Utility and Attention - a structural model of consideration

by

Keyvan Dehmamy and Thomas Otter

May 5, 2014

Goethe University, Frankfurt am Main

Abstract

Models of consumer decision making often condition on attention to the different offers or alternatives to choose from. However, in many environments offers not only compete through their utility but also for the attention of decision makers. In this case, it is important to distinguish between attention and utility - it makes a difference whether an offering is overlooked, or rejected conditional on awareness - for optimal marketing control and empirical measures of competition. We show how quantity choices, in contrast to multinomial choices, facilitate the empirical distinction between attention and utility. In our illustrative application we analyze choices from simulated store shelves. We find that the number of facings of a brand on the shelf influence attention but not utility from the brand. We then formulate a parametric model that identifies attention based considerations sets. We find overwhelming support for this model compared to a model that ignores attention and motivates choices from utility only.[1]

Keywords: Discrete-continuous demand, consideration-sets, awareness, data augmentation, hierarchical Bayes

[1]We are grateful for the comments from seminar participants at the 2012 Marketing Science Conference, the 2012 ART Forum, University of Zurich, Erasmus University Rotterdam, and the Winter Marketing Economics Summit 2014. We are indebted to the Modellers® for providing us with the data.

E-mail addresses: otter@wiwi.uni-frankfurt.de, dehmamy@wiwi.uni-frankfurt.de.

1 Introduction

Marketing theory and practice is concerned with drawing attention to offers that are expected to be evaluated favorably by consumers. Marketing activities both influence the amount of attention for an offer as well as the value of the offer experienced by consumers. Attention (to an offering) by a consumer can be viewed as a requirement for the consumer to perceive value in it, i.e., it is possible that a consumer passes by an offering of high value because of a lack of attention. In contrast, marketers might invest in costly but effective measures to create attention for an offer, and the offering can still be rejected by many consumers because of a lack of perceived value.

Models of consumer decision making often condition on attention to the different offers or alternatives to choose from, or implicitly equate utility and attention when they include variables that control attention as arguments to the utility function. The assumption that consumers are aware of all alternatives in a particular choice situation often is questionable, and especially in environments where offers compete for the attention of decision makers, e.g., on a supermarket shelf.

Including likely causes of attention, for example the number of facings of a brand on the shelf, as arguments in the utility function is hamstrung by two difficulties: First, the interplay between utility and attention at the source of consumers' actions is highly non-linear and discontinuous. Without attention, variables that control (indirect) utility such as e.g., price have no effect. Second, the absence or presence of attention is only a probabilistic function of variables that control attention, at least over some range of these variables. Neither aspect is handled well by standard utility functions.[2]

However, the distinction between attention and utility based on consumers' choices is difficult. Unlike asking a consumer about whether he is aware of a particular offer and how much he values the offer, consumers' choices only reveal the combination of awareness *and* a positive valuation when a product is chosen, and the lack of awareness *or* the lack of a sufficiently positive valuation when a product is not chosen. Therefore, empirical identification of utility and attention requires prior assumptions about functional forms, e.g., exclusion restrictions that state which observed covariates drive utility and which drive attention (and not utility).[3] Similarly, the distinction between a consideration stage and utility maximization conditional

[2]It is certainly mathematically possible to conceive of the absence of attention as an extreme utility shock, the distribution of which is controlled by variables that facilitate attention. However, the implication is that there are causes to utility that only determine if a purchase could potentially occur, with no bearing on how much someone values a specific offer, which in turn implies that there are states of the world where a utility maximizing consumer is indifferent between exactly identical offers at different prices.

[3]Technically, a model for attention and utility fit to multinomial choice data is a model of selection based on attention where the outcomes 'failed the attention stage' and 'failed because of insufficient utility, conditional on attention' cannot be distinguished from the data. As a consequence, and in contrast to situations where the selection outcome is observed independently, identification requires functional form assumptions even before considering unobserved dependence between attention and utility.

on consideration based on multinomial choice observations requires functional form assumptions that define and thus identify 'consideration' and 'utility' a priori (e.g., Terui et al., 2011; Sovinsky Goeree, 2008; van Nierop et al., 2010; Bronnenberg and Vanhonacker, 1996).

Draganska and Klapper (2011) show how additional data on the distribution of consideration sets can be used to improve inference about the utility function when the full information assumption is not met. We contribute by demonstrating how observations of choice and quantity, i.e., quantity choices help jointly identify standard utility maximization and a preceding selection process which determines the set of alternatives utility is maximized over. We show that under the assumptions that i) all consumers are 'in the market' (Fennell et al., 2003) and that ii) they determine what *and* how much to choose by maximizing (constrained) utility, the information about a selection step preceding utility maximization afforded by quantity choices approaches the information content in the combination of multinomial choices with independent observations of the selection outcomes. We believe that this result is useful because independent data on selections that precede conditional utility maximization are not generally available.[4]

[4]Technically, the assumption that someone is in the market defined by a particular product category implies that all consumers under consideration are responsive to price, conditional on the selection that may precede utility maximization. We thus rule out utility functions accommodating variables that only determine the possibility of choosing a positive quantity at all, such as e.g., dog ownership in the context of demand for dog food, and variation in such variables (see also Footnote 2). We thank Emir Kamenica for

In our illustrative application, we find strong empirical evidence for a selection step that precedes conditional utility maximization in a conjoint experiment designed to measure how shelf space allocation influences demand. We find that the number of shelf facings occupied by a brand on the shelf increases the probability for the brand to enter conditional utility maximization, but does not contribute to the brand's utility. Given this result and the nature of the variable controlling the selection process, i.e., the number of facings, we conclude that attention is at the source of the selection process.

The remainder of the paper is organized as follows: We first develop how quantity choices are informative about selections that precede utility maximization and utility functions. Next, we present our data and use our arguments to identify the presence of an attention based selection step that precedes utility maximization, before committing to a particular model. We then extend an existing model of conditional utility maximization by a model of attention and document strong empirical support for this extension. The Appendix summarizes computational details.

pointing this out to us (Marketing Economics Research Summit, 02-01-2014).

2 Information In Quantity Choices

2.1 Multinomial Choice

It will be useful to first briefly restate how the combination of multinomial choices with independent observations of selections, which determine which brands enter utility maximization, helps with the identification of basic model features. In the absence of correlated unobservables, the comparison between the set of regressors that predict the selection outcome and that predicting the multinomial choice outcome conditional on selection, identifies variables that exclusively drive the selection or the utility motivating multinomial choice, given selection. Alternatively, if theory points to at least one variable that exclusively drives the selection process, the combination of multinomial choices with independent observations of selections identifies conditional dependence between the selection decisions and multinomial choice, given selections (e.g., Wachtel and Otter, 2013).

However, when only the final multinomial outcomes are observed, the identification of a selection step that precedes utility maximization requires functional form assumptions, even if conditional independence between the selection and utility is assumed. Moreover, the degree of potential conditional dependence between the selection preceding utility maximization and utility is fundamentally unidentified in this case. Even exclusion restrictions are insufficient to jointly identify the selection equation and conditional dependence between the selection equation and utility. The function relating the excluded variable to the unobserved selection outcome can always be adjusted to accommodate the ‘indirect’ effect on utility through conditional dependence between the selection equation and the utility function. Next we show how quantity choices usefully improve on the (lack of) information present when only final multinomial outcomes are observed, approaching the information contained in the combination of independent observations of selection outcomes and conditional multinomial choice.

2.2 Quantity Choices - One Inside Good

We first discuss the stylized example of one brand, i.e., the inside good, and an outside good. For illustrative purposes, and in line with our empirical results, we refer to the selection that determines if a brand enters utility maximization as selection based on attention throughout.

The consumer always considers the outside good but does not necessarily pay attention to the brand available for purchase. The dotted line in Figure 1 depicts this consumer’s unobservable price demand curve conditional on attention to the brand.[5] The dashed line underneath is the observable price-

[5]The functional forms implied by the examples in Figure 1 are without loss of generality and immaterial to our argument.

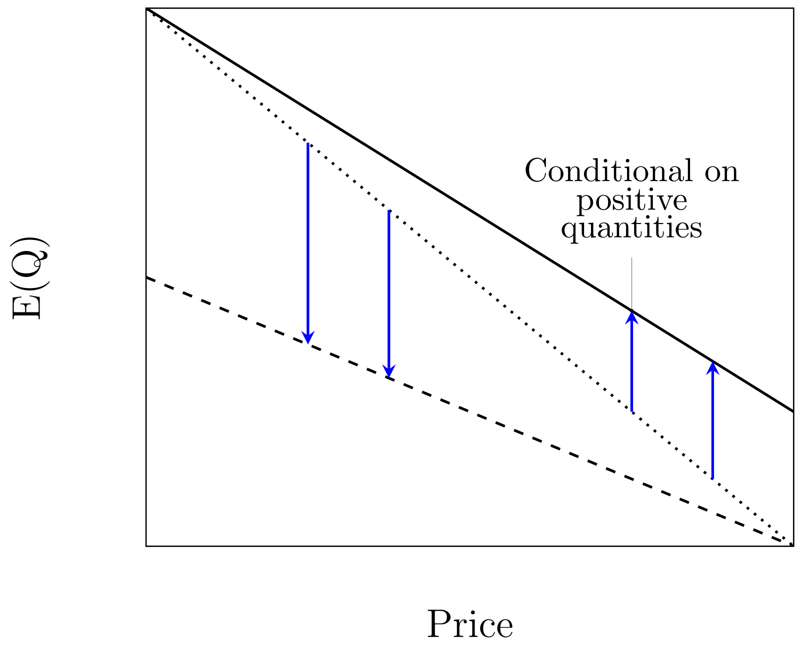

Figure 1: Demand curves under full and limited attention

demand curve that is marginal with respect to attention. For example, if the probability of paying attention to the brand is constant and equal to 0.5, the observable marginal demand is half of the unobservable conditional demand for every price. Finally, the solid line above is the (directly observable) price-demand curve, conditional on positive observed demand. It is constructed by omitting all observations with zero demand and therefore conditions on attention *and* sufficiently positive utility.[6]

Conditioning on realized positive demand implicitly selects observations based on relatively more positive realizations of stochastic arguments to con-

[6]Observed positive demand is impossible without attention to the brand. However, zero demand can occur if the brand's indirect utility is sufficiently low, even conditional on attention.

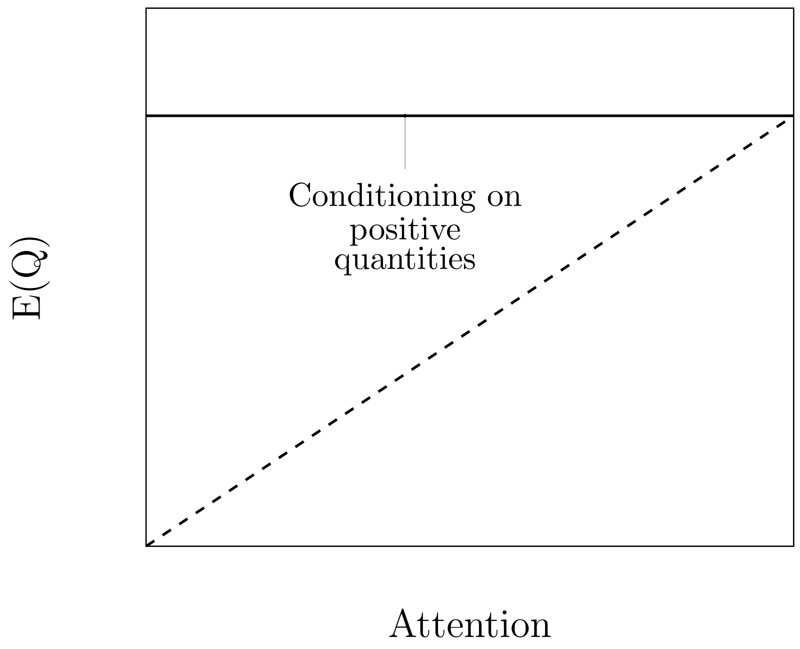

Figure 2: Demand as a function of increasing attention probabilities

ditional utility maximization. Because of this selection, the distance between a point on the solid line and the corresponding point on the unobserved conditional price-demand curve (the dotted line in Figure 1) is decreasing in the indirect utility of the brand.

In Figure 1 price influences both marginal demand (the dashed line) and quantity demand conditional on observing positive quantities, i.e., strictly positive demand (the solid line). In Figure 2 we changed the x-axis from price to a variable that only drives attention. The solid line in Figure 2 is again constructed by omitting all observations with zero demand, and the dashed line is the marginal demand curve. As the probability of attending to the brand increases, marginal demand increases. However, conditional on positive demand, i.e., conditional on knowing that the brand was attended to, the quantity demanded is constant and therefore independent of variables that only control attention.

An implicit assumption here is that unobserved, stochastic aspects of attention are orthogonal to utility. If this is not the case, the selection based on positive errors in the attention process, inherent to conditioning on positive quantities when observable covariates to attention are unfavorable, translates into selection based on utility errors which change expected positive quantities, even if observed covariates only drive attention.

However, under the assumption of conditional independence between utility and attention, finding an association between a variable and the collection of strictly positive demand observations implies that this variable is an argument to utility maximization. Conversely, finding independence between a variable and the collection of strictly positive realizations of demand implies that this variable should be excluded from the utility function. If the same variable is associated with marginal changes in demand, it defines an unobserved selection process that precedes utility maximization, i.e., it only determines if a brand enters utility maximization at all with no bearing on how much the brand is valued.

Alternatively, if theory points to at least one variable to be excluded from the utility function that is associated with changes in marginal demand, i.e., with the probability of paying attention and thus observing non-zero demand only, quantity choices facilitate the identification of unobserved dependence between attention and utility in analogy to the classic Heckman selection model (Heckman, 1979).

However, different from the classic selection model, the selection outcome is not fully independently observed in quantity choices. Observing zero demand can be rationalized by low utility or a lack of attention. Therefore, regressing the binomial outcomes 'zero demand' versus 'non-zero demand' on variables excluded from the utility function and other variables fails to estimate the selection equation consistently, in general.[7]

The partial observability of attention - positive quantities are conditional on attention - coupled with observable variance in positive quantities, however, enables the parametric identification of conditional dependence. The intuition for the identification is that conditional dependence between attention and utility gives rise to systematic, observable variation in strictly positive quantities as a function of the probability of attention, given attention and holding observed argument to the utility function constant. In summary, the information in quantity choices about a selection step that precedes utility maximization usefully improves on that in simple multino-

[7]The selection equation will be estimated consistently if attention and utility are conditionally independent and the fluctuations between zero and positive quantities because of unobserved utility shocks are orthogonal to observed covariates.

mial choices and approaches that in independent observation of the selection outcome and a multinomial choice outcome.

2.3 Quantity Choices - Multiple Inside Goods

The case of multiple brands is more complicated because attention is a finite resource such that drawing more attention to one brand usually comes at the price of a higher chance of overlooking another brand. First, the physical environment, e.g., a store shelf may be space constrained and allocating more shelf-facings to one brand to facilitate attention to this brand, in the limit, translates into dropping other brands from the shelf altogether. Second, human attention is of inherently limited capacity (Miller, 1956). Therefore, an increase in attention to one brand brought about by an increase in a variable that only influences attention may cause another brand to be overlooked, i.e, excluded from conditional utility maximization.

The complication relative to the earlier example with only one brand and an outside good arises because, when the set of brands attended to changes, observed positive quantities may adjust. However, if one is willing to make the standard assumption that utility from a brand at most depends on what other brands are chosen, but not on what other brands were in the choice set, quantity adjustments because of changes in the set of brands attended to can be identified from the data before committing to a model.[8]

Consider the case of two brands A and B competing in attention and (indirect) utility for a consumer's demand and an outside good the consumer always pays attention to. If a particular covariate controls attention but not utility, the function relating this covariate to strictly positive demand for e.g., brand A will be flat except for discrete jumps occurring when observed demand for brand B changes from positive to zero, or vice versa. Put differently, any response in positive quantities demanded from brand A to a covariate that influences attention only is fully mediated by the presence or absence of positive demand for brand B.

If a particular covariate therefore is correlated with changes in demand for brand A but, after controlling for the presence or absence of positive demand for brand B, is orthogonal to strictly positive demand for brand A, the analyst learns that this covariate can only influence attention and

[8]This assumption rules out all direct effects from attributes of other alternatives to the utility of a focal alternative. However, it does not require independence between stochastic utility components, or a specific distribution of random utility. If there were direct effects from attributes of other alternatives to the utility of a focal alternative, the utility of and therefore the optimal quantity from the focal alternative would depend on which of the non-chosen alternatives the consumer paid attention to. In contrast, the distribution of random utility does not depend on what alternatives a consumer pays attention to, even if random utility is distributed dependently. For example, given realizations of stochastic utility and observing a positive quantity for brand A and zero for brand B, the positive quantity for brand A would not adjust if we deleted brand B from the choice set, i.e., conditioned on a lack of attention to brand B. The lack of adjustment from deleting brand B is, conditional on error realizations, independent from utility or attention causing B to be rejected in the first place. In general, unless paying attention to a brand changes the budget or the (indirect) utility function, the specific reason, i.e., a lack of utility or of attention, for choosing nothing from a particular brand is immaterial for observed positive quantities that maximize utility conditional on attention (cf. Bronnenberg, 2013).

that it has to be excluded from the utility function, of course subject to the assumption of conditional independence between attention and utility.[9]

Conversely, all covariates found to shift positive quantities conditional on a fixed set of brands receiving positive demand need to be included in the (indirect) utility function, again subject to the assumption of conditional independence between attention and utility. However, an influence of such covariates on attention cannot be directly ascertained or ruled out.

3 Data

We illustrate our arguments using data from a commercial study designed to measure demand effects of shelf-space allocation.[10] Respondents saw virtual store shelves on a computer screen featuring different varieties, i.e., 'brands' of packaged pre-cut and washed salads. The respondents' task was to indicate which brands they would choose and how many packs of each brand. All study participants were in the market defined by the category of packaged pre-cut and washed salads, as per their past purchase behavior. Respondents were asked to imagine a typical shopping trip involving the intention to purchase from this category and to choose as many brands and packs per

[9]Only the presence or absence of positive demand for B, not the actual quantity demanded from B needs to be controlled for. Therefore, the argument is independent of the specific substitution relationships between brands in the demand system conditional on attention, i.e., perfect substitution, weak complementarity, or strong complementarity.

[10]We thank the Modellers® for providing us with the data.

brand from a shelf as they would in a real purchase situation. Therefore, respondents' stated preferences for brands and quantities on a particular shelf can be represented as a multivariate count vector.

Each respondent saw 16 different shelves in total. Respondents were asked to treat each of the 16 shelves as offers faced in one of 16 independent shopping trips, much like in a standard choice based conjoint experiment (Louviere and Woodworth, 1983). Every shelf had 5 rows and 12 columns, i.e., a total of 60 slots. The experiment featured 48 brands in total. The average shelf carried 37 different brands. The minimum number of brands presented on one shelf was 33 and the maximum 46.

All shelves presented to respondents were completely filled, i.e., did not have empty slots, and multiple facings of brands compensated for differences between the number of slots and the number of brands on a shelf. When a brand occupied multiple slots on a shelf, these slots were adjacent and horizontally aligned. The maximum number of slots allocated to one brand, i.e., the maximum number of facings was four.

Different respondents saw different sets of shelves. Overall the experiment comprised eight sets of 16 shelves each. Across the 8×16 choice-sets, variation in the number of facings on the set $\{1, 2, 3, 4\}$ is approximately orthogonal to brands, $\chi^2 = 25.75, df = 141$. Approximate orthogonality also holds for the relationship between the 60 shelf position and brands, $\chi^2 = 375.57, df =$

2773. Brand prices and volumes per pack are fixed within brands. The causal variables of interest are therefore a brands' number of facings on a shelf and its position. Overall 2300 consumers participated in this experiment. For computational reasons, our results are based on a randomly selected subset of 700 respondents. On average, sample respondents chose 2.32 different brands from a shelf with a standard deviation of 2.00. The average quantity, i.e., total number of packs chosen from a shelf is 3.08 with a standard deviation of 3.14.

4 Reduced Form Analysis

In the context of our illustrative application, the management question is if and how a brand's position and number of facings, the 'shelf-control variables', influence demand for that brand. Under the primitive assumption that respondents engage in constrained utility maximization conditional on awareness when choosing from a shelf, a researcher trying to measure the influence of these variables has a number of choices.

First, he could assume, preferably based on theory, that the shelf-control variables only influence respondents' attention to brands. Second, he could assume away the relevance of attention processes in this experimental context and postulate that respondents always attend to all brands presented. Under this assumption, the shelf-control variables need to be included as arguments to the (indirect) utility function. Third, he could assume that the shelf-control variables both influence attention to and the indirect utility from brands.

Up to prior theoretical arguments these choices are subjective. Obviously, the researcher could commit to (sets of) specific parametric models of attention and indirect utility, and use model comparisons to identify the causal mechanism by which the shelf-control variables affect demand.[11] We do not intend to criticize this approach. However, we illustrate in the following how the quantity choices we observe in our experimental data facilitate the direct identification of basic model features.

We emphasize that the identification argument we illustrate here does not hinge on the availability of experimental data, although approximate orthogonality between brands and shelf-control variables certainly helps with econometric identification and statistical efficiency.

Following the theoretical development presented in Section 2, we investigate the shelf-control variables as potentially shifting demand marginal with respect to attention, and as potentially shifting strictly positive demand. Remember that strictly positive demand is conditional on attention to the brands which exhibit positive demand, by definition.

[11] We present the results of one such approach in Section 7.

Without loss of generality, we characterize demand marginal to attention as the binary outcome of 'not choosing a brand from a shelf given its presence on the shelf' versus 'choosing any positive quantity of the same brand given its presence on the shelf'. The switch from zero to positive demand for a brand can be brought about by increasing attention to the brand or by increasing its indirect utility given attention. Therefore the distribution of these binary outcomes across brands on a shelf is marginal to attention.[12] Variation in strictly positive demand, in contrast and up to the assumption of conditional independence between attention and utility, directly points to variation in indirect utility because it occurs conditional on attention.

To learn about the mechanism by which the shelf control variables shift demand, if at all, we specify two regression equations. The first equation is a probit model that relates the collection of binary outcomes: 'not choosing a brand from a shelf given its presence on the shelf' versus 'choosing any positive quantity of the same brand given its presence on the shelf', to brand specific dummy variables, dummy variables for the number of facings of a brand, and a flexible parametric function of brand position on the shelf. We also include the number of different brands present on a shelf as control variable.

[12] Under the implicit assumption that attention processes are relevant at all, i.e., that a non-degenerate distribution of attention generated the data.

Positions are coded as signed horizontal and vertical distances from the shelf center. Deviations towards the right and upwards (towards the left and downwards) from the center are coded as positive (negative). When a brand occupied more than one slot on the shelf, i.e., was present with more than one facing, one of the corresponding slots was randomly chosen as its position.

The second equation is a linear model that regresses variation in strictly positive demand on the same set of covariates. We construct the dependent variable for this regression by omitting all the 'zero' observations from the probit regression and replacing the 'ones' with the actually observed positive quantities.

Table 1 summarizes the coefficients for the shelf control variables and Table 2 those for brand dummy variables. We use a flexible bi-variate second order polynomial function to capture the joint influence of signed vertical and horizontal distances from the center of the shelf with coefficients $\theta_v, \theta_h, \theta_{vh}, \theta_{v^2}, \theta_{h^2}, \theta_{vh^2}$, and θ_{v^2h}. The main effect of the number of facings is captured by dummy coding with coefficients $\theta_2, \theta_3, \theta_4$, where 'one facing' forms the base category. The coefficient θ_{nB} measures the effect of the number of brands on the shelf which is coded linearly.

The key result from Table 2 is that shelf control variables influence the probability of choosing a positive quantity from a brand versus choosing nothing, but do not influence positive quantities. Increasing the number of

Table 1: Regressing choice (probit) and strictly positive quantities (linear regression) on shelf characteristics. * denotes significance at the 10%, and ** at the 5% level.

Par	Choice(Probit)	Quantity(Linear)
θ_v	-0.00092	0.0025
	(0.0091)	(0.0109)
θ_h	0.01884**	0.0028
	(0.0021)	(0.0025)
θ_{vh}	0.00573**	0.0016
	(0.0018)	(0.0021)
θ_{v^2}	0.01954**	-0.0007
	(0.0045)	(0.0057)
θ_{h^2}	0.00276**	0.0004
	(0.0005)	(0.0006)
θ_{vh^2}	-0.00753**	-0.0008
	(0.0009)	(0.0012)
θ_{v^2h}	-0.00127**	-0.0001
	(0.0004)	(0.0004)
θ_2	0.15118**	0.015
	(0.0086)	(0.0105)
θ_3	0.23004**	0.0240
	(0.0128)	(0.0157)
θ_4	0.27030**	0.0290
	(0.0234)	(0.0262)
θ_{nB}	0.01079**	-0.0024*
	(0.0011)	(0.0013)

Table 2: Regressing choice (probit) and strictly positive quantities (linear regression) on shelf characteristics-brand specific constants. * denotes significance at the 10%, and ** at the 5% level.

Par	Choice(Probit)		Quantity(Linear)			Choice(Probit)		Quantity(Linear)	
	Mean	S.D.	Mean	S.D.		Mean	S.D.	Mean	S.D.
B_1	-2.18014**	0.05070	1.2146**	0.06041	B_{25}	-2.25342**	0.05436	1.2084**	0.06405
B_2	-2.25495**	0.06524	1.3368**	0.07858	B_{26}	-2.14762**	0.05217	1.2560**	0.06252
B_3	-1.84961**	0.04796	1.2554**	0.05730	B_{27}	-2.37269**	0.05324	1.2106**	0.06438
B_4	-1.85169**	0.04880	1.2580**	0.05583	B_{28}	-2.01303**	0.05074	1.2649**	0.05920
B_5	-2.59103**	0.05761	1.1451**	0.07324	B_{29}	-2.59956**	0.05971	1.1766**	0.07689
B_6	-2.40150**	0.05209	1.2663**	0.06545	B_{30}	-1.89273**	0.04977	1.2149**	0.05697
B_7	-2.55181**	0.05482	1.1789**	0.07041	B_{31}	-2.10283**	0.05291	1.4028**	0.06117
B_8	-1.93419**	0.04887	1.2234**	0.05724	B_{32}	-2.15080**	0.05525	1.1904**	0.06441
B_9	-2.33416**	0.05173	1.2023**	0.06189	B_{33}	-1.92231**	0.05152	1.2947**	0.06059
B_{10}	-2.26305**	0.05598	1.1673**	0.06287	B_{34}	-1.89680**	0.04966	1.2261**	0.05807
B_{11}	-2.03346**	0.04911	1.2582**	0.05708	B_{35}	-1.94264**	0.05079	1.2793**	0.05952
B_{12}	-1.60334**	0.05764	1.2649**	0.06444	B_{36}	-2.33240**	0.05180	1.2543**	0.06247
B_{13}	-1.86016**	0.04854	1.3884**	0.05700	B_{37}	-2.30464**	0.05094	1.2463**	0.06263
B_{14}	-2.02944**	0.04831	1.2388**	0.05842	B_{38}	-2.18482**	0.05659	1.2562**	0.06898
B_{15}	-2.34882**	0.05081	1.1667**	0.06109	B_{39}	-1.87121**	0.05015	1.3739**	0.05867
B_{16}	-2.23564**	0.04910	1.2033**	0.05858	B_{40}	-2.85373**	0.07874	1.3628**	0.11698
B_{17}	-2.02705**	0.04818	1.2642**	0.05747	B_{41}	-3.22208**	0.08496	1.3697**	0.12652
B_{18}	-2.15420**	0.04778	1.3240**	0.05789	B_{42}	-2.95957**	0.06391	1.2657**	0.08926
B_{19}	-2.36001**	0.05500	1.2218**	0.06589	B_{43}	-2.97640**	0.07039	1.2245**	0.09128
B_{20}	-2.04710**	0.04976	1.2513**	0.05830	B_{44}	-2.74078**	0.07480	1.2516**	0.07486
B_{21}	-2.08486**	0.04875	1.2591**	0.05872	B_{45}	-2.57492**	0.05932	1.2339**	0.07563
B_{22}	-2.12623**	0.05056	1.3026**	0.05856	B_{46}	-2.67466**	0.05950	1.3107**	0.07950
B_{23}	-1.97067**	0.04876	1.2267**	0.05767	B_{47}	-2.47339**	0.06261	1.3198**	0.07688
B_{24}	-2.27836**	0.05282	1.3832**	0.06313					

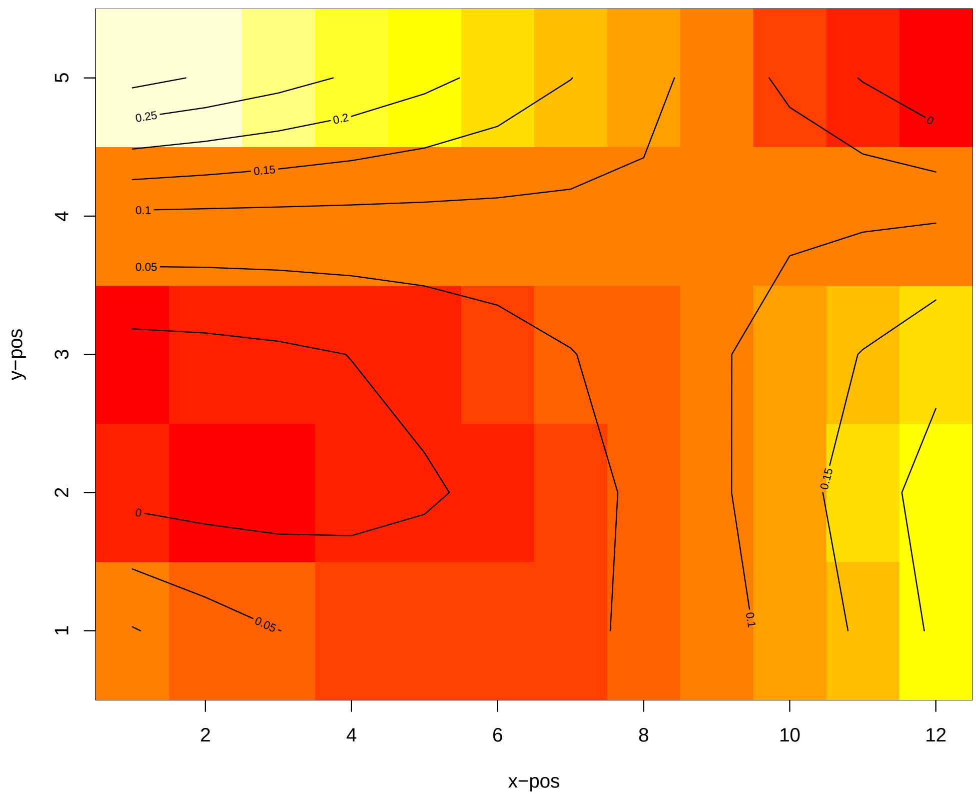

Figure 3: Heat-map illustrating the position effect on the probit scale

facings increases the probability of choosing a brand, however at decreasing marginal returns to additional facings. The position also has an effect on the probability of choosing a brand that we illustrate in the heat-map in Figure 4. When brands appear in the upper left or the lower right corner of the shelf, they are more likely to be chosen (see Figure 4). Finally, the coefficients in the last row of Table 1 indicate that a brand on a shelf which features a larger number of brands is more likely to be chosen but at a smaller quantity.

In isolation, the positive effect of shelf-facings on brand choice could occur even if responses were random. The position effect adds a habitual component, notwithstanding potentially random choice given this habit. However, distinct brand specific constants in the range of $[-3.22(0.08), -1.60(0.06)]$ in the probit model, and in the range of $[1.14(0.07), 1.40(0.06)]$ in the linear regression model for positive quantities imply that the representative consumer cares about differences between brands in this exercise, both in terms of choice probability and the optimal positive quantity given brand choice. Finally, the inverse relationship between quantity per brand and the number of brands on the shelf suggests that respondents allocate a fixed budget across fewer brands when there are fewer brands on the shelf which is strongly consistent with the assumption that respondents actually maximize utility when making their quantity choice decisions.

Given respondents' utility maximizing behavior and assuming conditional independence between attention and utility, the influence of shelf-control variables on choice probabilities and the lack thereof on positive quantities implies that the indirect utility function excludes shelf-control variables. Up to these assumptions, the shelf-control variables seem to contribute to demand only through facilitating awareness of a brand on the shelf. And the fact that we find two different, roughly independent variables that influence marginal demand but not strictly positive demand provides limited empirical support to assuming conditional independence between attention and utility similar to an over-identification test based on multiple instrumental variables (see Hausman, 1978).

The variation in brand specific constants in the probit model and in the linear regression model for positive quantities (see Table 2) imply that perceived differences between brands influence both demand marginal with respect to attention and strictly positive demand, i.e., demand conditional on attention. It follows that brand specific constants need to be included in the utility function, again assuming conditional independence between attention and utility. However, based on these results we cannot rule out the possibility that some brands are inherently better at attracting attention than other brands, in addition to providing differential utility conditional on attention. We leave this question for future research. Next we extend an existing model of conditional utility maximization by a model of attention and compare between full and limited attention models.

5 Evidence From a Parametric Model

We augment the multiple discrete-continuous extreme value (MDCEV) model introduced by Bhat (2005) by a model of attention to test our results in the context of a specific, and we believe reasonable, parametric setting. The MDCEV-model is based on a translated non-linear, additively separable utility function (see also Kim et al., 2002). The model contains the the standard single discrete-continuous model (e.g., Dubin and McFadden, 1984; Hanemann, 1984; Chiang, 1991; Chintagunta, 1993; Arora et al., 1998) and thus the standard multinomial-logit model as special cases.

5.1 The Utility Model

The utility function for brand j consists of a baseline utility $\psi(x_j)$, where x_j are brand characteristics, a measure of (diminishing) marginal returns to quantity q_j, $0 < \alpha_j \leq 1$, and a translation parameter γ_j. By the assumption of additive separability the total utility from a set of brands at quantities q_j is

$$U(q_0, q_1, \ldots, q_J) = \sum_j \psi(x_j)(q_j + \gamma_j)^{\alpha_j} \tag{1}$$

Diminishing marginal returns motivate the choice of more than one brand, and positive parameters γ_j rationalize corner solutions in the sense that some brands may not be chosen at all (see Kim et al., 2002).

We follow the standard approach to obtain a smooth likelihood function based on Equation 1 and assume that the baseline utility is subject to iid Type I extreme value error:

$$U(q_0, \ldots, q_j) = \sum_j \exp(\beta' x_j + \varepsilon_j)(q_j + \gamma_j)^{\alpha_j}. \tag{2}$$

Combining Equation 2 with the budget constraint $I = \sum_j q_j p_j$, where the index $j = 0$ refers to the outside good with price normalized to one, yields the Langrangian function:

$$\mathcal{L} = \sum_j \exp(\beta' x_j + \varepsilon_j)(q_j + \gamma_j)^{\alpha_j} - \lambda \left(\sum_j q_j p_j - I \right) \tag{3}$$

Finally, the quantity choices that maximize budget constrained utility, $\{q_j^*\}$, $j = 0, \ldots, J$, satisfy the Kuhn-Tucker conditions

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial q_j}(q_1^*, \ldots, q_J^*) &= 0, \quad \text{if } q_j^* > 0 \\ \frac{\partial \mathcal{L}}{\partial q_j}(q_1^*, \ldots, q_J^*) &< 0, \quad \text{if } q_j^* = 0 \end{aligned}$$

which in the context of Equation 3 take the following form:

$$p_j^{-1}\alpha_j \exp(\beta' x_j + \varepsilon_j)(q_j^* + \gamma_j)^{(\alpha_j - 1)} = \lambda, \quad \text{if } q_j^* > 0 \tag{4}$$

$$p_j^{-1}\alpha_j \exp(\beta' x_j + \varepsilon_j)(q_j^* + \gamma_j)^{(\alpha_j - 1)} < \lambda, \quad \text{if } q_j^* = 0. \tag{5}$$

Utilities are measured relative to an outside good with index $j = 0$ that we assume to be essential, i.e., γ_0 is set to zero. Hence, $\lambda = \alpha_0 \exp(\varepsilon_0)(q_0^*)^{(\alpha_0 - 1)}$. Substituting this expression for λ in Equations 4 and 5 and taking logarithms, the optimal quantities from (inside) brands $j = 1, \ldots, J$ can be expressed as

$$\begin{aligned} V_j + \varepsilon_j &= V_0 + \varepsilon_0, \quad \text{if } q_j > 0 \\ V_j + \varepsilon_j &< V_0 + \varepsilon_0, \quad \text{if } q_j = 0 \quad \text{where} \\ V_j &= \beta' x_j + \ln \alpha_j + (\alpha_j - 1)\ln(q_j^* + \gamma_j) - \log p_j. \end{aligned}$$

The expression is useful for generating data from the model, given parameters $\{\beta, \gamma, \alpha\}$, draws of the error terms, and a budget I by solving for the amount of money to be spent on the outside good such that the Kuhn-Tucker conditions hold, and the budget is exhausted.

Bhat (2005) showed that the probability density of data generated from this model can be expressed in closed form. Equation 6 is the probability density of choosing the positive quantities $q_0^*, q_1^*, \ldots, q_M^*$ from a subset of M

brands among J brands in the choice set.[13]

$$P(q_0^*, q_1^*, \ldots, q_M^*, 0, \ldots, 0) = \left[\prod_{i=0}^{M} c_i\right]\left[\sum_{i=0}^{M} \frac{1}{c_i}\right]\left[\frac{\prod_{i=0}^{M} e^{V_i}}{\left(\sum_{j=0}^{J} e^{V_j}\right)^M}\right] M!. \tag{6}$$

where c_i is defined as $\left(\frac{1-\alpha_i}{q_i^*+\gamma_i}\right)$.

Because we do not observe the budgets or equivalently the amount of money spent on the outside good in our data, we follow the standard approach of setting the budget equal to a respondent's maximal observed expenditure throughout the repeated measurements[14]. Next we extend this model that is conditional on attention to brands in the set $\{1, \ldots, J\}$ by a model of attention that determines the size and the composition of the set entering constrained utility maximization.

5.2 The Attention Model

We model attention to brands by a multivariate choice model. Specifically, we use the autologistic model (Besag, 1972, 1974) in Equation 7 to express the full conditional probability of paying attention to, or consid-

[13]The incoherence between the assumption of strictly continuous, i.e., infinitely divisible quantities and applications of the model to integer quantities is beyond the scope of this paper (see Lee and Allenby, 2014)

[14]For the choice sets with the respective maximum observed expenditures, we set the outside good consumption to 0.1, a relatively small value compared to the price of one unit of the inside goods.

ering brand j, $P(y_j = 1)$, given consideration of other brands in the set, $Y = (y_1, y_2, \ldots, y_J)'$. We only partially observe consideration sets $Y = (y_1, y_2, \ldots, y_J)'$ - we only know for sure that a brand was considered when we observe a positive quantity - and address the problem of unobserved attention to brands that were not chosen using data augmentation (see Section 6 and Section 9.1 in the Appendix for details).[15]

$$
\begin{aligned}
P(y_j = 1) &= \frac{\exp(C_j + \Phi_j Y)}{1 + \exp(C_j + \Phi_j Y)}, \quad \text{for } j = 1, \ldots, J \text{ with} \qquad (7) \\
\Phi_j &= (\phi_{j,1},\ \phi_{j,2},\ \ldots,\ \phi_{j,j-1},\ 0,\ \phi_{j,j+1},\ \ldots,\ \phi_{j,J})
\end{aligned}
$$

The 'consideration index' C_j in Equation 7 is a parametric function of covariates that influence attention, i.e., brand $j's$ position and number of facings in our example. The parameters in Φ_j account for inherently limited information processing (Miller, 1956). A negative value $\phi_{i,j}$ implies that paying attention to brand i inhibits attention to brand j. Because having paid attention to brand j cannot cause itself, the corresponding elements are fixed to zero, i.e., $\phi_{j,j} = 0 \quad \forall j \in \{1, \ldots, J\}$.

For a parsimonious specification and the lack of more detailed theory, we

[15] The consideration probability of the outside good with index $j = 0$ is fixed at 1.

impose the following equality constraints on the elements in Φ_j:

$$\phi_{i,j} = \phi_{k,l} = -\exp(\phi^*) \quad \forall i, j, k, l \in \{1, \ldots, J | i \neq j \text{ and } k \neq l\}$$

such that inhibitory effects are identically symmetric across all brands in all choice sets. Very large values ϕ^* imply that attention is essentially limited to one brand because of limited information processing capacity. Very small values ϕ^* imply that there are no capacity constraints.

Anticipating an eventually fully heterogeneous version of our model, we choose the following, relative to those in Section 4 more parsimonious, functional forms for the consideration index C_j. Brand $j's$ number of facings denoted by ζ_j enter linearly and brand $j's$ position as a bi-variate first order polynomial:

$$C_j = \delta_0 + \delta_1 \zeta_j + \delta_2 v_j + \delta_3 h_j + \delta_4 v_i h_i. \tag{8}$$

Positions are again coded as signed horizontal. h, and vertical, v, distances from the shelf center. Deviations towards the right and upwards (towards the left and downwards) from the center are coded as positive (negative). When a brand occupied more than one slot on the shelf, i.e., was present with more than one facing, one of the corresponding slots was randomly chosen as its position. The coefficient δ_0 measures a baseline level of attention. Next we sketch Bayesian inference for a hierarchical, fully heterogeneous version of our joint model of attention and constrained utility maximization. We report the details in the Appendix.

6 Bayesian Estimation

Conditional on consideration sets, we use Metropolis-Hastings steps to draw from the full conditional distributions of respondent $n = 1, \dots, N$ specific parameter vectors $\beta_n = (\beta_{n,1}, \dots, \beta_{n,J})$ and $\alpha_n = (\alpha_{n,0}, \dots, \alpha_{n,J})$ defined, up to a constant of proportionality, by the product of the likelihood in Equation 6 and the hierarchical priors for β_n and α_n.[16]

The hierarchical prior for β_n is multivariate normal with standard weakly informative subjective prior distributions (Rossi et al., 2005). That for α_n is non-standard, to simultaneously allow for flexibility across different brands and efficient pooling of information across respondents. We found this prior to improve on a log-normal prior in combination with a logit transformation in situations where respondents are heterogeneous in $\alpha_{n,j}$.

We divide the unit interval into ten equally spaced intervals $\{(0, 0.1]$, $(0.1, 0.2]$, ..., $(0.9, 1]\}$ and specify a subjective Dirichlet prior for the hier-

[16]Because the outside good forms the baseline for utility comparisons $\beta_{n,0}$ is fundamentally unidentified and fixed at zero. The translation parameter γ_0 for the outside good is fixed at zero, i.e., the outside good is treated as essential, and the remaining elements of the parameter vector γ_n fixed at 1 for empirical identification (Kim et al., 2002, 2007)

archical prior probabilities that an $\alpha_{n,j}$ is from a particular interval, $p_\alpha \sim$ Dirichlet$(1,\ldots,1)$. Given the interval, the prior for $\alpha_{n,j}$ is uniform on the interval. The subjective Dirichlet prior is updated for each brand separately, resulting in brand specific hierarchical prior probabilities over the 10 intervals a posteriori.

Again conditional on consideration sets, we use Metropolis-Hastings steps to draw from the full conditional distributions of respondent specific parameters $\delta_n = (\delta_{n,0}, \delta_{n,1}, \delta_{n,2}, \delta_{n,3}, \delta_{n,4})$ and ϕ_n^* in the attention model. We construct the full conditional distribution, up to a constant of proportionality, as the product of full conditional attention probabilities given in Equation 7 times the hierarchical prior density.[17] We use a standard multivariate normal hierarchical prior distribution with weakly informative subjective priors here and allow for hierarchical prior dependence between parameters in the utility function and those in the model of attention.

Finally, we augment the unobserved part of the consideration set. The full conditional attention probabilities in Equation 7 fully characterize a proper distribution of all consideration sets (Besag, 1972). Unfortunately, evaluating

[17]We note that this only results in a large sample consistent approximation to the full conditional posterior, because the product of full conditional attention probabilities in Equation 7 only defines a pseudo-likelihood (Besag, 1975). We leave exact Bayesian inference for this (sub-)model to future research. However, because the marginal posterior of parameter sets $\{\delta_n\}$ and $\{\phi_n^*\}$ is marginal with respect to unobserved attention to (large sets of) brands that were not chosen, we believe that 'overconfidence' from inference based on the pseudo-likelihood is negligible in our application.

this distribution for a particular consideration set is beyond reach for choice sets of the size encountered in our application.[18] However, in the context of our model, this unwieldy distribution is a prior distribution for the set of considered brands which is an argument to the likelihood of the observed quantity choices in Equation 6. And the full conditional attention probabilities in Equation 7 define a Gibbs-sampler to generate from this distribution. Thus, setting the proposal distribution of consideration sets equal to the prior distribution, we arrive at a simple Metropolis-Hastings update with acceptance probabilities defined by ratios of likelihoods defined in Equation 6 (see Section 9.1 in the Appendix for details).

7 Results

We compare three full-information models M1, M2, and M3 that assume attention to all brands on a shelf a priori to three limited information models M4, M5, and M6 that include our model of brand selection based on attention before utility maximization. Models M1 and M4 ignore shelf control variables, i.e., a brand's number of facings on the shelf and its position all together. Models M2 and M5 include a brand's number of facings, the former as a source of utility and the latter as a cause to attention. Finally, models

[18] For a shelf featuring 37 brands, for example, the normalizing constant of this distribution consists of $2^{37} \approx 1.37 \times 10^{11}$ terms.

M3 and M6 include both a brand's number of facings and its position, again the former as sources of utility and the latter as causes to attention (see Table 3). When we include shelf-control variables as sources of utility, we include them as covariates to a brand's baseline utility $\psi(x_j)$ (see Equations 1 and 2).

Table 3: Model Comparison

	Utility Index			Consideration Index			LML
Model	nF	po	Brand	Const.	nF	po	
M1			x				-86378.59
M2	x		x				-88522.53
M3	x	x	x				-100162.3
M4			x	x			-80359.76
M5			x	x	x		-73462.16
M6			x	x	x	x	-83931.53

The rightmost column in Table 3 reports log-marginal likelihoods for the different models (see Section 9.2 in the Appendix for computational details).[19] Models assuming full information, M1-M3, fit the data much worse than models that account for a selection of brands based on attention, M4-M6. The best full information model, M1, excludes shelf control variables and fits the data much worse than the worst limited information model M6. It appears that including shelf control variables as sources of utility takes away from the explanatory power of the full attention model. The best lim-

[19]Marginal likelihoods automatically penalize for differences in model complexity and can be directly compared across non-nested models. Assuming equal subjective prior model probabilities, ratios of marginal likelihoods define Bayes-factors (Rossi et al., 2005).

ited attention model, M5, includes a brand's number of facings as a cause to attention, but excludes position effects from the model. Apparently, the additional complexity from including the bi-variate first-oder polynomial we use to code a brand's position is no longer supported by the data, once heterogeneity is taken into account.

Table 4 summarizes posterior moments of the population mean effects of shelf control variables and of selected brand specific constants. Comparing the full information models M2 and M3 to the limited information models M5 and M6 we see a statistically credible change in the sign of the number of facings coefficient, δ_1. Both full information models M2 and M3 imply that increasing a brand's number of facings decreases this brand's utility, i.e., depresses demand for this brand overall. In contrast, both limited attention models M5 and M6 imply that increasing a brand's number of facings increases the probability that this brand is considered, i.e., enters utility maximization to the effect of increasing demand for this brand overall.[20] We also find a statistically credible sign change for the main effect of a brand's horizontal position comparing the coefficient δ_3 across models M3 and M6.

Across all limited information models M4 - M6, the posterior mean of ϕ^* suggests the relevance of inherent capacity constraints to attention such that

[20]The posterior distribution of the number of facings coefficients, $\delta_{1,n}$ in the utlity function under M2 (in the attention function under M5) has more than 90% (80%) of its mass in the negative (positive) domain (see Table 8 in the Appendix).

attention to brand i inhibits attention to brand j. In the best fitting model overall, M5, the size of the inhibitory effect from two additional brands in the consideration set at the posterior mean of ϕ^* roughly equals the positive effect of one additional facing on attention.

Table 6 reports posterior moments of the population means of $\alpha's$, that measure the rate at which marginal returns to quantity diminish.[21] Limited information models M4-M6 infer that utility from the outside good with index $j = 0$ diminishes faster than under full information (M1-M3). Apparently, the full information models need a large α_0, i.e., slowly diminishing marginal returns from the outside good, to help rationalize the choice of only a few brands from large sets of brands.

Finally Table 7 reports population covariances and correlations between the base line utilities of selected brands and parameters in the attention index. We find only small correlations between parameters δ_0 and δ_1 in the attention index and the brands' base line utilities, β. However, for some brands we find sizeably positive correlations between base line utilities β and ϕ^*, suggesting that respondents with high baseline preferences for these brands are more attention constrained and thus have smaller consideration sets than other respondents.

[21] We compute these values directly from the marginal posterior of random effects. An alternative approach computes the means implied by the brand-specific hierarchical multinomial prior over uniform α-bins. The corresponding population variances are reported in Table 9 in the Appendix.

Table 4: Posterior means of population mean parameters $\bar{\delta}, \bar{\beta}, \bar{\phi}^*$. Posterior standard deviations in parentheses.

Par	M1	M2	M3	M4	M5	M6
δ_0				2.4295	0.5001	0.5005
				(0.24)	(0.00)	(0.01)
δ_1		-0.5014	-0.9406		0.3351	0.4531
		(0.02)	(0.02)		(0.02)	(0.03)
δ_2			-0.6763			-0.2811
			(0.01)			(0.02)
δ_3			-0.0592			0.0592
			(0.00)			(0.01)
δ_4			-0.0211			-0.0564
			(0.00)			(0.00)
β_1	-5.3756	-1.8816	-2.1036	-10.1095	-5.5372	-0.7991
	(0.15)	(0.07)	(0.10)	(0.81)	(0.27)	(0.10)
β_{10}	-4.0166	-1.3251	-1.0845	-3.4835	-3.6387	-0.3745
	(0.11)	(0.06)	(0.08)	(0.33)	(0.20)	(0.10)
β_{20}	-2.5355	-1.3739	-1.6635	-2.7957	-1.6658	-0.3538
	(0.07)	(0.07)	(0.08)	(0.26)	(0.11)	(0.09)
β_{30}	-2.2924	-0.9415	-1.6796	-1.6213	-1.3788	-0.6449
	(0.09)	(0.07)	(0.07)	(0.22)	(0.11)	(0.08)
β_{40}	-3.0492	-0.9889	-1.2378	-4.3277	-3.8041	-0.7318
	(0.08)	(0.06)	(0.08)	(0.36)	(0.13)	(0.10)
ϕ^*				-0.8716	-1.8475	-0.2506
				(0.07)	(0.07)	(0.04)

Table 5: Posterior means of population variances in $\delta_n, \beta_n, \phi_n^*$. Posterior standard deviations in parentheses.

Par	M1	M2	M3	M4	M5	M6
δ_0				35.8941	0.0010	0.0005
				(2.69)	(0.00)	(0.00)
δ_1		0.1021	0.3604		0.4453	0.5529
		(0.01)	(0.02)		(0.03)	(0.03)
δ_2			0.0791			0.2345
			(0.00)			(0.01)
δ_3			0.0098			0.0318
			(0.00)			(0.00)
δ_4			0.0042			0.0150
			(0.00)			(0.00)
β_1	10.7440	2.7712	6.2837	382.2130	45.6233	6.9109
	(0.85)	(0.15)	(0.34)	(36.06)	(2.57)	(0.36)
β_{10}	6.2269	2.1769	3.9094	59.0138	22.8869	6.3702
	(0.37)	(0.14)	(0.23)	(4.46)	(1.25)	(0.33)
β_{20}	3.9774	3.2867	4.5025	40.5135	8.5287	4.8480
	(0.26)	(0.18)	(0.23)	(2.54)	(0.51)	(0.25)
β_{30}	4.9450	3.1603	4.4789	28.7196	7.3485	4.5930
	(0.29)	(0.20)	(0.24)	(1.77)	(0.39)	(0.29)
β_{40}	2.9068	1.8948	3.3254	77.1871	14.5966	6.4159
	(0.17)	(0.10)	(0.19)	(4.92)	(0.80)	(0.34)
ϕ^*				3.2614	2.8614	1.1990
				(0.19)	(0.15)	(0.07)

Population variances of these effects are reported in Table 5. Comparing the best fitting full attention model M1 to the overall best fitting model M5, we see that the latter infers relatively more heterogeneity in the brand specific utility constants β. In the full information model, systematic variance in attention to one and the same brand, because of changes in the shelf layout, can only be rationalized through shocks to the utility function. These shocks increase the error variance, i.e., lower the signal to noise ratio and thus make it harder to recover heterogeneous tastes. In the limited attention model without observed drivers of attention, M4, the posterior variances are even more extreme. However, because this model has to do without covariates to attention, it fails to clearly identify a posterior mode between the extremes of full attention to all brands (δ_0 large), and attention to only the chosen brands (δ_0 small), both at the individual and the population level thereby inflating marginal heterogeneity in brand specific utility constants, β.

Table 6: Posterior means of $\bar{\alpha}$. Posterior standard deviations in parentheses.

Par	M1	M2	M3	M4	M5	M6
α_0	0.7161	0.7716	0.7884	0.5588	0.6080	0.6779
	(0.15)	(0.15)	(0.13)	(0.14)	(0.17)	(0.19)
α_1	0.3911	0.2214	0.4517	0.3072	0.2797	0.5693
	(0.09)	(0.09)	(0.18)	(0.18)	(0.14)	(0.27)
α_{10}	0.1807	0.1102	0.1431	0.3117	0.1367	0.3114
	(0.05)	(0.07)	(0.10)	(0.15)	(0.12)	(0.26)
α_{20}	0.1536	0.1580	0.2050	0.2988	0.1618	0.3439
	(0.06)	(0.06)	(0.10)	(0.18)	(0.08)	(0.24)
α_{30}	0.1514	0.1431	0.1808	0.2639	0.2048	0.3587
	(0.05)	(0.05)	(0.09)	(0.12)	(0.10)	(0.20)
α_{40}	0.0853	0.0520	0.6199	0.5835	0.2294	0.7336
	(0.04)	(0.03)	(0.18)	(0.24)	(0.18)	(0.22)

Table 7: Posterior mean of selected population covariances (upper diagonal) and correlations (lower diagonal) in M5

Parameter	δ_0	δ_1	β_1	β_{10}	β_{20}	β_{30}	β_{40}	ϕ^*
δ_0	0.0010	0.0012	-0.0023	0.0031	0.0039	0.0031	-0.0028	0.0003
δ_1	0.0565	0.4453	-1.0190	-0.0080	-0.1354	0.0172	0.4341	-0.1146
β_1	-0.0108	-0.2260	45.6233	14.2461	7.9677	7.0470	6.9879	7.1465
β_{10}	0.0202	-0.0025	0.4408	22.8869	4.0736	3.2815	6.7967	2.3604
β_{20}	0.0420	-0.0694	0.4038	0.2915	8.5287	2.0910	2.4565	2.1316
β_{30}	0.0373	0.0094	0.3846	0.2531	0.2641	7.3485	3.4234	1.4947
β_{40}	-0.0233	0.1703	0.2704	0.3716	0.2202	0.3304	14.5966	1.1529
ϕ^*	0.0052	-0.1017	0.6253	0.2915	0.4312	0.3258	0.1781	2.8614

8 Discussion

In this paper we show how observing quantity choices facilitate the identification of a selection step that determines which of the brands available in a choice environment enter utility maximization. We first develop how the information content in quantity choices exceeds that in multinomial choices and compare the information in quantity choices to independent observations of a selection that precedes utility maximization. We find that quantity choices directly identify drivers of demand that operate before utility maximization, and thus have to be excluded from the utility function, assuming conditional independence. We also find that quantity choices identify conditional dependence when exclusion restrictions are known a priori. Both results suggest a useful information advantage from quantity choices relative to observing only multinomial outcomes.

The distinction between utility maximization and a preceding selection step is related to the differentiation between informative and persuasive effects of advertising (e.g. Ackerberg, 2001).[22] Actions that only inform consumers about the availability of a particular brand and therefore only deter-

[22] Ackerberg (2001) successfully identifies advertising for a newly introduced brand of yogurt as informative about the availability of its inherent product characteristics only, i.e., as not having an image, prestige, or generally persuasive effect by contrasting the effects on more and less experienced users. Here we follow his example and use the term informative in the sense of informative about the availability of a particular alternative and not in the sense of (additional) information about this alternative's features or characteristics (c.f. Johnson and Myatt, 2006)

mine the probability of choosing a positive quantity from that brand need to be excluded from the utility function that jointly determines if and how much to optimally choose from this brand conditional on awareness. If some action also influences the optimal quantity choice, conditional on awareness, it qualifies as persuasive and needs to be included in the utility function. Therefore, observed quantity choices should be useful in the context of measuring advertising effects.[23]

In the context of our illustrative application, the research question is if the number of facings and the position of a product on a shelf only help alert respondents to the presence of the product on the shelf or additionally connote utility, i.e., persuade respondents. We find strong empirical support in favor of informative effects, and contradicting persuasive effects.

An implication of full information models is that more variety weakly increases consumers' utility. Under full information, the limited provision of variety we see in the real world has to be motivated by supply side constraints only, such as e.g., the fixed costs associated with creating, producing and offering additional variety. The limited attention model explored in this

[23]Terui et al. (2011) investigate direct, persuasive effects and awareness effects of advertising in a multinomial probit choice model augmented by a consideration stage. They find support for an indirect, awareness effect of advertising on consideration sets in a fully parametric setting, and assuming conditional independence between consideration and utility maximization. Similarly, van Nierop et al. (2010) find shelf space, display, and feature advertising to influence the consideration stage, again in a fully parametric model and assuming conditional independence between consideration and utility maximization (see also Bronnenberg and Vanhonacker, 1996).

paper contributes a demand side motivation for the limited provision of variety we see in practice, with potentially interesting implications for retailers' assortment decisions and competition.

If consumers' attention is inherently limited, consumers will fail to realize the full set of substitution possibilities among the different offers in a particular market. This failure translates into market power for brands, even if they are poorly differentiated under full information. In this way, limited information can rationalize a manufacturer's decision to offer more of the same variety that may already exist in a market under a different brand name, instead of contributing to increasing the total variety supplied in market.

Moreover, the action of alerting consumers to the presence of a brand often cannot be targeted well, such as e.g., investing in more shelf facings at a large, national retailer. Now, if a brand competes for consumers' inherently limited attention, by definition, at the cost of attention to other brands the retailer may be offering, horizontally differentiated 'niche' offerings face a greater challenge than under full information. First, the attention generated through additional, costly shelf facings only translates into purchases by a smaller subset of customers, by the definition of a niche offering. Second, the attention drawn to the brand depresses demand for other brands simply because attention is limited and therefore other, potentially more suitable offerings may not have received attention.

The flipside of this argument is that a manufacturer who can invest in attention to his brand, simultaneously decreasing attention to other brands because of consumers' limited bandwidth, will benefit relatively more from providing a less differentiated, 'mainstream' offering than under full information. A formal analysis of the implications of inherently limited consumer attention for competitive market outcomes is however beyond the scope of this paper and left for future research.

9 Appendix

9.1 Augmentation of Consideration Sets

Equation 7 expresses the full conditional (prior) probability that a respondent attends to a particular brand in a choice set. The set of recursively updated full conditional distributions for all brands define a Gibbs sampler for consideration sets. Obviously, brands chosen by a respondent are in the consideration set with probability 1. Therefore we can partition the consideration set $\{y_i\}_{i=1}^{J} = Y$ into an observed part Y^o comprising brands chosen by a respondent, and an unobserved part Y^* indicating if the remaining, non-chosen brands received attention or not for each choice set, $Y = Y^o \cup Y^*$.[24] We propose the unobserved part of the consideration set, conditional on δ,

[24] Respondents always attend to the outside good with index $j = 0$.

ϕ^*, and Y^o by Gibbs-sampling from the full conditional probabilities defined in Equation 7. The consideration set Y is an argument to the likelihood function $\mathcal{L}(q|\beta, \alpha, \gamma, Y)$ defined in Equation 6. The proposed move from Y^* to $(Y^*)'$ is then accepted with probability

$$\alpha = \min\left\{1, \frac{\mathcal{L}(q|\beta, \alpha, \gamma, Y^o \cup (Y^*)')}{\mathcal{L}(q|\beta, \alpha, \gamma, Y^o \cup Y^*)} \times \frac{\Pi((Y^*)'|\delta, \phi^*, Y^o)}{\Pi(Y|\delta, \phi^*, Y^o)} \times \frac{Q(Y)}{Q((Y^*)')}\right\},$$

where $\mathcal{L}(q|\beta, \alpha, \gamma, Y)$ is defined in Equation 6 and $\Pi(Y^*)$ and $Q(Y^*)$ denote prior and proposal densities of the consideration set's unobserved part Y^*, respectively. We noted earlier that computing the (joint) probability $\Pi(Y^*|\delta, \phi^*, Y^o)$ implied by recursively sampling from Equation 7 is practically impossible, when the number of (non-chosen) brands on a shelf is large. However, because we propose $(Y^*)'$ by recursively sampling from Equation 7, our proposal distribution $Q(Y^*)$ is equal to the prior $\Pi(Y^*|\delta, \phi^*, Y^o)$. As a consequence, the proposal and the prior distributions cancel in the Metropolis-Hastings ratio and the acceptance probability simplifies to

$$\alpha = \min\left\{1, \frac{\mathcal{L}(q|\beta, \alpha, \gamma, Y^o \cup (Y^*)')}{\mathcal{L}(q|\beta, \alpha, \gamma, Y^o \cup Y^*)}\right\}.$$

9.2 Computation of Marginal Likelihoods

A full attention model is that special case of a limited attention model that puts unit mass on the full consideration set, i.e.,

$$\mathcal{L}(q, Y^{full}) = \mathcal{L}(q|Y^{full})\Pi(Y^{full}) = \mathcal{L}(q|Y^{full}) \times 1 = \mathcal{L}(q|Y^{full}).$$

A limited attention model acknowledges that all we learn from the data directly is that the brands that were chosen from a choice set were part of the consideration set, Y^o. However, the respondent may have paid attention to additional brands that form the unobserved part of the consideration set Y^*. The unobserved parts of the consideration set are parameters that need to be integrated out to arrive at the marginal density of what is observed, i.e., q and Y^o:

$$\int \mathcal{L}(q, Y^o \cup Y^*)dY^* = \int \mathcal{L}(q|Y^o \cup Y^*)\Pi(Y^o \cup Y^*)dY^*$$

$$= \int \mathcal{L}(q|Y^o \cup Y^*)\Pi(Y^*|Y^o)\Pi(Y^o)dY^*.$$

$$= \mathcal{L}(q|Y^o)\Pi(Y^o)$$

Therefore, full attention models can be compared to limited attention models based on comparing $\mathcal{L}(q|Y^{full})\Pi(Y^{full}) = \mathcal{L}(q|Y^{full})$ to $\mathcal{L}(q|Y^o)\Pi(Y^o)$.

We noted earlier that computing $\Pi(Y)$ is beyond practical reach for larger sets of brands. When computing $\Pi(Y^o)$, we take advantage of the fact that the number of brands chosen from a particular choice set, i.e., shelf, tends to be small.

Using the basic identity (Rossi et al., 2005),

$$\frac{1}{\mathcal{L}(q, Y^o)} = \int \frac{g_1(\alpha, \beta) g_2(\delta, \phi^*) g_3(Y^* | \delta, \phi^*)}{\mathcal{L}(q|\alpha, \beta, Y^o, Y^*) \Pi(Y^o | Y^*, \delta, \phi) \Pi(Y^* | \delta, \phi^*) \pi(\alpha, \beta, \delta, \phi^*)} \times p(\alpha, \beta, Y^*, \delta, \phi | q, Y^o) d\{\alpha, \beta, Y^*, \delta, \phi^*\}. \tag{9}$$

where $p(\alpha, \beta, Y^*, \delta, \phi | q, Y^o)$ is the joint posterior distribution, and setting $g_1(\alpha, \beta) g_2(\delta, \phi^*) = \pi(\alpha, \beta, \delta, \phi^*)$ and $g_3(Y^* | \delta, \phi^*) = \Pi(Y^* | \delta, \phi^*)$, Equation 9 simplifies to

$$\frac{1}{\mathcal{L}(q, Y^o)} = \int \frac{1}{\mathcal{L}(q|\alpha, \beta, Y^o, Y^*) \Pi(Y^o | Y^*, \delta, \phi^*)} \times p(\alpha, \beta, Y^*, \delta, \phi^* | q, Y^o) d\{\alpha, \beta, Y^*, \delta, \phi^*\}. \tag{10}$$

Next we explain how to compute $\Pi(Y^o | Y^*, \delta, \phi^*)$.

For the sake of illustration let's assume that Y^o contains only two brands,

A and B. Then

$$\Pi(Y^o|Y^*, \delta, \phi) = \frac{\exp(C_A + C_B + \{C_{Y^*}\} + Y'_{A,B,Y^*} ltrig(\Phi) Y_{A,B,Y^*})}{D} \tag{11}$$

where C_j is brand $j's$ consideration index defined in Equation 8, the subscripts in Y_{A,B,Y^*} indicate which brands are in the consideration set, and D is a normalizing constant defined as follows (Besag, 1972):

$$\begin{aligned} D = & \exp(\{C_{Y^*}\} + Y'_{Y^*} ltrig(\Phi) Y_{Y^*}) + \\ & \exp(C_A + \{C_{Y^*}\} + Y'_{A,Y^*} ltrig(\Phi) Y_{A,Y^*}) + \\ & \exp(C_B + \{C_{Y^*}\} + Y'_{B,Y^*} ltrig(\Phi) Y_{B,Y^*}) + \\ & \exp(C_A + C_B + \{C_{Y^*}\} + Y'_{A,B,Y^*} ltrig(\Phi) Y_{A,B,Y^*}) \end{aligned} \tag{12}$$

Thus, using MCMC draws of Y^* for integration in computing $\Pi(Y^o|\delta, \phi^*)$ from $\Pi(Y^o|Y^*, \delta, \phi^*)$ saves us from having to explicitly compute $\Pi(Y^*|\delta, \phi^*)$ and $\Pi(Y^o, Y^*|\delta, \phi^*)$ which is practically impossible given the algebraic structure of the normalizing constant in Equation 12.

Table 8: Comparison between the deciles of the means of posterior distribution of coefficients on number of facings (δ_0)

	Deciles								
Model	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
M2	-0.90	-0.76	-0.67	-0.59	-0.50	-0.43	-0.34	-0.25	-0.10
M5	-0.44	-0.10	0.11	0.24	0.37	0.51	0.69	0.83	1.05

Table 9: Posterior variances in α_n. Posterior standard deviations in parentheses.

Par	M1	M2	M3	M4	M5	M6
α_0	.0291	.0244	.0178	.0279	.0313	.0372
	(.0011)	(.0008)	(.0004)	(.0015)	(.0014)	(.0003)
α_1	.0216	.0145	.0357	.0496	.0248	.0791
	(.0010)	(.0011)	(.0006)	(.0026)	(.0009)	(.0006)
α_{10}	.0084	.0083	.0101	.0360	.0189	.0741
	(.0014)	(.0011)	(.0005)	(.0023)	(.0016)	(.0017)
α_{20}	.0089	.0062	.0113	.0414	.0096	.0634
	(.0013)	(.0012)	(.0004)	(.0033)	(.0008)	(.0006)
α_{30}	.0078	.0053	.0087	.0262	.0136	.0430
	(.0016)	(.0007)	(.0001)	(.0023)	(.0011)	(.0006)
α_{40}	.0063	.0018	.0356	.0835	.0413	.0523
	(.0016)	(.0006)	(.0005)	(.0034)	(.0016)	(.0008)

References

D. A. Ackerberg. Empirically distinguishing informative and prestige effects of advertising. *RAND Journal of Economics*, 32(2):316–333, 2001.

N. Arora, G.M. Allenby, and J. L. Ginter. A hierarchical Bayes model of primary and secondary demand. *Marketing Science*, 17:29–44, 1998.

J. Besag. Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society, B*, 34(1):75–83, 1972.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, B*, 36(2):192–263, 1974.

J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24(3): 179–195, 1975.

C. R. Bhat. A multiple discrete-continuous extreme value model formulation and application to discretionary time-use decisions. *Transportation Research Part B*, 39:679–707, 2005.

B. Bronnenberg and W. Vanhonacker. Limited choice sets, local price response, and implied measures of price competition. *Journal of Marketing Research*, 33(1):163–173, 1996.

B. J. Bronnenberg. The provision of convenience and variety by the market. *Available at SSRN 2326363*, 2013.

J. Chiang. A simultaneous approach to whether to buy, what to buy and how much to buy. *Marketing Science*, 4:297–314, 1991.

P. K. Chintagunta. Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Science*, 12:194–208, 1993.

M. Draganska and D. Klapper. Choice set heterogeneity and the role of advertising: an analysis with micro and macro data. *Journal of Marketing Research*, XLVIII:653–669, August 2011.

J. A. Dubin and D. L. McFadden. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, 52(2):345–362, 1984.

G. Fennell, G. Allenby, S. Yang, and Y. Edwards. The effectiveness of demographic and psychographic variables for explaining brand and product category use. *Quantitative Marketing and Economics*, 1:223–244, 2003.

W. M. Hanemann. Discrete/continuous models of consumer demand. *Econometrica*, 52(3):541–561, May 1984.

J. A. Hausman. Specification tests in econometrics. *Econometrica*, 46:1251–1271, 1978.

J. J. Heckman. Sample selection bias as a specification bias. *Econometrica*, 47:153–162, 1979.

J. P. Johnson and D. P. Myatt. On the simple economics of advertising, marketing, and product design. *The American Economic Review*, 96(3): 756–784, June 2006.

J. Kim, G. M. Allenby, and P. E. Rossi. Modeling consumer demand for variety. *Marketing Science*, 21:229–250, 2002.

J. Kim, G. M. Allenby, and P. E. Rossi. Product attributes and models of multiple discreteness. *Journal of Econometrics*, 138:208–230, 2007.

S. H. Lee and G. M. Allenby. Modeling indivisible demand. *Marketing Science*, 2014. Forthcoming.

J. J. Louviere and G. G. Woodworth. Design and analysis of simulated choice or allocation experiments: An approach based on aggregate data. *Journal of Marketing Research*, 20:350–367, 1983.

Marketing Economics Research Summit. Personal communication with Emir Kamenica. 02-01-2014.

G. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63:81–97, 1956.

P. E. Rossi, G. M. Allenby, and R. McCulloch. *Bayesian Statistics and Marketing*. John Wiley and Sons, 2005.

M. Sovinsky Goeree. Limited information and advertising in the U.S. personal computer industry. *Econometrica*, 76(5):1017–1074, 2008.

N. Terui, G. M. Allenby, and M. Ban. The effect of media advertising on brand consideration and choice. *Marketing Science*, 130(1):74–91, 2011.

E. van Nierop, B. Bronnenberg, R. Paap, M. Wedel, and P. H. Franses. Retrieving unobserved consideration sets from household panel data. *Journal of Marketing Research*, 47(1):63–74, 2010.

S. Wachtel and T. Otter. Successive sample selection and its relevance for management decisions. *Marketing Science*, 32(1):170–185, February 2013.