Working Paper

Explaining Dynamic Heterogeneity in Brand Choice through Cross-Category Factors

Ryan Dew, Yuhao Fan*

December 5, 2023

Abstract

Understanding individual customers' sensitivities to prices, promotions, brand, and other aspects of the marketing mix is fundamental to a wide swath of marketing problems, including targeting and pricing. An important aspect of how consumers behave is dynamics: preferences are not stable over time, and individual-level preference parameters often evolve heterogeneously. Prior work has developed methods for capturing dynamic heterogeneity independently across product categories, but ignored the possibility of *correlated* dynamics *across* categories. Preference dynamics in one category may be indicative of changes in other categories, especially if those changes are driven by external factors. In this work, we propose a framework for capturing such correlated dynamics by means of a Bayesian nonparametric dynamic latent factor model, wherein individual-level preference parameters are modeled as weighted combinations of a common set of cross-category latent factors, and nonparametric deviations from those weighted factors. The resulting specification effectively decomposes variation in individual-level dynamics into latent cross-category trends and individual-specific trajectories. We apply our model to grocery purchase data, with two main results: first, we show that a surprising degree of dynamic heterogeneity can be explained by only a few latent trends, and attempt to characterize those trends. Second, we characterize the exact correlations that exist across parameter dynamics, many of which would be difficult to predict a priori. Managerially, the proposed framework not only provides insight into how consumer preferences are changing across product categories, but improves predictive ability by means of its ability to effectively leverage cross-category data.

Keywords: dynamic heterogeneity, Bayesian nonparametrics, Gaussian processes, choice models

*Ryan Dew is the corresponding author (`ryandew@wharton.upenn.edu`). He is an Assistant Professor of Marketing at the Wharton School, University of Pennsylvania. Yuhao Fan is a Data Scientist at URBN. Special thanks to Kevin Liu and Nikhil Kona for excellent research assistance .

1. Introduction

We know that customers' sensitivities to the marketing mix are not static. Dynamics in sensitivities can stem from a number of things, including changes in people's personal lives, consumer learning, and fluctuations in the macroeconomic environment. These changes may lead to shifts in consumers' sensitivities to marketing mix variables and brand preferences. For example, a customer who has recently graduated from college and started a job may become less price sensitive or less promotion sensitive, as a result of an increase in her budget constraint. She may also begin exploring more brands, resulting in a shift in her brand preferences toward more premium brands. These changes can have important implications for marketers, as understanding these shifts in preferences can help marketers optimize their targeting strategies and their allocation of marketing mix variables. Yet, while many of these dynamics can be rationalized using standard economic models of utility maximization under constraint or consumer learning, it is extremely challenging to account for such dynamics in the indirect utility models commonly used by marketers to optimize the marketing mix.

This challenge has led to recent methodological work in marketing on dynamic heterogeneity specifications that allow for modeling individual-level dynamics in preference parameters (Dew et al., 2020), including in brand choice applications. Yet, while that work solves the statistical problem of estimating a time-varying distribution of unobserved heterogeneity, it ignores the potentially important correlations that may exist across product categories. As illustrated by the simple example above, changes in consumer's budget or economic circumstances may impact many categories at once, inducing correlations in the dynamics of brand choice parameters. Such correlations can be extremely valuable to marketers: in many shopping contexts, including in both traditional consumer packaged goods and modern online settings, marketers may observe an abundance of information about a customer in one product category, but no or relatively little information in another. Understanding to what degree a shift in a consumer's preferences and sensitivities to marketing mix variables persists across product categories, or is idiosyncratic to a single product category, can help marketers leverage data in data-rich categories to make predictions about data-scarce categories. Furthermore, by leveraging this cross-category information, marketers can better target customers in categories in which they have not yet purchased. Yet, despite the obvious value of understanding cross-category preference dynamics, the marketing and choice modeling literature has provided only limited solutions to estimating these dynamics from data.

In this paper, we develop a Bayesian semiparametric methodology for estimating individual-level, dynamic, correlated model parameters, which we apply in the context of an indirect utility model for brand choice. Our model leverages latent, multi-output Gaussian processes to capture the dynamics of customers' preference parameters in a parsimonious way, that allows for sharing of statistical information both across individuals, and within individuals across time and across product categories. We achieve this by fusing work in marketing on cross-category choice and dynamic heterogeneity with work in machine learning on semiparametric dynamic factor models (Teh et al., 2005) and multi-output Gaussian processes, resulting in a novel specification that contributes to both the marketing and machine learning literatures. We term our specification Multi-Category Dynamic Heterogeneity (MCDH). While Bayesian nonparametrics have seen increased use in marketing, including Gaussian processes, ours is the first in the field to adapt multi-output Gaussian processes.

To illustrate the potential gains from capturing correlated, individual-level preference dynamics, we apply our MCDH model to both simulated and real consumer choice data. Foremost, we show that our specification is identified, and can be easily applied to typical marketing mix problems, in a relatively scalable fashion. But more importantly, we also show MCDH easily outperforms both classic and modern benchmarks in out-of-sample tasks, including the recently proposed GPDH framework, which is a sophisticated, Bayesian nonparametric model that also captures dynamic heterogeneity. Substantively, we show that correlated dynamics matter in two important ways: first, we show that the model is indeed able to leverage cross-category information to yield more reliable preference estimates, which can be especially beneficial when estimating marketing mix sensitivity data in relatively sparser product categories. Moreover, we show that these statistical gains have practical consequences, in terms of the precision of estimating important quantities like price elasticity. Finally, we show that the dynamics uncovered from the model reveal interesting insights about potential drivers of dynamics, including large, macro-level events like the 2008 great recession.

We organize the rest paper as follows. In Section 2, we review relevant literature to our research. We develop our MCDH model in Section 3 and implement the model through simulation in Section 4. In Section 5, we describe the brand choice data used in our application. In Section 6, we describe the high-level results of our application, and in Section 7, we discuss the forecasting performance of our model compared to a comprehensive set of benchmarks. In Section 8, we show that the MCDH model gives more reliable estimates of price elasticities. Finally, we conclude and discuss future research in Section 9.

2. Literature

The cross-category nature of preferences has been extensively examined in marketing, in two main streams of research. The first stream focuses on the correlation of customers' preference parameters across categories. The first in this stream is Ainslie & Rossi (1998), who investigate whether households have similar sensitivities to price and feature promotion across categories. They use a variance components decomposition approach to model households' sensitivities to these variables across categories, and find large cross-category correlations in sensitivities to price (0.32) and to feature promotion (0.58). Seetharaman et al. (1999) further extend the work and show that households' state dependence behaviors also show large correlation across categories. Hansen et al. (2006) build on this framework and estimate customers' preference for store brands across categories. They find that customers who prefer store brands in one category are more likely to prefer store brands in other categories. This stream of work suggests that customers behave similarly across categories in many respects, and that information about preference parameters in one category can be leveraged to understand preferences in another.

The second stream of research investigates complementarity and substitutability between categories. Manchanda et al. (1999) models multicategory purchase incidence and examines the cross-effects that changing prices in one category may have on purchasing in another category. However, their work only models category purchase incidence, and does not model the brand choices within each category. Building on that, Mehta (2007) models both category purchase incidence and brand choice decisions, and Song & Chintagunta (2007) integrate purchase quantity decisions into the framework. Lee et al. (2013) further build on the previous work and allow the cross-effects between categories to be asymmetric.

Our research is aligned more closely with the first stream of research. We investigate the cross-category dynamics of customers' preference parameters. A key shortcoming of the extant models in this stream of research is that they assume customers' preferences are static. Yet, we know from more recent work that preferences are dynamic, and that there are important cross-category linkages. For example, Gordon et al. (2013) investigate the relationship between price elasticity and the business cycle. They find that on average, customers are more price sensitive during economic downturns, with a few notable exceptions, driven primarily by these categories' insignificant shares of wallet. While this work establishes the importance of capturing dynamics in preferences, their specification of heterogeneity is quite restrictive: while consumers are assumed to differ in their initial sensitivities to marketing variables, the way these sensitivities evolve over time is assumed to be the same across people.

In the literature on explicitly modeling customer preference dynamics, most authors have focused on single-category dynamics. Many papers in this line of research have included time-varying individual-level parameters, including Kim et al. (2005), Liechty et al. (2005), Sriram et al. (2006), Lachaab et al. (2006), and Guhl et al. (2018). However, while these papers adopt different specifications for the individual-level time-varying parameters, they all impose a common restriction on heterogeneity: they restrict an individual-level parameter's deviation from the population mean to stay the same across different time periods, such that a preference paraemter β_{it} is modeled as $\beta_{it} = \mu_t + \theta_i$. To account for flexible heterogeneous evolution of such parameters, Dew et al. (2020) introduce the idea of dynamic heterogeneity, or a continuously evolving distribution of unobserved heterogeneity. They model dynamic heterogeneity using Gaussian processes, in what they term the Gaussian Process Dynamic Heterogeneity (GPDH) specification, which allows individual-level parameters to differ from the population mean parameter at different time periods in a parsimonious, hierarchical Bayesian fashion. They show that the GPDH model nests the traditional random coefficient model, yields more accurate and statistically efficient population and individual-level estimates, and performs better in terms of fit and forecasting. Moreover, the authors find that ignoring heterogeneity in dynamics may bias parameter and elasticity estimates.

Building on this body of work, we develop a cross-category, individual-level, Bayesian nonparametric model that captures the dynamics of customers' preference parameters in a given category, as well as the cross-category nature of the dynamics in preferences. By building a flexible, cross-category dynamic preference model, we allow information learned about how customers' preference parameters have evolved in one category to be transferred across categories, and inform predictions of preference evolution in another category by sharing common latent factors that affect all of customers' preference parameters. Such information sharing allows us to forecast customer's brand choice more accurately and produces more reliable estimates of price elasticities.

3. Model

In this section, we briefly introduce required background knowledge about Gaussian processes, and then describe our modeling framework in detail.

3.1. Gaussian process

A Gaussian process (GP) is a stochastic process $f(.)$ defined on an input space, which, in our case, is time $t \in R^+$. A GP is defined by a mean function $m(t)$ and a covariance function $k(t, t')$, such that for a fixed set of inputs $\boldsymbol{t} = \{t_1, t_2, \ldots, t_T\}$,

$$f(\boldsymbol{t}) \sim \mathcal{N}(m(\boldsymbol{t}), K(\boldsymbol{t})),$$

where $m(\boldsymbol{t})$ is the mean function evaluated at all inputs, a $T \times 1$ vector, and $K(\boldsymbol{t})$ is a $T \times T$ matrix formed by evaluating the covariance function $k(t, t')$ pairwise at all of the inputs. In short, the mean function specifies the prior expectation of the value of the process for each input t, and the covariance function specifies how correlated the process is across pairs of inputs, t and t'. Since a GP defines a probability distribution over outputs, given any inputs, it serves as a natural, nonparametric prior over function spaces (Rasmussen & Williams, 2006). In this capacity, is typically denoted $f(t) \sim \mathcal{GP}(m(t), k(t, t'))$.

In most GP applications, mean functions play a limited role, and are usually assumed to be constant, allowing the covariance function to capture the properties of the functions generated by the GP priors. Different covariance functions, also called kernels, can capture different broad features of the functions being modeled, including smoothness, differentiability, and amplitude. A valid kernel is a function $k : \mathbb{R}^2 \to \mathbb{R}$, such that the covariance matrix $K(\boldsymbol{t})$ generated by the kernel function is positive semidefinite for any set of inputs $\boldsymbol{t}$. Many different kernels have been proposed in the GP literature, the simplest and most popular of which is the squared exponential kernel, which is the kernel we will use in this paper. The squared exponential kernel function is defined by two hyperparameters, an amplitude parameter σ and a length-scale parameter ρ, with functional form:

$$k_{\mathrm{SE}}(t, t'|\sigma, \rho) = \sigma^2 \exp\left[-\frac{(t-t')^2}{2\rho^2}\right].$$

The amplitude parameter σ governs how far the function draws from a GP with this kernel can be from the mean function, while the length-scale parameter ρ captures the smoothness of the departures from the mean function, and is sometimes called simply the smoothness parameter. The simple but expressive nature of the SE kernel has led to its widespread use across many application areas, including in prior marketing studies (Dew et al., 2020; Dew & Ansari, 2018).

3.2. Multi-Category Dynamic Heterogeneity (MCDH)

Having described Gaussian processes, we now describe our brand choice model with cross-category dynamic heterogeneity. Following the standard random (indirect) utility specification, we assume that the utility that the customer i gets from purchasing product j in category c at time t is:

$$u_{icjt} = \sum_{p=1}^{P} \beta_{icpt} \times x_{pcjt} + \epsilon_{icjt},$$

where ϵ_{icjt} is independently and identically distributed as an extreme value distribution, leading to the traditional multinomial logit model for choice probabilities. x_{pcjt} can include marketing mix variables such as prices and promotions as well as brand level dummy variables. β_{icpt}

thus measures customer i's sensitivity to variable p in category c at time t. For identification purposes, we normalize the brand dummy variable of the brand with highest share of each category to be 0 across all time periods for all individuals.

For ease of notation, we collapse all the brand loyalty terms and preference sensitivity terms for a given customer, β_{icpt}, into a time-varying vector β_{it}, which is size $K \times 1$. K enumerates over each category c and each preference parameter p. Our goal is thus to come up with a parsimonious specification for β_{it} that can be feasibly estimated from a typical marketing mix data. To achieve this goal, we first make a conceptual pivot, and consider β_{it} as a vector-valued *function* of time, $\beta_i(t)$. We then model this vector-valued function using a *multi-output* Gaussian process (MOGP). Briefly, just as GPs provide a principled way of placing a prior over real-valued functions, multi-output GPs provide a means of placing a prior over vector-valued functions, where the goal is to model a multivariate output simultaneously by exploiting the correlations across the multivariate output, thus outperforming modeling each element of the output independently (Álvarez et al., 2012).

The specific MOGP formulation we adapt to model $\beta_i(t)$ is based on the the popular linear model of coregionalization (Schmidt & Gelfand, 2003) and the semiparametric latent factor model introduced by (Teh et al., 2005). Specifically, we model the k^{th} element of $\beta_i(t)$, $\beta_{ik}(t)$, as the sum of L independent, latent, time-varying functions, such that

$$\beta_{ik}(t) = \alpha_k + \sum_{l=1}^{L} \omega_{ikl} \times u_l(t), \tag{1}$$

where α_k is a parameter that captures the population, time-invariant mean of the k^{th} element in the $\beta_i(t)$, and the $u_l(t)$ are latent functions drawn from $\mathcal{GP}(0, k_{\mathrm{SE}}(t, t'|1, \rho_l))$ which are shared across people.[1] The ω_{ikl} serve as a individual-specific weights on each of the latent factors. We assume $\omega_{il} = (\omega_{i1l}, \ldots, \omega_{iKl}) \sim \mathcal{N}(0, \Sigma_\omega)$, where Σ_ω is a $K \times K$ unrestricted covariance matrix.

We model Σ_ω as $\Sigma_\omega = \mathrm{diag}(\tau)\Lambda_\omega\mathrm{diag}(\tau)$. Λ_ω is the correlation matrix that measures the dynamic correlations between the preference sensitivities, while the vector τ measures the variation of preference sensitivities across customers. We model $\tau \sim \mathcal{N}^+(0, 1)$ and $\Lambda_\omega \sim LKJ(2)$,

[1] We restrict the amplitude to 1 for identification, since we do not restrict the values of the weight terms ω.

where LKJ(2) stands for the LKJ correlation distribution with shape parameter 2 (Lewandowski et al., 2009), which is a weakly informative prior favoring the identity matrix.

Intuitively, since independent latent functions are convolved through the weights, the result is a set of correlated functions, comprising the components of $\beta_i(t)$. This structure allows us to capture the correlations of brand loyalties of a customer within a category, as well as the price sensitivities across different categories.

There are two key sources of model specification in this setting: the kernel function used in estimating the GPs, and the number of latent functions, L. As noted above, we chose to use the simple and common squared exponential kernel for each of the latent factors, which assumes a priori that these functions are relatively smooth. Other common kernels in the literature, including the Matern class of kernels, make different assumptions, which may be more desirable in different contexts. We tested several classes, and found our model relatively robust to this choice. The more crucial part of model specification is setting L. Intuitively, more latent functions will naturally allow us fit more complex patterns of preference parameters, yet also risks overfitting. Additional factors may capture smooth, predictable time variation, but they may also capture noisy, difficult to generalize patterns in the training data. Thus, to choose L, we will rely on forecasting accuracy, and favor smaller values for a more parsimonious model. We explore this issue more in the next section.

The usage of a few latent functions to capture dynamics across all customers has many advantages. Foremost, it provides a source of dimensionality reduction: instead of modeling each $\beta_{ik}(t)$ as a separate function draw, we let $\beta_{ik}(t)$ be the sum of weighted latent functions that are shared across all customers and all preference parameters. This structure pools information across different customers, and across different points in time. Moreover, we also build information-sharing into the weights: since the weight vector ω_{il} is modeled hierarchically, with a full covariance matrix, it allows for information sharing across preference parameters, including those of different product categories. Moreover, since the covariance is shared across people, it provides another source of pooling. Together, these mechanisms allow us to turn the complex problem of estimating β_{icpt} into a simpler problem, with considerable information sharing across units of analysis. In turn, this yields increased precision in model estimates, as we will demonstrate empirically in later sections.

Finally, from a substantive perspective, modeling the dynamics in preferences through a set of common latent factors has another advantage: if there are common temporal shocks to many consumers' preference parameters, these may be naturally captured in one or several of the latent factors, and consumers' individual weights on those factors may thus indicate to what degree they were affected by those temporal shocks. This is part of the motivation for allowing each of the latent factors to have its own length-scale, ρ_l. Recall that the length-scale parameter determines the smoothness of the function draws. Hence, by allowing each function to have its own length-scale, we intuitively allow the model to detect both long-run trends and short-run shocks to preferences, which may be captured on different factors. While we cannot *guarantee* such findings, if we can map real-world phenomena onto the latent factors, this provides a powerful tool for understanding the dynamics, and the inter-linkages among consumers. Moreover, these latent factors allow us to understand the correlated dynamics that may exist in customers' different preference sensitivities. How much weights a latent factor contributes to different preference sensitivities informs us how these preference sensitivities are correlated dynamically.

3.3. Related Work

Our model relates to three streams of research: multi-output GPs, models of dynamic preferences, and models of dynamic latent factors. In this section, we briefly clarify the links and distinctions between the present work and these other models.

Our multi-output GP approach to modeling correlated, dynamic coefficients is directly inspired by similar models that have been proposed in machine learning and geostatistics. In geostatistics, similar specifications have a long history in cokriging applications, where the goal is to model potentially correlated mineral deposits or other environmental factors across space and time (Wackernagel, 2013). In that literature, the idea of linearly combining independent factors to learn correlated processes is called the linear model of coregionalization (LMC). In machine learning, Teh et al. (2005) propose a similar idea to the LMC in their semiparametric latent factor model (SLFM). Our specification for correlated dynamics closely follows their specification. However, crucially, both in geostatistics and in machine learning, models like the LMC and SLFM are almost exclusively used in regression or classification tasks, where the outcomes of interest are directly observed, and modeled (up to some noise) by the linear combination of GPs. In our application, the task is much more difficult: our “outcome” of interest is, in fact, a model coefficient capturing consumer sensitivities, which are then combined with observed variables (i.e., the marketing mix) to model utilities which determine choice. The jump from modeling the regression function directly to modeling latent quantities like model parameters is non-trivial. The other key difference between our work and these more typical applications is scale: while most geostatistical and machine applications consider a handful of related processes, our model focuses on modeling correlated preference parameters for *many* consumers, each of which has *many* sensitivities. To handle this scale, we introduce the hierarchical structure on the weights, that allows for correlated patterns within individuals, and sharing of information across individuals.

In marketing, there are two areas of research that are directly relevant to ours, among which Dew et al. (2020)‘s recent work is perhaps the most relevant piece. That paper introduces the idea of dynamic heterogeneity, and illustrates the potential biases in model estimates and decision quantities, including price elasticities, that can arise when dynamics in the distribution of cross-sectional heterogeneity are ignored. Our work provides an alternative method for modeling such dynamic heterogeneity, that also allows for correlations across preference parameters. Both Dew et al. (2020) and our work use nonparametric functions to model dynamic heterogeneity of customers’ preference sensitivities, but Dew et al. (2020) treats dynamics of preference sensitivities in different product categories independently. On the contrary, we explicitly model the correlation of preference sensitivities across categories and allow information sharing of preference sensitivities within and across product categories. We show the gains by our approach in a direct comparison in later sections. Our work is also connected to research in marketing using dynamic latent factor models, including early work by Du & Kamakura (2012). In their work, Du & Kamakura (2012) decompose an observed collection of time series of search volume for a set of brands into a linear combination of dynamic latent factors. The MCDH model we propose is different from the work by Du & Kamakura (2012) in three important aspects. The first two distinctions closely mirror those described above: our model adapts the ideas of correlated latent factors to model sensitivities, not a directly observed time series, and our model does this for a massive number of model parameters, spanning customers, categories, and preference parameters. Notably, our model specification has another important distinction from theirs: while they model their latent factors using a relatively complex state space model, our model uses a small number of relatively simple GPs, which provide increased pooling across units of time.

3.4. Estimation

We estimate our MCDH model in a fully Bayesian fashion, assigning weakly informative prior distributions for each hyperparameter not yet specified. We perform inference using Hamiltonian Monte Carlo (HMC), via the No-U-Turn sampler (NUTS) in Stan, leveraging recently introduced parallel computing methods for scalability (Betancourt, 2018). We jointly sample all the parameters in the model, including the mean functions, the hyperparameters, and the individual-level functions. The joint density of the data y and the full set of model parameters, $\Theta = \{\beta, \mu, \eta, \gamma, \omega, \theta, \sigma, \rho\}$, is:

$$
\begin{aligned}
p(y, \Theta | X) = &\prod_{n=1}^{N} p(y_n | X_n, \{\beta_{i_n c_n p}(t)\}_{p=1}^{P}) \times \\
&\prod_{i=1}^{I} \prod_{k=1}^{K} \prod_{l=1}^{L} p(\beta_{ik}(t) | \alpha_k, \omega_{ikl}, u_l(t)) p(\omega_{ikl} | 0, \Sigma_\omega) p(u_l(t) | 1, \rho_l) \times \\
&p(\Sigma_\omega) p(\rho) p(\alpha_k),
\end{aligned}
$$

where $n = 1, \ldots, N$ denotes the n^{th} observation in the data, X_n indicates the marketing mix variables and brand dummy variables observed in the n^{th} observation, i_n indicates the customer ID of the n^{th} observation, and c_n indicates the product category of the n^{th} observation.

4. Simulation

In this section, we explore two aspects of our model using simulations. First, we explore the flexibility of simulated preference sensitivities of the MCDH model, using just a few number

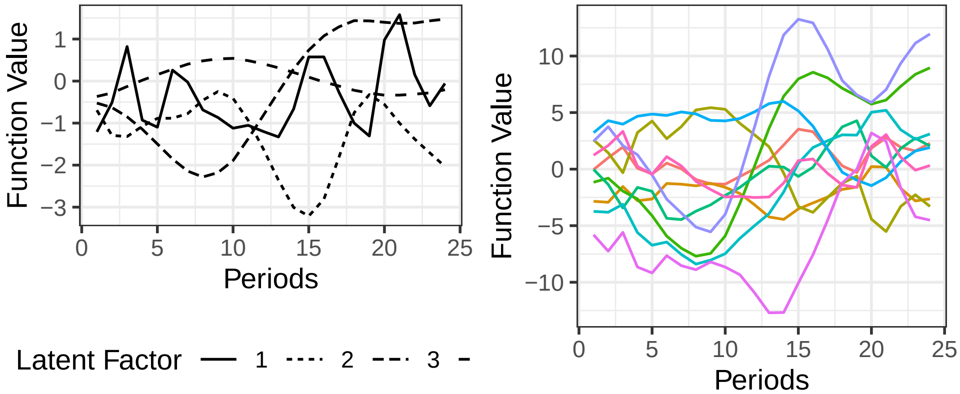

Figure 1: Simulation of four latent factors and the preference parameters generated by them. Latent factors are plotted on the left, and each line type represents a different latent factor. Preference parameters are plotted on the right, and each color represents a different preference parameter.

of latent factors. Then, we simulate brand choice data using the MCDH model to explore the identification of the model.

One may think that we need a large number of latent factors in order to capture flexible preference dynamics. Here we illustrate that only using four latent factors already allows us to capture very flexible dynamics of preference parameters. We use four latent factors that are drawn from GP, such that $u_l(t) \sim \mathcal{GP}(0, k_{\text{SE}}(t, t'|1, \rho_l))$, where ρ_l takes value $1, 2, 4, 8$. We fix α_k at 0 and draw ω_{ilk} independently from $\mathcal{N}(0, 2)$. We then compute $\beta_{ik}(t)$ from these simulated values based on equation 1.

In Figure 1, we plot the latent factors that we simulate on the left and the ten random sampled preference parameters using four latent factors on the right. We can see that the ten sampled preference parameters represent very different patterns of dynamics. Despite the few number of latent factors, the model achieves such flexibility through the amplifier ω_{ikl}. If the absolute value of ω_{ikl} is the largest, u_l plays a bigger role compared to the rest of the latent factors. In Figure 1, we see that preference dynamics are very different in terms of smoothness, depending on which u_l is playing a bigger role. Moreover, ω_{ikl} can take on negative values, so even if two preference dynamics have similar shapes, they can develop in opposite direction. Therefore, just a few latent factors can yield great flexibility for preference parameters in the

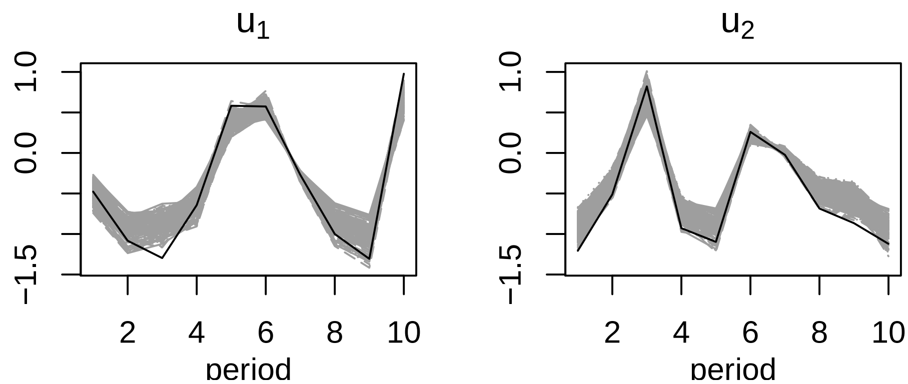

Figure 2: Simulation recovery of the dynamic latent factors. The shaded bands are post burn-in posterior draws and the solid lines are the function values used in simulation. The x-axis marks the time and y-axis marks the function values.

MCDH model.

We also explore the identification and properties of the model through a simulation and recovery exercise. We simulate a data set where 100 customers repeatedly choose among different 6 brands in each of the 5 categories. We assume each customer chooses 20 times in each period for a total of 10 periods. Besides the brand dummies, we assume that there is only one observable variable that affects customers' brand choices. We estimate our MCDH model with this simulated data. We plot the recovery of the latent factors as well as the randomly sampled customers' preferences sensitivities in the Figure 2 and Figure 3. It is worth noting that the latent factors $u_l(t)$ are identified up to the sign. However, this does not affect the identification of the $\beta_{ik}(t)$, which is the most important aspect of the model.

The hyperparamters of the GPs that generate $u_l(t)$ are all very well recovered, and the true values 1 fall around the modes of the posterior distribution, which ranges between (0.6,1.3). In Figure 2, we examine the latent function recovered. We can see that the estimated the function values track the values of the generating process very closely. In Figure 3, we plot the customer-level dynamics from 9 randomly sampled customers. The estimated function values still track the true values very well, but the post burn-in posterior draws are now much wider. This is expected as all choice data contribute to identifying the latent factors, while a customer's own choice data across categories help to identify the customer-level dynamics. The latent factors

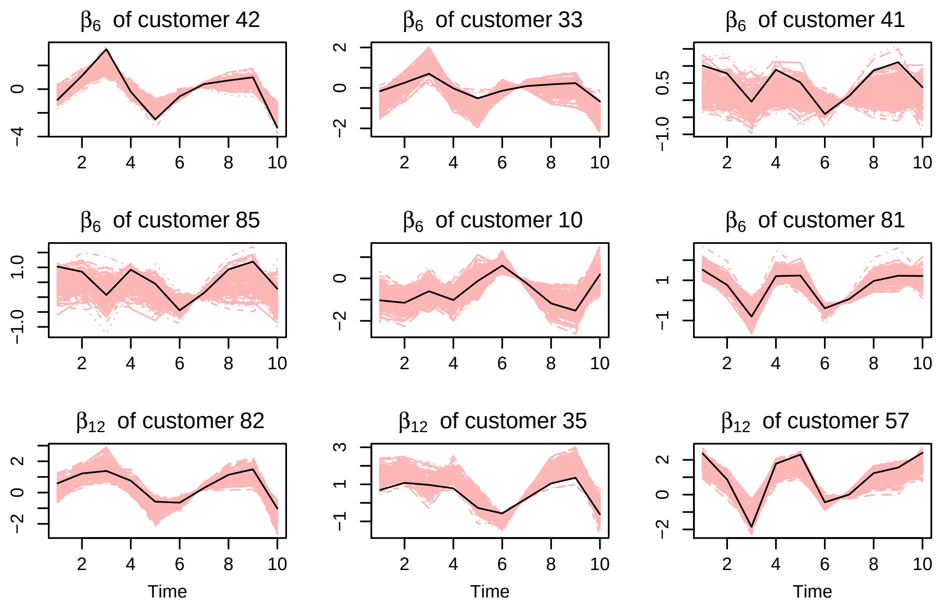

Figure 3: Simulation recovery of the cross-category dynamics of customer-level preference sensitivities. Each panel plots the dynamics of a preference sensitivity of a randomly sampled customer. The shaded bands are post burn-in posterior draws and the solid lines are the function values generated in simulation. The x-axis marks the time and y-axis marks the function values.

are more precisely recovered than the individual-level cross-category dynamics, because of the relatively more abundant data across customers.

5. Data

We apply the MCDH model to brand choice in the IRI consumer packaged good panel data, from January 2001 to December 2012. The IRI data records the UPC level transactions of customers of IRI's BehaviorScan Program.[2] This period covers the Great Recession, which started from December 2007 and ended in June 2009, according to the National Bureau of Economic Research. We expect to detect interesting patterns of dynamics of preference sensitivities and price elasticities during this period. In our analysis, we focus on six product categories: coffee, paper towels, potato chips, soda, toilet tissues, and frozen pizza. Since we model customer choices at the brand level, we aggregate the UPC-level choice information into brands. For each

[2]The IRI data has been used in many papers, including Gordon et al. (2013), which contains a detailed discussion of the data.

product category, we keep the transaction data of the top six brands. The top six brands make up 70% to 90% of the market shares in each product category in our data, and store brands are in the top six brands in each categories.

To derive the price information of each brand, we aggregate the UPC-level price information. For each week in each store, we aggregate the price of each UPC bought, weighed by the share of the UPC compared to all UPCs of the brand in that week. Thus, in our cleaned data, price of a brand varies at the store and week level. Price and brand dummies are the only explanatory variables that we use in this application. We standardize the price data before inputting them into the model.

Following Gordon et al. (2013), we model the time variation of preference sensitivities at the quarterly level. For this analysis, we only include the customers who were active in both the starting 6 quarters and ending 6 quarters of our sample and were active in all of the above categories for at least 15 quarters. We arrive at a sample of 141 customers after filtering based on this selection rule. Over the span of the 48 quarters, these customers on average made 723 purchases across all six categories. Soda is the most purchased category, with an average of 292 purchases over the sample period across all customers, and paper towel is the least purchased category, with an average of 64 purchases. All customers purchased in at least one category in each of the 48 quarters. We estimate the model with the first 44 quarters (11 years) of the data and leave the last 4 quarters (year 2012) of the data out for evaluating the forecasting performances of the model.

6. Results

6.1. Latent Factors

Before diving into analyses based on the MCDH, we first need to determine the number of dynamic latent factors to use. While there are many approaches for doing so, including information criteria measures, we chose to do this selection using out-of-sample predictive validity. Using holdout sample mean hit rate as the metric for forecasting accuracy, we find that using four latent functions yields the highest mean hit rate. Thus, we focus our empirical results here

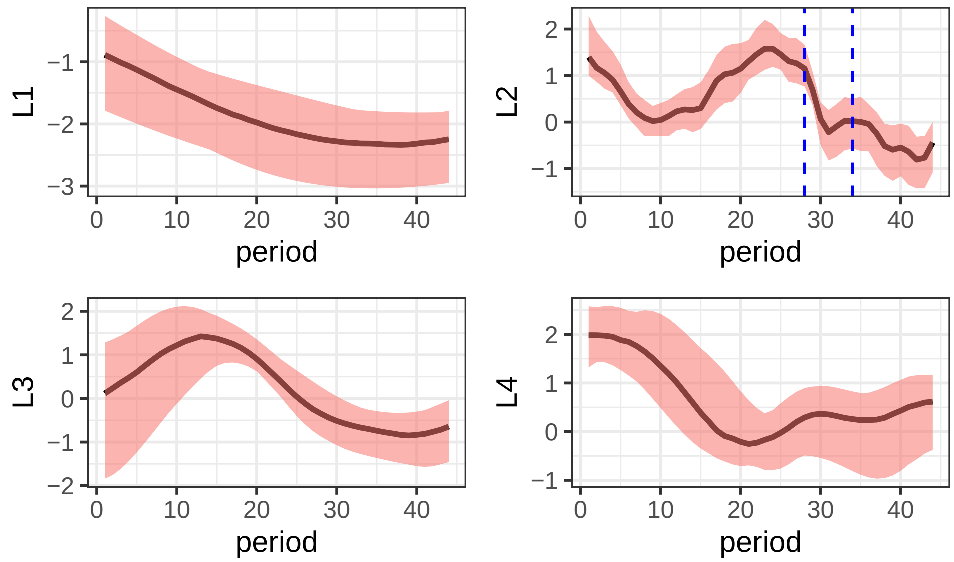

Figure 4: Posterior draws of latent functions: $u_l(t)$. The shaded bands are post burn-in posterior draws and the solid lines are the posterior medians. The x-axis marks the time and y-axis marks the function values. Blue dashed line marks the start and the end of the Great Recession.

on the case that uses four latent functions.

The four latent functions are plotted in Figure 4. Most of the latent functions are relatively smooth and capture long-term trends. The first latent function captures a slowly decreasing trend in customers' preference sensitivities. The third latent function captures an increasing trend in the first half of our sample periods and a decreasing trend in the second. The fourth latent function captures a initially faster decreasing but later slower increasing trend. The second latent function is particularly interesting. The second latent function is more jagged than any other ones. Moreover, it shows this sudden change right around the onset of the Great Recession and a slow recovering trend at the end of the sample period. This latent function captures more quarter-to-quarter perturbation than other latent factors. Moreover, because it captures such a trend that matches what we expect from the Great Recession, we will refer to it as the recession latent factor moving forward. However, as the latent factors are only identified up to a sign, we need to interpret it with some caution. The dynamics in the preference sensitivities may not follow the dynamics in the latent factors and have the opposite trajectory to the dynamics in latent factors, as the ω_{ikl} can take on negative values.

6.2. Correlation Among Preference Sensitivities

In this subsection, we explore the dynamic correlations between customers' price sensitivities and store brand loyalties across categories. In Table 1 and 2, we report the posterior median estimates of two submatrices of Λ_ω. Table 1 records the estimates of correlations between ω_{pl} where p is a subset of the vector K that indicates price sensitivities parameters. Table records the estimates of correlations between ω_{ql} where q is a subset of the vector K that indicates store brand loyalty parameters. These correlation estimates measures how much weight on the latent factors, ω_{ikl} are correlated for different preference sensitivities k. The larger the correlation estimate, the more that two preference sensitivities rely on the same latent factors. Thus, these correlation estimates capture how similar two preference sensitivities' evolution over time are, and we refer to them as the dynamic correlation estimates.

We find that customers' dynamic price sensitivities are positively correlated across all categories and customers' preferences for store brands are also positively correlated across most categories. The dynamic correlation pattern makes intuitive sense. For example, price sensitivity in toilet tissues and paper towels categories, which are considered household items, are highly correlated. Moreover, price sensitivities in chips and soda, which are considered snacks, are highly correlated as well.

Table 1: Dynamic correlation between price sensitivities. We report the posterior median estimate here. Bold text indicates the 95 percent credible interval does not contain 0.

	Chips	Coffee	Frozen Pizza	Paper Towel	Soda	Toilet Paper
Chips	**1**	0.13	**0.4**	0.19	**0.38**	**0.41**
Coffee	0.13	**1**	0.14	**0.29**	**0.26**	**0.42**
Frozen Pizza	**0.4**	0.14	**1**	**0.2**	0.11	**0.21**
Paper Towel	0.19	**0.29**	**0.2**	**1**	**0.18**	**0.33**
Soda	**0.38**	**0.26**	0.11	**0.18**	**1**	**0.41**
Toilet Paper	**0.41**	**0.42**	**0.21**	**0.33**	**0.41**	**1**

Compared to price sensitivities, sensitivities in store brand loyalty across product categories are dynamically correlated to a lesser degree. The highest correlation of sensitivities in store brand is between toilet tissues and paper towels. This is not surprising as there is a greater similarity between these two categories than any other categories that we study here.

Table 2: Dynamic correlation between store brand Loyalty. We report the posterior median estimate here. Bold text indicates the 95 percent credible interval does not contain 0.

	Chips	Coffee	Frozen Pizza	Paper Towel	Soda	Toilet Paper
Chips	**1**	**0.24**	0.03	-0.13	0.02	-0.1
Coffee	**0.24**	**1**	**0.21**	0.09	0.06	0.01
Frozen Pizza	0.03	**0.21**	**1**	0.06	**0.23**	**0.16**
Paper Towel	-0.13	0.09	0.06	**1**	**0.21**	**0.43**
Soda	0.02	0.06	**0.23**	**0.21**	**1**	**0.34**
Toilet Paper	-0.1	0.01	**0.16**	**0.43**	**0.34**	**1**

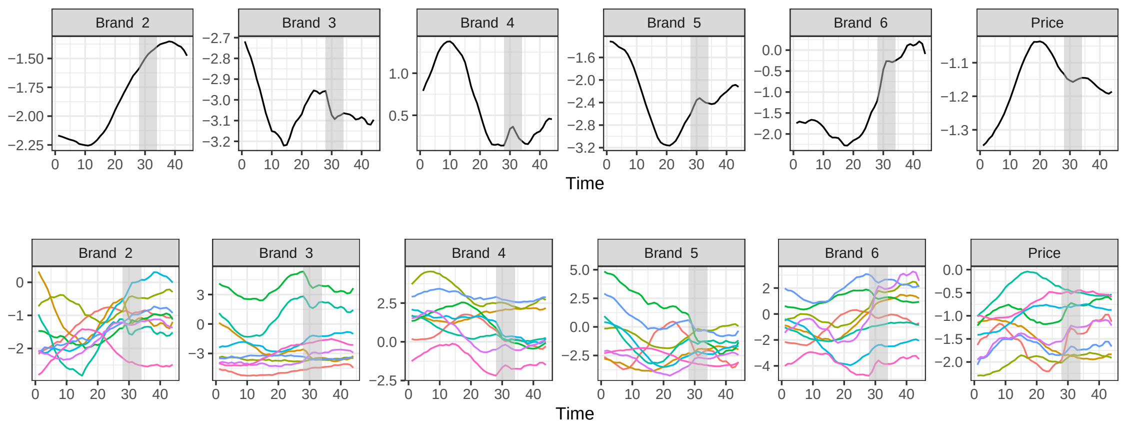

Figure 5: Dynamics of the estimated preference sesnsitivities from the paper towel category. The upper panel display the category level dynamics. The lower panel display the individual level dyanmics, where each colored line represents a different customer.

People tend to group toilet tissues and paper towels into household items while other other product categories in our sample do not fall into the bigger category with toilet tissues and paper towels.

6.3. Brand-level and Individual-level Preference Sensitivities

In this subsection, we explore the estimated dynamics of the preference sensitivities using the paper towel category as a case study. The MCDH yields interesting insights not only at the category level, but also at the individual level. On the category level, we notice that the brand loyalty for the store brand increased substantially. By the end of the sample period, store brand becomes the top three brands in the category. Most customers became less price sensitive in the product category at the start, and then gradually became slightly more price sensitive. The

Great Recession did not seem to affect customers' price sensitivities much in paper towels.

More interestingly, we see that, although individual-level dynamics of preference sensitivities are centered around category-level dynamics, they are largely unrestricted and take on different shapes. There is large dynamic heterogeneity in customers' preference sensitivities. Customers' preference sensitivities varies substantially in terms of the shape and the trend. We see that customers brand loyalty and price sensitivities evolve quite a lot over time but in a smooth fashion. Leveraging such information will allow us to target customers at the individual level. We discuss how our targeting strategy may differ had we not used the MCDH model more in Section 8.

7. Forecasting Performance

In this section, we compare the predictive ability of MCDH against a host of both classic and modern benchmarks, to establish its validity, and show the gains from capturing correlated dynamics in brand choice.

7.1. Benchmarks

For comparison, we estimated the following benchmarks:

Multinomial Logit. The first model we run is the standard multinomial logit model. Under this model, customer $i's$ utility of choosing brand j in category c is

$$u_{icj}(t) = \beta_{icj} + \beta_{icp} P_{icjt} + \epsilon_{icjt},$$

where β_{ic1} is restricted to be 0 across all customers and categories. We collapse all the β_{icj} and β_{icp} into a single vector β_i and assume it is drawn from a normal distribution with a diagonal matrix, such that the k^{th} element of β_i, $\beta_{ik} \sim \mathcal{N}(\mu_{\beta_k}, \sigma_{\beta_k})$. In this model, there is no information sharing across a customer's brand loyalties toward different brand or across a customer's price sensitivities in different categories.

Multinomial Logit with Information Sharing. The second model we run is the multinomial logit model with information sharing. Under this model, customer $i's$ utility of choosing brand j in category c is specified in the same way as model 1, but now $\beta_i \sim \mathcal{N}(\mu_\beta, \Sigma_\beta)$, where Σ_β is a covariance matrix that captures the correlation across a customer's different preference parameters.

Fixed Offsets. The third model is what Dew et al. (2020) referred to as the fixed offsets model. In this specification, there are dynamics, but the individual-level patterns are restricted. Here, customer $i's$ utility of choosing brand j in category c at time t is modeled as

$$U_{icj}(t) = \beta_{icj}(t) + \beta_{icp}(t)P_{icjt} + \epsilon_{icjt}.$$

We collapse $\beta_{icj}(t)$ and $\beta_{icp}(t)$ into a 36×1 vector $\beta_i(t)$. Then, the k^{th} element of $\beta_i(t)$ is modeled as $\beta_{ik}(t) \sim \mathcal{N}(\mu_k(t), \sigma_k)$. $\mu_k(t)$ is the population time-varying preference parameter and is drawn from a Gaussian Process $\mathcal{GP}(\mu_k, k_{SE}(t, t'|\sigma, \rho))$. Therefore, this model allows the population level preference parameter to be time-varying and an individual's preference parameter to deviate from the population one, but the deviation is restricted to be constant across all time periods. Moreover, we do not allow such deviation to pool information across a customer's different preference parameters. Specifications like this, where a dynamic mean is modeled with static heterogeneity, have been used many times in the marketing literature (Kim et al. (2005), Liechty et al. (2005), Sriram et al. (2006), Lachaab et al. (2006), Guhl et al. (2018)).

Fixed Offsets with Information Sharing. The fourth model is the offset model with information sharing. In this model, we keep the utility specification and population level time-varying preference parameter from the offset model. The only thing we change is that we model $\beta_i(t) \sim \mathcal{N}(\mu(t), \Sigma_\beta)$. In other words, we allow individual deviation from the population level function to be correlated across different preference parameters here.

Gaussian Process Dynamic Heterogeneity The last model we consider is the GPDH model implemented in Dew et al. (2020). This is the state-of-art choice model that models dynamics in customer's preference sensitivities. However, the GPDH model does not allow any information sharing among product categories or different preference parameters. The utility specification and population is the same as in the offset model. $\beta_{ik}(t)$ is a draw from a Gaussian Process, such that $\beta_{ik}(t) \sim \mathcal{GP}(\mu_k(t), k_{se}(t, t'|\sigma_k, \rho_k))$. $\beta_k(t)$ is modeled using autoregressive moving average (ARMA) time series. We use the ARMA(1) specification as in Dew et al. (2020), such that $\mu_k(t) = \mu_{kt} = \alpha_{0k} + \alpha_{1k}\mu_{kt-1} + \alpha_2\zeta_{kt-1} + \zeta_{kt}, \zeta_{kt} \sim \mathcal{N}(0, \tau_k^2)$.

7.2. Performance

We report in Table 3 the forecasting performance of the MCDH model compared to the benchmark models, which are all estimated using 44 quarters of data, and evaluated based on forecasting performance on a holdout sample of 4 quarters. Four different metrics are used to evaluate the forecasting performances, including the average hit rate, precision, recall, and specificity. The MCDH outperforms the benchmarks in almost all categories on all metrics The only exception was in soda, where it is slightly outperformed by the GPDH.

Theoretically, the MCDH should outperform the benchmarks when the focal product category: (1) has interesting dynamics in customers' sensitivities; (2) is dynamically correlated with other categories; and (3) has relatively sparse data. When (2) and (3) apply, the information sharing structure of MCDH should improve predictions in the focal category, relative even to complex, dynamic models like GPDH that do not pool information across categories. To explore this, we visualize the difference between mean hit rate of MCDH and GPDH in two ways in Figure 6. The first factor we are interested in is the number of observation in a category. We believe our model should outperform the GPDH model, when the number of observations is lower, again because the MCDH allows for more information sharing across observations. The relationship between hit rate and number of observations is shown in the left panel of Figure 6.

However, this information sharing should only matter in categories that are actually correlated, as estimated by MCDH. To measure the strength of information between categories, as estimated by MCDH, we create a metric that measures the amount of pooling that each product category is getting from the rest. We create the pooling variable by computing the average of the absolute correlations between the preference parameter of a category and the preference

Table 3: Forecasting perfomance of all models. Bolded text indicates the best performance model in the product and metric pair. We report mean hit rate, and macro precision, recall, and specificity.

Category	Model	Hit Rate	Precision	Recall	Specificity
Chips	Logit	0.613	0.633	0.613	0.439
Chips	Logit + info	0.615	0.634	0.615	0.44
Chips	Offset	0.599	0.631	0.599	0.453
Chips	Offset + info	0.591	0.628	0.591	0.452
Chips	GPDH	0.605	0.639	0.621	0.467
Chips	MCDH	**0.639**	**0.657**	**0.639**	**0.496**
Coffee	Logit	0.606	0.674	0.614	**0.695**
Coffee	Logit + info	0.603	0.671	0.611	0.692
Coffee	Offset	0.611	0.674	0.621	0.687
Coffee	Offset + info	0.61	0.676	0.621	0.69
Coffee	GPDH	0.606	0.661	0.635	0.642
Coffee	MCDH	**0.645**	**0.686**	**0.655**	0.69
Frozen Pizza	Logit	0.566	0.651	0.568	0.771
Frozen Pizza	Logit + info	0.566	0.649	0.567	0.77
Frozen Pizza	Offset	0.564	0.651	0.565	0.767
Frozen Pizza	Offset + info	0.567	0.653	0.568	0.768
Frozen Pizza	GPDH	0.55	0.599	0.533	0.705
Frozen Pizza	MCDH	**0.631**	**0.693**	**0.633**	**0.802**
Paper Towel	Logit	0.46	0.489	0.465	0.694
Paper Towel	Logit + info	0.464	0.494	0.471	0.697
Paper Towel	Offset	0.444	0.489	0.451	0.69
Paper Towel	Offset + info	0.449	0.495	0.457	0.698
Paper Towel	GPDH	0.433	0.503	0.454	0.696
Paper Towel	MCDH	**0.499**	**0.508**	**0.503**	**0.708**
Soda	Logit	0.474	0.479	0.474	0.682
Soda	Logit + info	0.476	0.482	0.476	0.685
Soda	Offset	0.453	0.471	0.454	0.675
Soda	Offset + info	0.454	0.474	0.455	0.676
Soda	GPDH	**0.501**	**0.512**	**0.505**	**0.715**
Soda	MCDH	0.498	0.501	0.498	0.701
Toilet Paper	Logit	0.588	0.602	0.588	0.791
Toilet Paper	Logit + info	0.591	0.606	0.591	0.794
Toilet Paper	Offset	0.586	0.615	0.586	0.809
Toilet Paper	Offset + info	0.591	0.62	0.591	0.813
Toilet Paper	GPDH	0.579	0.603	0.583	0.799
Toilet Paper	MCDH	**0.611**	**0.632**	**0.611**	**0.825**

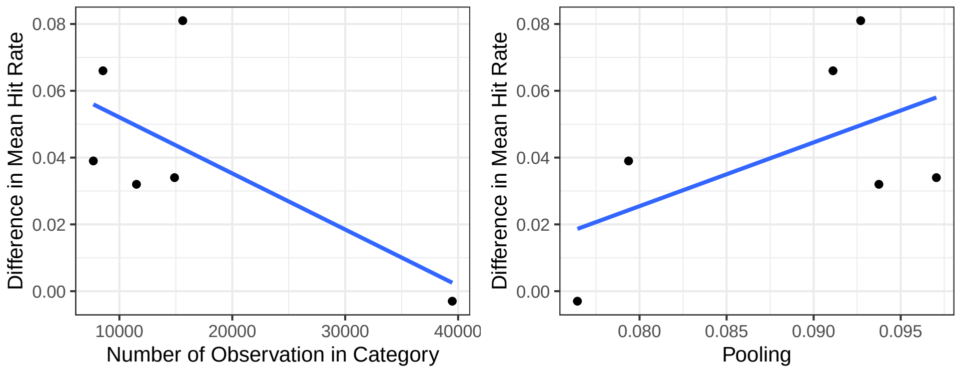

Figure 6: Difference in mean hit rate between MCDH and GPDH. Each point represents a product category. The line is the smoonthed conditional mean of the difference in mean hit rate, conditional on the variable on the x axis.

parameters of all other categories. To define the pooling variable of category c mathematically, let J_c be the set of row indices (or equivalently, column indices) of Λ corresponding to the parameters in category c, and let $|J_c|$ be the cardinality of that set. For instance, if category 2 has 5 parameters, which are elements 6-10 of $\beta_i(t)$, then $J_c = \{6, 7, 8, 9, 10\}$, and $|J_c| = 5$. Equivalently, denote $J_{-c} = \{1, \ldots, K\} \setminus J_c$. We define the magnitude of pooling for category c as

$$\text{Pooling}_c = \frac{1}{|J_c|} \sum_{i \in J_c} \frac{1}{|J_{-c}|} \sum_{j \in J_{-c}} |\Lambda_{ij}|.$$

Intuitively, this measure just computes the absolute strength of correlation between each parameter in the focal category, and each parameter not in the focal category, and averages these together. A higher Pooling_c indicates the category c is more strongly correlated with other categories more generally. We show the relationship between hit rate and this metric relationship in the right panel of Figure 6.

Figure 6 shows that the fewer observations we have in a product category, the better forecasting performance MCDH model has over the GPDH model. We have about 40,000 observations in the soda categories, which is more than twice as many as the number of observations in any other categories, and GPDH outperforms over MCDH slightly in the soda category. Moreover, figure 6 shows that the more pooling a product category has from other product categories, the better the forecasting performance will be compared to GPDH model. These

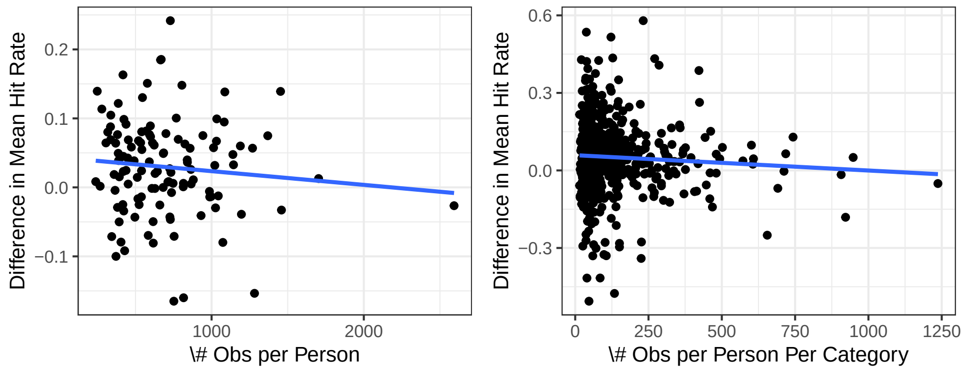

Figure 7: Difference in mean hit rate between MCDH and GPDH. Each point represents an individual (on the left) or an individual category pair (on the right). The line is the smoonthed conditional mean of the difference in mean hit rate, conditional on the variable on the x axis. The pattern is robust to whether we include the rightmost outlier or not.

finding makes intuitive sense, as the major advantage of MCDH over GPDH is the ability to leverage information from other product category to improve estimation in the focal product category. When a product category does not pool much information from others, or when we have many observations in a product category that it does not need to pool information from other categories, the advantage of MCDH over GPDH diminishes.

The previous analysis focused on the category-level. However, we can also do this analysis at the individual-level. Recall that MCDH provides many sources of information sharing across individuals. Thus, we would expect to see gains following similar patterns for individual-level predictions. We visualize the relationship between the number of observations of an individual and the individual-level hit rate in Figure 7. We find that on average, when there is relatively scarce observation of an individual or an individual category pair, MCDH performs better than GPDH. This finding is again in line with our intuition that MCDH will outperform GPDH when there is a larger need for pooling information across categories or across individuals.

8. Price Elasticities

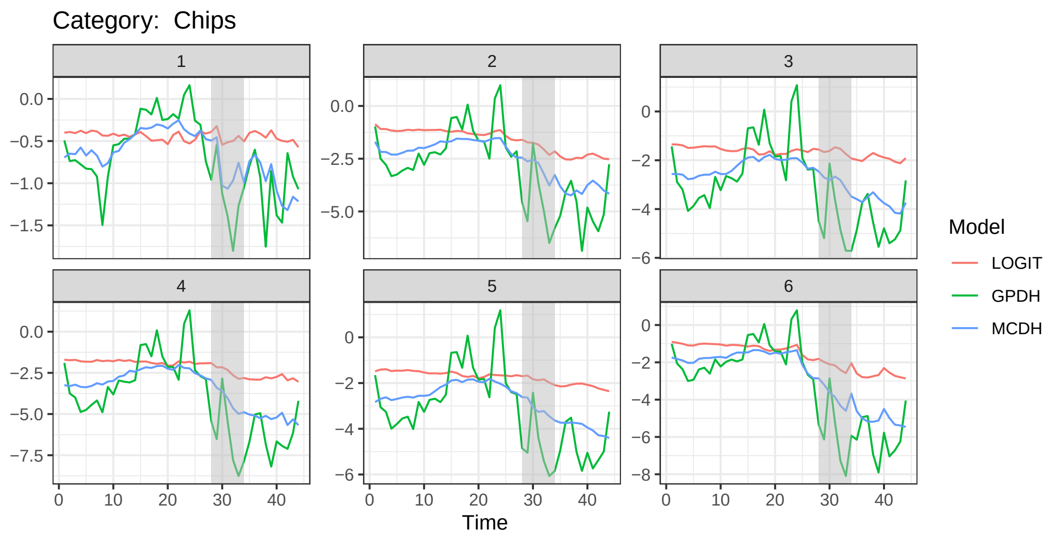

Figure 8: Dynamics of category-level price elasticities. Each panel represents the a brand in the chips category. Each color line represents the price elasiticities computed from a competing model. The grey vertical area marks the start and end of the Great Recession.

8.1. Category-level Price Elasticities

While MCDH outperforms the benchmarks in terms of forecasting performance, it also enables richer and more precise insights that are useful for managers. One such insight is the estimation of dynamic price elasticities. To study the category-level dynamics of price elasticities, we focus on a single category, chips, as a case study. Computing price elasticities under our framework is simple, and directly follows the standard MNL formulation: $\epsilon_{ijt} = \beta_{i\text{Price}_c}(t) \times \text{Price}_{\text{ijt}} \times [1 - p_{ijt}]$, where i indexes individuals, c category, and j choice alternatives within c. In Figure 8, we plot the average price elasticity across individuals, broken out by categories. Besides the elasticities from our MCDH model, we also plot the price elasticities computed from the GPDH and the standard logit models. The elasticities computed from the logit model are rather static, and suffer from the attenuation bias previously documented by Dew et al. (2020), where price elasticities estimated from static models are underestimated compared to models that capture dynamics in preference parameters. The logit model misses the fact that customers were more price elastic during the Great Recession. Comparing the price elasticities computed from MCDH and GPDH model, we see that MCDH model give us smoother price elasticities estimates. The ability to pool information across categories, through a parsimonious set of

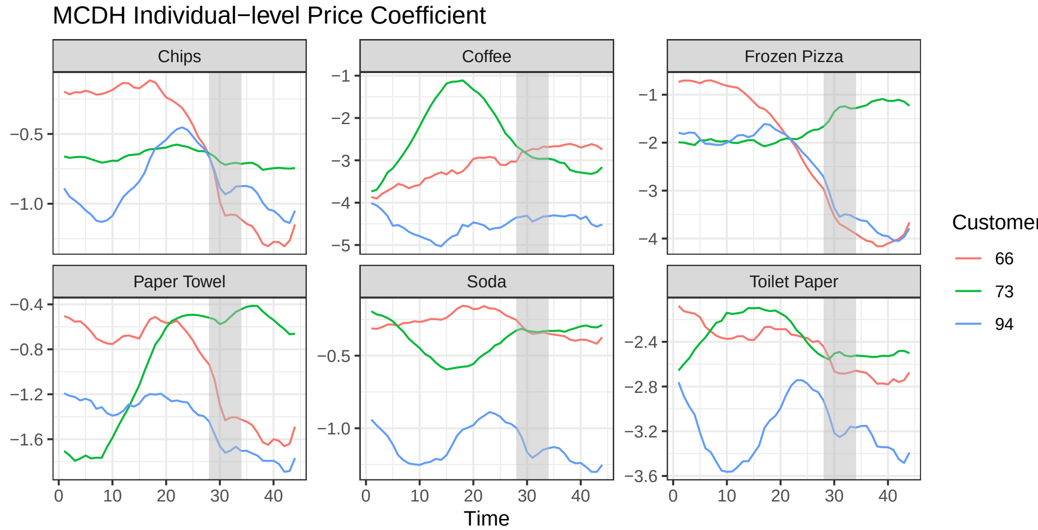

Figure 9: Dynamics of individual-level price coefficients in MCDH. Each panel represents the a category. Each color line represents the price coefficient of a customer. The grey vertical area marks the start and end of the Great Recession.

latent functions regularizes the function estimates. In the GPDH model, the price elasticities estimates are not only more jagged, they also become positive in several periods. The noise in the data and the insufficient information sharing structure in GPDH results in unreliable estimates of price elasticities.

8.2. Individual-level Price Coefficients

Now we move from the category-level price elasticities and zoom in on the individual-level price coefficients (i.e., sensitivities) in the MCDH model. As a case study, we plot these coefficients for three randomly sampled customers in Figure 9. The price sensitivities of these three customers display very different dynamics. For example, customer 73's price sensitivities are relatively constant throughout the entire calibration window, while customer 66 and 94 feature strong dynamics, including dramatic changes around the time of the Great Recession, in almost all categories. Notably, while there are some general trends, there are also important individual-level idiosyncracies in the dynamics: although each of these customer's parameters were computed by linear combinations of four latent functions, the individual-level curves are very different, in important and useful ways. Compare, for instance, the three consumers in the frozen pizza category: we have one consumer who experienced a decline in price sensitivity

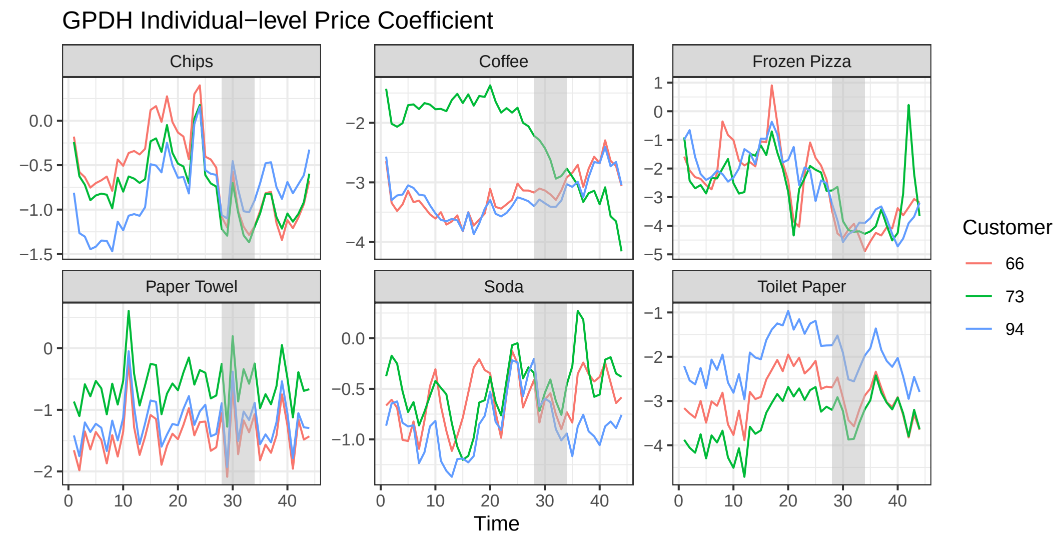

Figure 10: Dynamics of individual-level price coefficients in GPDH. Each panel represents the a category. Each color line represents the price coefficient of a customer. The grey vertical area marks the start and end of the Great Recession.

(i.e., became much more price sensitive) gradually throughout the entire calibration window, another consumer whose price sensitivity remained rather constant, with a slight uptick toward the end, and a third whose price sensitivity was constant but then dramatically declined at the time of the Great Recession.

These dynamics have important implications for targeting. Consider, for instance a targeted couponing strategy based on price sensitivity. A strategy based on a static model would certainly have ignored extremely relevant trends, like for instance that consumers 66 and 73 essentially switched places in terms of the price sensitivity in paper towels, or that consumer 66 has become much more price sensitive across the board, particularly in chips, frozen pizza, and paper towels.

While dynamic insights are one of the benefits of MCDH, recently introduced models like GPDH also allow for insights about individual-level dynamics. The benefit of MCDH relative to these models is not just that it allows for individual-level dynamics, but it can estimate these dynamics more precisely. To compare our insights to GPDH, we plot the price coefficients of the same customers above, but estimated from GPDH, in Figure 10. We see the estimates are much noisier, and suggest less obvious trends in behaviors. This is because the GPDH model relies solely on pooling information within a category, through a category-specific mean, which itself may exhibit noisy dynamics. The increased smoothness of the MCDH estimates is not merely a statistical gain, though: again, targeting strategies based on MCDH versus GPDH may be very different. For instance, price sensitivities exhibiting many period-to-period changes that may not be reflective of customers' true changes of price sensitivities. More importantly, too much period-to-period variation in GPDH may also fail to pick up the long term trends of customers' price sensitivities. Take the paper towel category as an example. The GPDH estimates a lot of period-to-period variation in price sensitivities, but with an apparently constant long-term trend, while the MCDH shows that customer 73 becomes less price sensitive in paper towel gradually and customer 66 becomes more price sensitive in paper towel as time goes on. These differences are thus not just a matter of statistical efficiency, but practical relevance.

9. Conclusion

In this work, we develop a flexible framework, the MCDH model, to capture cross-category, dynamic, individual-level marketing sensitivities in the context of brand choice. MCDH leverages latent, multi-output Gaussian processes and a hierarchical correlation structure across customers to accurately infer dynamics in customers' sensitivities to marketing variables. At the heart of the model is a simple specification of dynamic latent factors that underpin consumers' sensitivities to marketing variables, which can parsimoniously capture common trends in markets, and share information across product categories. By allowing information sharing across individuals, categories, time, and different preference sensitivities, MCDH provides more precise estimates that can be used for the optimization of the marketing mix. Using IRI grocery purchase data, we show that the latent factors can uncover interesting dynamics that exist in customers' price sensitivities, and that the MCDH considerably improves in forecasting ability over modern benchmarks.

Our work contributes to the literature both methodologically and substantively. Methodologically, our MCDH specification is new to the literature in marketing and machine learning. To our knowledge, it is the first application of multi-output GPs to marketing problems. More generally, ours is also the first to use MOGP to specify sensitivities, rather than a regression function. We show the clear gains in doing so: the model parameters retain their classic interpretations (e.g., as price sensitivities), but are specified in a dynamic, flexible, correlated fashion. We also show that the MCDH model outperforms previously developed brand choice models in various forecasting metrics, and achieves more reliable price elasticity estimates than previously developed brand choice models. Substantively, our work documents interesting dynamics that took place in consumer preferences surrounding the Great Recession. We also find correlations in *dynamics* in both customers' price sensitivities and in their store brand loyalties, but that such correlations are stronger in price sensitivities than in store brand loyalties.

Our work has some limitations that provide opportunities for future research. First, while still interesting, our data is also quite limited: we focus on a subset of the IRI data, where customers were highly active over our sample period and made purchases at the start and the end of our sample period. Doing so allows us to rule out that the observed dynamic heterogeneity was driven by customer attrition. However, this precludes us from studying more sparse settings where the information sharing of the MCDH may even further improve insights and performance. In general, we see considerable room for application of MCDH even beyond grocery purchasing data, including to dynamic and sparse settings like e-commerce purchasing.

Finally, the MCDH specification we proposed is simple, and as we have documented, works well empirically. However, the literature on MOGP models is vast, including a wealth of specifications from geostatistical and machine learning applications. Our specification was inspired by perhaps the simplest of the MOGP models: the semiparametric latent factor model. Future work may examine more complex specifications for the cross-category correlations, or even allow the cross-category correlation matrix to itself evolve. Such an approach is suggested in recent work by Liu et al. (2018). Moreover, the kernel we use in capturing the latent factors is perhaps the most simple kernel used in the GP literature: the squared exponential. While we find that our results are not particularly sensitive to this choice, recent work on MOGP suggests more complex structures that can be used to better share information across the latent processes (i.e., the different preference parameters), including the asymmetric kernels described in Liu et al. (2018). All of these suggest rich possibilities for extending this work, and further improving our ability to understand correlated, dynamic heterogeneity.

References

10 Ainslie, A., & Rossi, P. E. (1998). Similarities in Choice Behavior Across Product Categories. *Marketing Science*, *17*(2), 91–106. https://doi.org/10.1287/mksc.17.2.91

Álvarez, M. A., Rosasco, L., & Lawrence, N. D. (2012). Kernels for vector-valued functions: A review. *Foundations and Trends® in Machine Learning*, *4*(3), 195–266. https://doi.org/10.1561/2200000036

Betancourt, M. (2018). A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv:1701.02434 [Stat]*. http://arxiv.org/abs/1701.02434

Dew, R., & Ansari, A. (2018). Bayesian Nonparametric Customer Base Analysis with Model-Based Visualizations. *Marketing Science*, *37*(2), 216–235. https://doi.org/10.1287/mksc.2017.1050

Dew, R., Ansari, A., & Li, Y. (2020). Modeling Dynamic Heterogeneity Using Gaussian Processes. *Journal of Marketing Research*, *57*(1), 55–77. https://doi.org/10.1177/0022243719874047

Du, R. Y., & Kamakura, W. A. (2012). Quantitative Trendspotting. *Journal of Marketing Research*, *49*(4), 514–536. https://doi.org/10.1509/jmr.10.0167

Gordon, B. R., Goldfarb, A., & Li, Y. (2013). Does Price Elasticity Vary with Economic Growth? A Cross-Category Analysis. *Journal of Marketing Research*, *50*(1), 4–23. https://doi.org/10.1509/jmr.11.0162

Guhl, D., Baumgartner, B., Kneib, T., & Steiner, W. J. (2018). Estimating time-varying parameters in brand choice models: A semiparametric approach. *International Journal of Research in Marketing*, *35*(3), 394–414. https://doi.org/10.1016/j.ijresmar.2018.03.003

Hansen, K., Singh, V., & Chintagunta, P. (2006). Understanding Store-Brand Purchase Behavior Across Categories. *Marketing Science*, *25*(1), 75–90. https://doi.org/10.1287/mksc.1050.0151

Kim, J. G., Menzefricke, U., & Feinberg, F. M. (2005). Modeling Parametric Evolution in a Random Utility Framework. *Journal of Business & Economic Statistics*, *23*(3), 282–294. https://www.jstor.org/stable/27638821

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*(1), 57–81. https://doi.org/10.1007/s11129-006-6559-x

Lee, S., Kim, J., & Allenby, G. M. (2013). A Direct Utility Model for Asymmetric Complements. *Marketing Science*, *32*(3), 454–470. https://doi.org/10.1287/mksc.2013.0782

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001.

Liechty, J. C., Fong, D. K. H., & DeSarbo, W. S. (2005). Dynamic Models Incorporating Individual Heterogeneity: Utility Evolution in Conjoint Analysis. *Marketing Science*, *24*(2), 285–293.

https://doi.org/10.1287/mksc.1040.0088

Liu, H., Cai, J., & Ong, Y.-S. (2018). Remarks on multi-output gaussian process regression. *Knowledge-Based Systems*, *144*, 102–121.

Manchanda, P., Ansari, A., & Gupta, S. (1999). The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions. *Marketing Science*, *18*(2), 95–114. https://doi.org/10.1287/mksc.18.2.95

Mehta, N. (2007). Investigating Consumers' Purchase Incidence and Brand Choice Decisions Across Multiple Product Categories: A Theoretical and Empirical Analysis. *Marketing Science*, *26*(2), 196–217. https://doi.org/10.1287/mksc.1060.0214

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.

Schmidt, A. M., & Gelfand, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research: Atmospheres*, *108*(D24). https://doi.org/https://doi.org/10.1029/2002JD002905

Seetharaman, P. B., Ainslie, A., & Chintagunta, P. K. (1999). Investigating Household State Dependence Effects across Categories. *Journal of Marketing Research*, *36*(4), 488–500. https://doi.org/10.2307/3152002

Song, I., & Chintagunta, P. K. (2007). A Discrete–Continuous Model for Multicategory Purchase Behavior of Households: *Journal of Marketing Research*. https://doi.org/10.1509/jmkr.44.4.595

Sriram, S., Chintagunta, P. K., & Neelamegham, R. (2006). Effects of Brand Preference, Product Attributes, and Marketing Mix Variables in Technology Product Markets. *Marketing Science*, *25*(5), 440–456. https://doi.org/10.1287/mksc.1050.0188

Teh, Y. W., Seeger, M., & Jordan, M. I. (2005). *Semiparametric Latent Factor Models*. 8.

Wackernagel, H. (2013). *Multivariate geostatistics: An introduction with applications*. Springer Science & Business Media.