170 grams). The data contain almost 2 million customers per week (mean = 1934321, sd = 178550) and we use this weekly number of customers to define the market volume (M_t) for computing the market shares of the four brands and the outside good. This calculation is based on the assumption that choices refer to single cans of tuna (Jiang et al., 2009).

Table 4 contains descriptive statistics of the DFF data. SK has the largest demand and thus also the highest market share, whereas HH has the lowest demand and the smallest market share. BB and CS are approx. at the same level between HH and SK. The prices of CS and SK are $ 0.8 per can, BB and HH are on average $ 0.02 and $ 0.11 cheaper. A similar picture emerges for the costs. The profit margin for DFF in this category is on average 29 %. The brands differ slightly in the intensity of their promotion activity, where CS has the highest value, followed by SK, BB, and HH.

All variables vary significantly over time, which is typical for data at the retailer-level and, of course, helps to identify marketing-mix effects. The DFF dataset, and the canned tuna product category in particular, offers even more advantages in light of our research questions. (1) The data extend over a long period, which facilitates the identification of time-varying parameters (Sriram & Kalwani, 2007). (2) Heterogeneity (Jiang et al., 2009), endogeneity (Meza & Sudhir, 2006), and dynamics in the form of seasonal effects (Chevalier et al., 2003) have already been addressed using this dataset. A joint consideration of all three aspects using a flexible modeling approach for the time-varying parameters appears as a logical extension of the existing research. (3) The canned tuna market can be adequately modeled with a small number of brands, which reduces the modeling effort of the analysis. BB, CS, HH, and SK account for approx. 80 % of the market. (4) Canned tuna brands are not heavily advertised, so it is not an issue that no data on advertising are available. This would not be the case for other product categories, e.g., beer or laundry detergent.[9]

5.2. Instruments

For the CF approach, we need suitable instrumental variables. Following (Chintagunta, 2002; Meza & Sudhir, 2006), and (Sriram, Balachander, & Kalwani, 2007), we choose the wholesale price as an instrument for the retailer's price. Several reasons support this decision (see Chintagunta, 2002, p. 147 for more details): Wholesale prices are generally highly correlated with the retailer's prices because for the retailer the wholesale prices can be interpreted as costs. When costs change, the retailer will probably adjust prices.[10] However, the UMCs are most likely uncorrelated with costs, e.g.,

[9] Please note that practical reasons mainly drive the choice of the particular dataset. Even though the dataset is more than 20 years old it works well as a simple case study to further illustrate the proposed approach using real data. Given that the dataset is publicly available and has been analyzed before also facilitates reproducibility and interpretation of the results. However, we do not claim that the substantive insights about tuna preferences necessarily still apply today.

[10] See also the discussion in Train (2009, p. 311): More than 80 % of interviewed managers in an empirical study indicate that they are using a fixed markup on the cost for pricing ("cost-plus pricing").

the retailer will not usually adjust the shelf space allocation when wholesale prices change. On the other hand, Rossi (2014) critically remarks that wholesale prices would not be exogenous if manufacturers adjust this variable to market conditions. In our case, this is unlikely because DFF with 25% of all supermarket sales does not dominate the retail market in Chicago (Chintagunta, Dubé, & Singh, 2003). Hence, it seems unlikely that (sales shifts of) DFF should have an impact on wholesale prices, which have to be set at the market level in any case (not the retail chain level), according to Chintagunta (2002). Furthermore, the wholesale price has proven to be a valid instrument throughout several empirical studies: Nijs, Srinivasan, and Pauwels (2007) investigated determinants of retailer prices and profits and concluded that wholesale prices play an important role. Since the authors examined the same DFF data as we do, this particularly applies to our study. Fong, Simester, and Anderson (2011) cooperated with a retailer for their study. Price elasticities estimated from data of a joint field experiment were compared with price elasticities calculated from the retailer's nonexperimental data. As expected, the estimated elasticities were biased due to endogeneity. However, using the wholesale price as instrument solved the problem and the authors conclude: *"The wholesale price instrument eliminates far more bias than it introduces, and appears to be an effective way to estimate price sensitivity"* (Fong et al., 2011, p 17).[11]

Also, the promotion variable is assumed to be exogenous and qualifies as an instrument for the price of the retailer. This approach is not unusual in marketing (see, e.g., Chintagunta, 2001 and Sriram et al., 2007), and DFF store managers confirmed according to Chintagunta, Dubé, and Singh (2002, p. 206), that promotion activities are typically planned several weeks in advance. Therefore, it seems unlikely that the promotion variable and the error term correlate at the demand level (Chintagunta et al., 2003). However, promotion activities and prices are generally highly correlated as retailers coordinate their prices with their planned promotional activities.

Finally, an interaction term between the wholesale price and promotion variable explains pass-through differences in promotion weeks, and brand-specific intercepts account for the brands' different price levels (see Table 4). Since we use log-prices in the demand model, the wholesale price is also logarithmized. On the one hand, this approach improves the data fit; on the other hand, this approach makes it easier to interpret the results (in terms of plausibility) since in this case the pass-through elasticity is estimated directly.

To get a first impression about the quality of the instruments and to check their plausibility, a pooled regression of the prices is performed using the aforementioned instrumental variables. Eq. (36) summarizes the results (with standard errors in parentheses):

$$\begin{aligned}\ln(\text{price}_{jt}) &= \underset{(0.021)}{\mathbf{0.027}} \cdot \mathbb{I}(j = \text{BB}) + \underset{(0.020)}{\mathbf{0.055}} \cdot \mathbb{I}(j = \text{CS}) \\ &\quad - \underset{(0.024)}{\mathbf{0.048}} \cdot \mathbb{I}(j = \text{HH}) + \underset{(0.021)}{\mathbf{0.061}} \cdot \mathbb{I}(j = \text{SK}) + \underset{(0.034)}{\mathbf{0.397}} \cdot \ln(\text{cost}_{jt}) \\ &\quad - \underset{(0.038)}{\mathbf{0.117}} \cdot \text{promotion}_{jt} + \underset{(0.060)}{0.111} \cdot \ln(\text{cost}_{jt}) \cdot \text{promotion}_{jt}, \\ &\quad \sigma_\zeta = 0.096,\ R^2 = 0.548,\ N = 1348. \end{aligned} \tag{36}$$

With an R^2 value of 0.548, the instruments have sufficient explanatory power. The costs impact prices significantly positive, the pass-through elasticity has a plausible value (see Nijs et al., 2010), and promotional activities reduce prices significantly. However, the interaction term is not significant at the 5% level. In sum, we conclude that the instrumental variables are appropriate for our empirical study.

5.3. Results

Before discussing the estimation's results, a brief discussion of modeling or estimation issues that occurred when studying the canned tuna data is in order. A model specification in which all parameters are heterogeneous and correlated with one another, leads to poorly converging and unstable MSL estimates. Especially in combination with time-varying parameters, preliminary estimates showed that only price parameter heterogeneity is important. For this reason, below only the price variable is specified heterogeneously.[12]

Seasonality plays a role in the sales of canned tuna. Chevalier et al. (2003) and Meza and Sudhir (2006) use indicator variables and show that the price elasticity increases during Lent (in absolute terms). Lent refers to the period of 6 weeks before Easter Sunday, where religious Christians abstain from eating meat. This leads to a period of peak demand for meat substitutes, such as (tuna) fish. One possible explanation for the increase in price elasticity during this period is that consumers engage in additional price search and buy tuna more often (Meza & Sudhir, 2006). Combining Lent indicators with time-varying parameters did not provide satisfactory results in this paper, but this is not surprising because of the flexibility of random walk processes. It is more promising to consider the seasonality in the variance of the UMC. This is very easy within an SSM by specifying the variance as a function of the Lent indicators $\sigma^2_{\varepsilon\, jt} = \sigma^2_{\varepsilon\, j} \cdot \exp\left(\tau \cdot \mathbb{I}(t = \text{Lent})\right)$. For $\tau \neq 0$ we have (temporal) conditional heteroscedasticity, and the KF then works in the weeks during Lent "differently". Indeed, this is reasonable since Chevalier et al. (2003) and Meza and Sudhir (2006), conclude that the canned tuna demand is more volatile during Lent (consequently $\tau > 0$).

We tried more complex price models than Eq. (36), e.g., with brand-specific cost, promotion, and interaction effects, but the fit improved only marginally. Since the joint MSL estimation of the price and the demand model is computational somewhat demanding, a more parsimonious model specification (without brand-specific effects) is advantageous. Preliminary analyses (without price parameter heterogeneity) revealed that the different price models do not significantly influence the overall results. For this reason, we opt for the simpler version with brand intercepts and without brand-specific effects.

A meaningful estimation of the TVP-MXL(IV) model is not possible for the tuna data. Although the likelihood optimization converges, the parameter estimates are not reasonable. In particular, the price parameter varies over time but has a positive sign in many weeks, which is economically implausible. Moreover, the price parameter heterogeneity is extremely high. This result underlines that the TVP-MXL(IV) model is not suited to investigate time-varying endogenous price effects. Hence, we do not include this model in the model comparison.

Table 5 contains the estimation results of the constant parameters for the TVP-MXL and the TVP-MXL(CF) model using $\mathscr{R} = 1000$ Halton draws. Many parameters are significantly different from zero and have plausible signs. Moreover, both models share similarities and differences in values often lie within the estimates' uncertainty.

In particular, the standard deviations for time-varying parameters and error terms are significantly positive in both models. Therefore, UMC and dynamics play a major role in the canned

[11] Nevertheless, the validity of an instrument needs to be checked for each application separately and we do not argue that wholesale prices are necessarily valid instruments.

[12] This is not unusual in applications in the marketing literature, see e.g., (Chung, Derdenger, & Srinivasan, 2013; Draganska & Klapper, 2007; Kim, Blattberg, & Rossi, 1995).

Table 5
Estimation results.

	TVP-MXL		TVP-MXL(CF)	
Parameter	Estimate	Std. Err.	Estimate	Std. Err.
$\bar{\alpha}_{\mathrm{BB}\,0}$	$\mathbf{-6.568}$	0.205	$\mathbf{-6.339}$	0.203
$\bar{\alpha}_{\mathrm{CS}\,0}$	$\mathbf{-6.416}$	0.199	$\mathbf{-6.227}$	0.198
$\bar{\alpha}_{\mathrm{HH}\,0}$	$\mathbf{-7.557}$	0.285	$\mathbf{-7.214}$	0.283
$\bar{\alpha}_{\mathrm{SK}\,0}$	$\mathbf{-5.676}$	0.172	$\mathbf{-5.501}$	0.172
$\bar{\alpha}_{\mathrm{promo}\,0}$	−0.041	0.182	0.128	0.183
$\bar{\beta}_0$	$\mathbf{-7.034}$	0.963	$\mathbf{-5.777}$	0.916
$\sigma_{\varepsilon\,\mathrm{BB}}$	**0.333**	0.018	**0.368**	0.036
$\sigma_{\varepsilon\,\mathrm{CS}}$	**0.396**	0.021	**0.410**	0.024
$\sigma_{\varepsilon\,\mathrm{HH}}$	**0.284**	0.017	**0.285**	0.018
$\sigma_{\varepsilon\,\mathrm{SK}}$	**0.391**	0.018	**0.402**	0.024
τ	**1.363**	0.173	**1.301**	0.170
$\sigma_{\eta_{\alpha\,\mathrm{BB}}}$	**0.078**	0.018	**0.085**	0.018
$\sigma_{\eta_{\alpha\,\mathrm{CS}}}$	**0.068**	0.021	**0.069**	0.021
$\sigma_{\eta_{\alpha\,\mathrm{HH}}}$	**0.121**	0.020	**0.118**	0.020
$\sigma_{\eta_{\alpha\,\mathrm{SK}}}$	**0.048**	0.012	**0.043**	0.012
$\sigma_{\eta_{\alpha\,\mathrm{promo}}}$	**0.044**	0.015	**0.040**	0.014
σ_{η_β}	**0.423**	0.049	**0.434**	0.048
σ_{υ_β}	0.787	0.807	1.127	0.690
ρ_{BB}			$\mathbf{-0.546}$	0.111
ρ_{CS}			$\mathbf{-0.252}$	0.106
ρ_{HH}			−0.020	0.164
ρ_{SK}			−0.208	0.118
$\mathscr{L}$	6455.150		6465.512	

Bold parameters are significantly different from zero (at least) on the 5% level (two-sided). The $\mathscr{L}$-value refers only to the log-likelihood value of the demand model and the Jacobian, so both models are comparable. However, the TVP-MXL(CF) model is estimated in one step including the price model. As the results for the price model are nearly identical to those of the price regression (36), we omit them here for brevity's sake.

tuna data, whether endogeneity is considered or not. τ is also significantly positive and thus there are more extreme demand shocks during the peak-demand period of Lent. For both models, the UMC variance increases more than threefold (TVP-MXL(CF) model: $\exp(1.301) \approx 3.673$). Heterogeneity is only significant in the TVP-MXL(CF) model (one-sided at the 10% level). Because σ_{υ_β} has a reasonable magnitude in both models and is essential for realistic substitution patterns (Dubé et al., 2002), the heterogeneous model specification is used nevertheless.

The results regarding endogeneity are in the focus of our study. The ρ_j parameters for BB and CS differ from zero significantly. Moreover, the joint null hypothesis that there is generally no endogeneity problem is rejected based on a likelihood ratio test with a highly significant χ^2 statistic of $-2 \cdot (6455.150 - 6465.512) = 20.724$, with $k = 4$ degrees of freedom. This test also shows that the TVP-MXL(CF) model is statistically superior to the TVP-MXL model because both models are nested and only differ in their endogeneity specification. The negative sign of the parameter is worth mentioning. Prices and UMC (Villas-Boas & Winer, 1999) are often positively correlated, but negative correlations are possible and imply that unobserved promotion activities (e.g., coupons, expanded shelf-space allocation, or favorable shelf locations) accompany price reductions (Park & Gupta, 2009).

The top part of Table 5 shows the initial values ($t = 0$) of the utility parameters. Before further analyzing the estimated paths of these time-varying parameters, an endogeneity bias can be recognized from the starting values already. Prima facie, the TVP-MXL model's brand utility parameters are too low and the price sensitivity (without an endogeneity correction) is too high. This stems from the previously mentioned negative correlation ($\rho_j < 0$), but tests revealed the difference for none of the parameters is significant if tested separately. Fig. 2 visualizes the paths of the time-varying parameters for both models. All utility parameters have smaller values in the TVP-MXL model than in the TVP-MXL(CF) model. On the one hand, these biases match those of the starting values; on the other hand, the already discussed parallel shifts of the parameters can be seen again. However, the figure shows that all parameters vary over time. This variation is also different from parameter to parameter in form and intensity and does not correspond to monotonous trends. This demonstrates the superiority of the random walk specification, wherein no assumption is needed a priori (Lachaab et al., 2006). Fig. 2 also contains 95% confidence intervals for each parameter. Two aspects are noteworthy. (1) The confidence intervals overlap for both models at any time. However, this does not indicate whether the endogeneity bias is significant. Such a test is performed using the ρ_j parameters, as mentioned above and described in Section 2.3.2. (2) No data are available during weeks 331 to 372. The KF and the KS can handle these missing values without problems and still estimate values. However, the confidence intervals are wider during this period because the estimates are more uncertain.

The brand intercepts for BB, CS, and SK evolve similarly: decreasing in the first 150 weeks and subsequently increasing again. In absolute terms, SK has the highest values; BB and CS are at a comparable level. HH often has the lowest values (except weeks 100–150), but this is not surprising because HH is DFF's store brand. Furthermore, HH has a different path over time compared to BB, CS, and SK. After a relatively stable trend in the first 240 weeks there is a significant dip, but the brand intercept recovers within a year. The promotion parameter has basically a rising trend over the entire period of the data. However, there are two short downturns, after 100 and 200 weeks, respectively. The price parameter is subject to very strong fluctuations, which is also apparent from the high σ_{η_β} value of approx. 0.4. The (mean) price parameters of the TVP-MXL(CF) model are between -5.777 and -1.740, with an average value of -3.853. The average value of the TVP-MXL model's price parameter is -5.177, so the endogeneity bias over time is on average -1.324 (sd $= 0.212$).

As mentioned before, some authors use an interaction term between the Lent indicator and the price variable to model seasonal variations in the price effect. A Lent indicator and price parameters of the TVP-MXL(CF) model correlate with -0.176. This replicates the results of Chevalier et al. (2003), who find a higher price sensitivity (in absolute terms) during the Lent period. In a GLS dummy regression of the price parameter (based on the TVP-MXL(CF) model accounting for estimation uncertainty) on the Lent indicator, Lent has a highly significant negative effect of -0.644. However, the low R^2 value of only 3.7% indicates that the Lent indicator explains only a small fraction of the time variation. This emphasizes the benefits of a more flexible specification of the time-varying parameters.

It should be noted that $\bar{\beta}_t$ is only the average value of the normally distributed price parameter. For the TVP-MXL(CF) model, the value of 1.127 for σ_{υ_β} implies that 90% of the population values lie between -1.851 and 1.851, when the mean value is set to zero. This highlights that price parameter variations caused by heterogeneity and dynamics can have quite similar magnitudes, although the causes are entirely different. In sum, the model estimation for DFF's canned tuna category reveals that heterogeneity and endogeneity are relevant and all parameters vary over time.

5.4. Implications for optimal retail-pricing

To examine the practical implications of using the "correct" model, retail price optimizations based on the TVP-MXL(CF) and TVP-MXL model are performed and compared with each other. We use a similar setup as in Kim et al. (1995) and Vilcassim and Chintagunta (1996). We assume: (1) the retailer maximizes profits in the canned tuna category for each week in all stores. (2) Retailer competition is ignored. (3) The retailer only optimizes prices. The

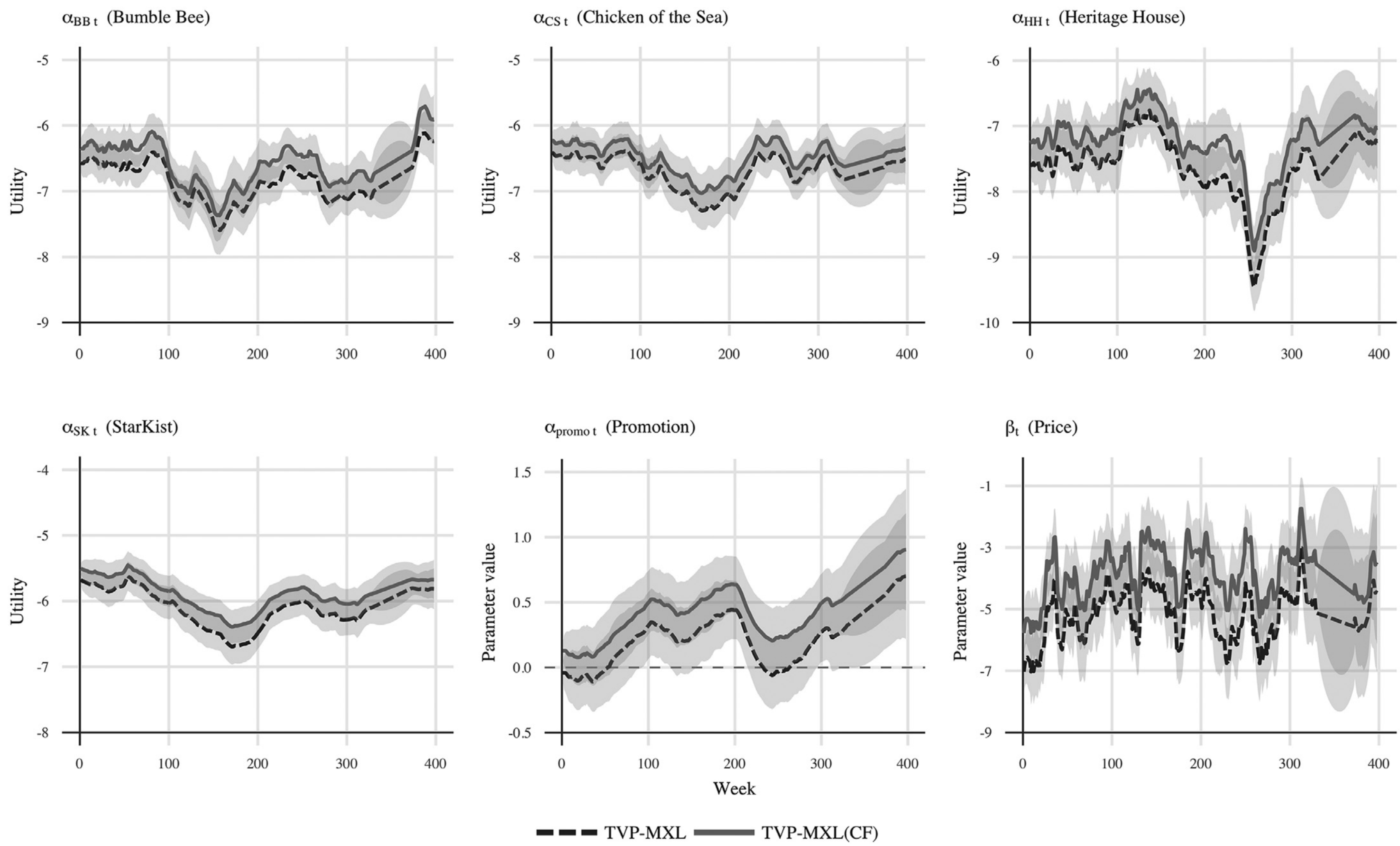

Fig. 2. Estimated parameter paths over time.

values for the promotion and cost variables are known to the retailer, but they are considered as given. (4) The comparisons are based on the first 300 weeks[13] of the data. We perform the optimization over several weeks to cover a broad range of values for all exogenous variables and utility parameter values. (5) The actual prices serve as the basis for a status quo benchmark.

For the retailer's optimization problem, we have for each t:

$$\max_{\mathbf{p}_t} \ \pi_t = M_t \cdot \sum_{j=1}^{J} (p_{jt} - c_{jt}) \cdot s_{jt}(\mathbf{p}_t). \tag{37}$$

Optimal prices within a week are linked because of the structure of the logit model, one brand's price has an effect on all other brands. However, Eq. (37) can be maximized separately for each t, because prices do not have intertemporal effects in the model. The first order condition for an optimum is:

$$s_{jt}(p^*_{jt}) + \sum_{k=1}^{J} \frac{\partial s_{kt}(p^*_t)}{\partial p_{jt}} \cdot (p^*_{kt} - c_{kt}) \equiv 0. \tag{38}$$

Eq. (38) can be written compactly as a matrix equation:

$$\mathbf{p}^*_t = \mathbf{c}_t - \mathbf{\Delta}_t(\mathbf{p}^*_t)^{-1} \cdot \mathbf{s}_t(\mathbf{p}^*_t). \tag{39}$$

The vector $\mathbf{p}^*_t$ contains the optimal prices for all J brands at the time t. The matrix $\mathbf{\Delta}_t(\mathbf{p}^*_t)$ summarizes all first derivatives of the market shares with respect to the vector of (optimal) prices (Chintagunta et al., 2003). $\mathbf{s}_t(\mathbf{p}^*_t)$ is the vector of the market shares at time t evaluated at the optimal price $\mathbf{p}^*_t$ and the J values of the variable costs are found in $\mathbf{c}_t$. Because $\mathbf{\Delta}_t(\mathbf{p}^*_t)$ and $\mathbf{s}_t(\mathbf{p}^*_t)$ depend on the optimal prices, Eq. (39) is only an optimality condition. However, this equation can be computed iteratively as a contraction mapping to obtain a numerical solution for the optimum (Morrow & Skerlos, 2011). Starting with an arbitrary vector $\mathbf{p}^{(0)}_t$ (e.g., average prices) we compute the following equation iteratively:

$$\mathbf{p}^{(h+1)}_t = \mathbf{c}_t - \mathbf{\Delta}_t\left(\mathbf{p}^{(h)}_t\right)^{-1} \cdot \mathbf{s}_t\left(\mathbf{p}^{(h)}_t\right). \tag{40}$$

We iterate until $\left\| \mathbf{p}^{(h+1)}_t - \mathbf{p}^{(h)}_t \right\| < 1 \cdot 10^{-14}$, save the solution as $\mathbf{p}^*_t$ for week t, and repeat the optimization for all $t = 1, \ldots, T$. The algorithm is easy to implement, robust, and converges quickly.[14]

To investigate the consequences of accounting for endogeneity, optimizations based on the TVP-MXL(CF) and the TVP-MXL model are carried out. The optimal prices and actual prices are then plugged into the TVP-MXL(CF) model (the "true" model) for the calculation of the (hypothetical) market shares and profit margins (see, e.g., Dubé, Hitsch, Rossi, & Vitorino, 2008). This ensures the comparability of the results since the evaluation of the different prices refers to the same model. Table 6 summarizes the results of the optimal retail-pricing exercise based on the different models. The top part shows the results based on the actual prices, which serve as a benchmark. The market shares generated by the TVP-MXL(CF) model are quite close to the true values. Actual profit margins are between 27 % and 28.5 % and the retailers profit for the canned tuna is on average $ 8283 per week.

The middle part of the table presents the results of the price optimization based on the TVP-MXL(CF) model. The optimal prices are (on average) only slightly lower than the actual prices, with

[13] Because of many missing values at the end of the investigation period, only 300 of the total 337 weeks are taken into account.

[14] Note that Morrow and Skerlos (2011) studied also more sophisticated versions of fixed-point iterations with better properties for problems with hundreds of alternatives in the choice set. Because of the rather simple setup ($J = 4$), the fixed-point iteration based on markups works well in our application.

Table 6
Optimal retail-pricing.

Brand	Price	Market share (cond.)**	Market share (uncond.)**	Profit margin
		Actual prices		
Bumble Bee	0.764 (0.108)*	22.055 (12.887)	0.529 (0.488)	28.356 (10.322)
Chicken of the Sea	0.792 (0.086)	22.060 (12.743)	0.557 (1.063)	26.977 (9.016)
Heritage House	0.687 (0.067)	19.483 (11.603)	0.467 (0.625)	27.096 (9.911)
StarKist	0.795 (0.096)	36.401 (15.716)	0.929 (1.204)	28.484 (8.968)
Retailer Profit		8282.926 (4977.618)		
		Optimal prices based on the TVP-MXL(CF) model		
Bumble Bee	0.727 (0.086)	23.059 (8.718)	0.695 (0.528)	25.497 (4.981)
Chicken of the Sea	0.775 (0.069)	19.948 (7.183)	0.618 (1.017)	25.932 (5.154)
Heritage House	0.661 (0.063)	19.722 (10.285)	0.641 (1.035)	24.882 (4.490)
StarKist	0.761 (0.077)	37.271 (9.255)	1.138 (1.555)	25.800 (5.152)
Retailer Profit		9820.554 (5556.297)		
		Optimal prices based on the TVP-MXL model		
Bumble Bee	0.673 (0.067)	22.945 (8.385)	0.880 (0.589)	19.804 (2.727)
Chicken of the Sea	0.714 (0.053)	20.188 (7.058)	0.789 (1.096)	19.869 (2.729)
Heritage House	0.618 (0.053)	19.275 (10.015)	0.792 (1.139)	19.726 (2.645)
StarKist	0.702 (0.056)	37.592 (9.008)	1.447 (1.590)	19.852 (2.821)
Retailer Profit		9519.483 (5594.615)		

(*) Mean and standard deviation in parentheses, calculated over weekly values. The small differences compared to the values in Table 4 are explained by the shorter time period of the data (300 weeks). (**) All market share calculations are based on the TVP-MXL(CF) model. Market shares and profit margins ((price – cost)/price) are specified in percentage values. The conditional market shares refer only to the brands within the category, while the unconditional market shares also include the outside good. Prices and profits are in scaled $ and refer to 6 oz cans.

differences between $ 0.017 (CS) and $ 0.037 (BB).[15] A similar picture emerges for the profit margins. The lower optimal prices reduce the attractiveness of the outside good. Also, the conditional market shares change, at the optimum, CS loses about two percentage points and drops to the level of HH, whereas BB, the former direct competitor of CS, increases its market share by one percentage point. SK expands its market leadership (approx. 0.9 percentage points) and thus remains the most popular tuna brand at DFF. The retailer profit is $ 9820.6 per week for optimal prices. This means by using the optimal prices based on the TVP-MXL(CF) model the retailer could increase profits by $ 1538.7 per week (≈ 18.6%). For a whole year, this adds up to almost $ 80000. Consequently, this is the maximum amount such a model-based optimization may cost yearly (including data collection and handling, model evaluation and implementation) to be profitable for the retailer.

The lower part of the table illustrates what happens when endogeneity is ignored in the optimization. The optimal prices based on the TVP-MXL model are much lower (e.g., almost $ −0.1 for BB and SK) than the actual prices. Because of this, the profit margins are also smaller, i.e., below 20%. This is the consequence of the endogeneity bias since it led to an overestimation of the price sensitivity. Due to considerable price reductions, the market shares of the four brands grow strongly compared to the outside good. The profit of the retailer is approx. $ 9519.5. Although this is a higher profit than in the status quo scenario, the increase is about 3% smaller than in the TVP-MXL(CF) model scenario. If again a whole year is considered, accounting for endogeneity has a value of $ 15657.2.

The normative analysis demonstrates that the optimal prices for the retailer based on the TVP-MXL model as well as the actual prices lead to suboptimal profits. In particular, ignoring price endogeneity costs the retailer a five-digit amount per year. The retailer should, therefore, better employ the TVP-MXL(CF) model for optimal pricing decisions.

[15] Given the small differences between the actual and the optimized prices (based on the TVP-MXL(CF) model) on average, we conclude that the optimization would leave the retailer's current price image unchanged (Weber, Steiner, & Lang, 2017).

Before closing this section, we want to highlight the applicability of the proposed approach.[16] As we rely only on information that should be readily available to retailers (i.e., weekly aggregated sales, prices, and costs at the brand-level) the estimation should be easy. Because of the time-varying parameters, the model is able to address changes in the data and this should foster the trust of practitioners in the results even if they are derived from past data (i.e., have the best inference at time T given the information before T). Assuming that the constant parameters remain stable, the retailer does not even have to re-estimate the whole model if new data arrives (e.g., every week), only the time-varying parameters have to be updated using the KF and KS recursions. New optimal prices can be obtained by running the optimization for the new periods and because this step is easy, simple, and fast the effort of implementing the proposed approach in a real-life scenario for retailers carrying many categories should be fairly straightforward.

6. Summary and conclusions

The present paper covers logit models for aggregate sales data. In addition to consumer heterogeneity and price endogeneity, these models allow the temporal variation of all parameters. Accounting for dynamics is particularly important in marketing (Leeflang et al., 2009); therefore, the assumption of constant utility functions over time is questionable (Lachaab et al., 2006). Since prices in market data are often endogenous (Villas-Boas & Winer, 1999), it seems appropriate to explicitly consider both phenomena (dynamics and endogeneity) when modeling aggregate demand.

If an endogenous variable has a time-varying effect, standard methods (2SLS, GMM, etc.) do not lead to valid results. As a suitable solution to the endogeneity problem, a CF approach is suggested and examined comprehensively. The key findings from a simulation study and an empirical application are as follows. (1) The proposed CF approach is able to estimate time-varying parameters consistently for aggregate logit models while taking into account heterogeneity as well as endogeneity. (2) Biases might be

[16] We thank a reviewer for the idea of including this discussion.

smaller when endogeneity is ignored compared to treating it incorrectly. (3) A wrong endogeneity correction causes biases in all dynamic effects. These biases affect not only the absolute parameter values at a given point in time but also the temporal paths. (4) We demonstrate the relevance of a logit model with heterogeneity, endogeneity, and parameter dynamics empirically for data from the retailer DFF in the product category canned tuna. (5) For DFF the (appropriate) price optimization has a profit potential of around \$ 80000 per year. (6) If the price optimization ignores endogeneity (TVP-MXL model), biases in the estimated price sensitivities lead to suboptimal retail prices and yearly profits are more than \$ 15600 higher (+3.1 %) based on the TVP-MXL(CF) model.

The approach postulated here also has limitations and there are several extensions for future research. (1) A specification with time-varying parameters in the price model to make cost effects more flexible would be promising (Kim & Kim, 2011). (2) The model ignores quantity choice and a corresponding extension would be useful (Nair, Dubé, & Chintagunta, 2005). (3) The MSL estimator is not well suited for large models with many brands and explanatory variables. In such cases, a Bayesian approach would be more promising (Jiang et al., 2009; Petris et al., 2009). (4) All time-varying parameters have been modeled as random walk processes, but alternative dynamic models are also possible. In case prior knowledge and data are available, a reparameterization of the time-varying parameters would be useful to understand long-term marketing-mix effects. (5) In case no suitable instruments are available an instrument-free approach would be helpful. For this, the copula approach of Park and Gupta (2012b) in combination with the KF appears to be a fruitful direction for future research.

In sum, we conclude that the proposed MXL-TVP(CF) model is well suited to model consumer (brand) choice taking heterogeneity, endogeneity, and in particular dynamics into account. Because the model only uses aggregate data, it is very useful for demand estimation and price optimization at the retailer level.

Acknowledgments

The constructive comments of three anonymous reviewers are gratefully acknowledged. I also thank Winfried Steiner and Hartwig Steffenhagen for guidance and support, and Juliet Reiss and Sebastian Gabel for detailed feedback on earlier versions of this paper. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 is gratefully acknowledged.

References

Ackerberg, D., Benkard, C. L., Berry, S., & Pakes, A. (2007). Econometric tools for analyzing market outcomes. In J. J. Heckman, & E. E. Leamer (Eds.), *Handbook of econometrics Vol. 6A, North-Holland, Amsterdam* (pp. 4172–4276).

Allenby, G. M., & Rossi, P. E. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics, 89*(1/2), 57–78.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics, 25*(2), 242–262.

Berry, S. T., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica, 63*(4), 841–890.

Besanko, D., Gupta, S., & Jain, D. (1998). Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Science, 44*(11), 1533–1547.

Bodapati, A. V., & Gupta, S. (2004). The recoverability of segmentation structure from store-level aggregate data. *Journal of Marketing Research, 41*(3), 351–364.

Casado, E., & Ferrer, J.-C. (2013). Consumer price sensitivity in the retail industry: Latitude of acceptance with heterogeneous demand. *European Journal of Operational Research, 228*(2), 418–426.

Chevalier, J. A., Kashyap, A., & Rossi, P. E. (2003). Why don't price rise during periods of peak demand? evidence from scanner data. *American Economic Review, 93*(1), 15–37.

Chintagunta, P. K. (2001). Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Science, 20*(4), 442–456.

Chintagunta, P. K. (2002). Investigating category pricing behavior at a retail chain. *Journal of Marketing Research, 39*(2), 141–154.

Chintagunta, P. K., Dubé, J.-P., & Goh, K. Y. (2005). Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models. *Management Science, 51*(5), 832–849.

Chintagunta, P. K., Dubé, J.-P., & Singh, V. (2002). Market structure across stores: An application of a random coefficients logit model with store level data. In P. H. Franses, & A. L. Montgomery (Eds.), *Econometric models in marketing: 16* (pp. 191–221). Amsterdam: Elsevier.

Chintagunta, P. K., Dubé, J.-P., & Singh, V. (2003). Balancing profitability and customer welfare in a supermarket chain. *Quantitative Marketing and Economics, 1*(1), 111–147.

Chintagunta, P. K., Jain, D. C., & Vilcassim, N. J. (1991). Investigating heterogeneity in brand preferences in logit models for panel data. *Journal of Marketing Research, 28*(4), 417–493.

Chung, K. Y. C., Derdenger, T. P., & Srinivasan, K. (2013). Economic value of celebrity endorsements: Tiger woods' impact on sales of nike golf balls. *Marketing Science, 32*(2), 271–293.

Cook, S. R., Gelman, A., & Rubin, D. R. (2006). Validation of software for Bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics, 15*(3), 675–692.

Draganska, M., & Klapper, D. (2007). Retail environment and manufacturer competitive intensity. *Journal of Retailing, 83*(2), 183–198.

Dubé, J.-P., & Chintagunta, P. J. (2003). Comment (on Bayesian analysis of simultaneous demand and supply). *Quantitative Marketing and Economics, 1*(3), 293–298.

Dubé, J.-P., Chintagunta, P. K., Petrin, A., Bronnenberg, B., Goettler, R., Seetharaman, P. B., et al. (2002). Structural applications of the discrete choice model. *Marketing Letters, 13*(3), 207–220.

Dubé, J.-P., Fox, J. T., & Su, C.-L. (2012). Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica, 80*(5), 2231–2267.

Dubé, J.-P., Hitsch, G. J., Rossi, P. E., & Vitorino, M. A. (2008). Category pricing with state-dependent utility. *Marketing Science, 27*(3), 417–429.

Erdem, T., Keane, M. P., & Sun, B. (2008). The impact of advertising on consumer price sensitivity in experience goods markets. *Quantitative Marketing and Economics, 6*(2), 139–176.

Foekens, E. W., Leeflang, P. S. H., & Wittink, D. R. (1999). Varying parameter models to accommodate dynamic promotion effects. *Journal of Econometrics, 89*(1–2), 249–268.

Fong, N. M., Simester, D. I., & Anderson, E. T. (2011). Private label vs. national brand price sensitivity: Evaluating non-experimental identification strategies. *Technical Report*. MIT, Sloan School of Management.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Greene, W. H. (2011). *Econometric analysis* (7th ed). Pearson, Edinburgh Gate.

Guadagni, P. M., & Little, J. D. C. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science, 2*(3), 203–238.

Gupta, S., Chintagunta, P. K., Kaul, A., & Wittink, D. R. (1996). Do household scanner data provide representative inferences from brand choices: A comparison with store data. *Journal of Marketing Research, 33*(4), 383–398.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica, 46*(6), 1251–1271.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica, 46*(4), 931–959.

Hruschka, H. (2010). Considering endogeneity for optimal catalog allocation in direct marketing. *European Journal of Operational Research, 206*(1), 239–247.

Hruschka, H. (2017). Functional flexibility, latent heterogeneity and endogeneity in aggregate market response models. *Marketing ZFP – Journal of Research and Management, 39*(3), 17–31.

Jedidi, K., Mela, C. F., & Gupta, S. (1999). Managing advertising and promotion for long-run profitability. *Marketing Science, 18*(1), 1–22.

Jiang, R., Manchanda, P., & Rossi, P. E. (2009). Bayesian analysis of random coefficient logit models using aggregate data. *Journal of Econometrics, 149*(2), 136–148.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering, 82*(Series D), 35–45.

Kalouptsidis, N., & Psaraki, V. (2010). Approximations of choice probabilities in mixed logit models. *European Journal of Operational Research, 200*(2), 529–535.

Keane, M. P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics, 15*(3), 310–327.

Kim, B.-D., Blattberg, R. C., & Rossi, P. E. (1995). Modeling the distribution of price sensitivity and implications for optimal retail pricing. *Journal of Business & Economic Statistics, 13*(3), 291–303.

Kim, C.-J. (2006). Time-varying parameter models with endogenous regressors. *Economics Letters, 91*(1), 21–26.

Kim, C.-J. (2010). Dealing with endogeneity in regression models with dynamic coefficients. *Foundations and Trends in Econometrics, 3*(3), 165–266.

Kim, Y., & Kim, C.-J. (2011). Dealing with endogeneity in a time-varying parameter model: joint estimation and two-step estimation procedures. *Econometrics Journal, 14*, 487–497.

Kiviet, J. F. (2011). Monte carlo simulation for econometricians. *Foundations and Trends in Econometrics, 5*(1/2), 1–181.

Kök, A. G., & Fisher, M. L. (2007). Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research, 55*(6), 1001–1021.

Kopalle, P. K., Mela, C. F., & Marsh, L. (1999). The dynamic effect of discounting on sales: Empirical analysis and normative pricing implications. *Markeitng Science, 18*(3), 317–332.

Lachaab, M., Ansari, A., Jedidi, K., & Trabelsi, A. (2006). Modeling preference evolution in discrete choice models: A Bayesian state-space approach. *Quantitative Marketing and Economics, 4*(3), 57–81.
Leeflang, P. S. H., Bijmolt, T. H., van Doorn, J., Hanssens, D. M., van Heerde, H. J., Verhoef, P. C., & Wieringa, J. E. (2009). Creating lift versus building the base: Current trends in marketing dynamics. *International Journal of Research in Marketing, 26*(1), 13–20.
Mela, C. F., Gupta, S., & Lehmann, D. R. (1997). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research, 34*(2), 248–261.
Meza, S., & Sudhir, K. (2006). Pass-through timing. *Quantitative Marketing and Economics, 4*(4), 351–382.
Morrow, W. R., & Skerlos, S. J. (2011). Fixed-point approaches to computing Bertrand–Nash equilibrium prices under mixed-logit demand. *Operations Research, 59*(2), 444–460.
Nair, H., Dubé, J.-P., & Chintagunta, P. K. (2005). Accounting for primary and secondary demand effects with aggregate data. *Marketing Science, 24*(3), 444–460.
Nevo, A. (2000). A practitioners guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy, 9*(4), 513–548.
Nijs, V., Misra, K. K., Anderson, E. T., Hansen, K., & Krishnamurthi, L. (2010). Channel pass-through of trade promotions. *Marketing Science, 29*(2), 250–267.
Nijs, V. R., Srinivasan, S., & Pauwels, K. (2007). Retail-price drivers and retailer profits. *Marketing Science, 26*(4), 473–487.
Pagan, A. (1984). Econometric issues in the analysis of regressions with generated regressors. *International Economic Review, 25*(1), 221–247.
Park, S., & Gupta, S. (2009). Simulated maximum likelihood estimator for the random coefficient logit model using aggregate data. *Journal of Marketing Research, 46*(4), 531–542.
Park, S., & Gupta, S. (2012a). Comparison of SML and GMM estimators for the random coefficient logit model using aggregate data. *Empirical Economics, 43*(3), 1353–1372.
Park, S., & Gupta, S. (2012b). Handling endogenous regressors by joint estimation using copulas. *Marketing Science, 31*(4), 567–586.
Pennings, C. L., & van Dalen, J. (2017). Integrated hierarchical forecasting. *European Journal of Operational Research, 263*(2), 412–418.
Petrin, A., & Train, K. E. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research, 47*(1), 3–13.
Petris, G., Petrone, S., & Campagnoli, P. (2009). *Dynamic linear models with R.* New York: Springer.
Reiss, P. C., & Wolak, F. A. (2007). Structural econometric modeling: Rationales and examples from industrial organization. In J. J. Heckman, & E. E. Leamer (Eds.), *Handbook of econometrics: 6A*. North-Holland, Amsterdam: Elsevier. 4280–4412.
Rossi, P. E. (2014). Invited paper–even the rich can make themselves poor: A critical examination of iv methods in marketing applications. *Marketing Science, 33*(5), 655–672.
Schweppe, F. C. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transaction on Information Theory, 11*(1), 61–70.
Shugan, S. M. (2002). In search of data: An editorial. *Marketing Science, 21*(4), 1–3.
Sriram, S., Balachander, S., & Kalwani, M. U. (2007). Monitoring the dynamics of brand equity using store-level data. *Journal of Marketing, 71*(2), 61–78.
Sriram, S., Chintagunta, P. K., & Neelammegham, R. (2006). Effects of brand preference, product attributes, and marketing mix variables in technology product markets. *Marketing Science, 25*(5), 440456.
Sriram, S., & Kadiyali, V. (2009). Empirical investigation of channel reactions to brand introductions. *International Journal of Research in Marketing, 26*(4), 345–355.
Sriram, S., & Kalwani, M. U. (2007). Optimal advertising and promotion budgets in dynamic markets with brand equity as a mediating variables. *Management Science, 53*(1), 46–60.
Train, K. E. (2009). *Discrete choice methods with simulation* (2nd ed). New York: Cambridge University Press.
Van Heerde, H. J., Mela, C. F., & Manchanda, P. (2004). The dynamic effect of innovation on market structure. *Journal of Marketing Research, 41*(2), 166–183.
Van Heerde, H. J., & Neslin, S. A. (2008). Sales promotion models. In B. Wierenga (Ed.), *Handbook of marketing decision models* (pp. 107–162). New York: Springer.
Vilcassim, N. J., & Chintagunta, P. K. (1996). Investigating retailer product category pricing from household scanner panel data. *Journal of Retailing, 71*(2), 103–128.
Villas-Boas, J. M., & Winer, R. S. (1999). Endogeneity in brand choice models. *Management Science, 45*(10), 1324–1338.
Weber, A., Steiner, W. J., & Lang, S. (2017). A comparison of semiparametric and heterogeneous store sales models for optimal category pricing. *OR Spectrum, 39*(2), 403–445.
Winer, R. S. (1986). A reference price model of brand choice for frequently purchased products. *Journal of Consumer Research, 13*(2), 250–256.
Yang, S., Chen, Y., & Allenby, G. M. (2003). Bayesian analysis of simultaneous demand and supply. *Quantitative Marketing and Economics, 1*(3), 251–275.