The perils of ignoring uncertainty in market simulations and product line optimization

Scott Ferguson
Department of Mechanical and Aerospace Engineering
North Carolina State University

Abstract

Quantitative market research models facilitate the creation of market simulators and the formulation of product line optimization solutions. Results from market simulators provide insight into how a population might respond to new product offerings, guiding decisions about product configuration and price. When physical product lines are created, the results from these simulations can also inform production and resource allocation decisions. The work presented in this paper highlights consequences of ignoring uncertainty associated with market-driven product line optimization problems, with a specific focus on parameter uncertainty. A two-objective optimization problem is introduced that maximizes revenue from the product line under a nominal model while also maximizing the worst case revenue from an uncertainty set of models. Here, the nominal model represents the mean of the posterior distribution of a hierarchical Bayes mixed logit model while the uncertainty set is represented by 800 draws from the posterior distribution. A third objective is also introduced that minimizes the variation of First Choice Share within the product line. The importance of this objective is demonstrated by illustrating the variation in share captured by each product when considering the models in the uncertainty set. This variation is discussed in the context of production and resource allocation decisions.

Introduction

Consider a manufacturer who is interested in creating a line of products for a heterogeneous market. The decision (design) variables for such a problem are product content (configuration) and product price. Configuration and pricing decisions can be informed by a market simulator that becomes the engine driving the product line optimization problem. Strategies for formulating and solving product line optimization problems have been presented at previous Sawtooth Software conferences [1–4], and even more references can be found in the literature [5–8]. These works have also shown that product line optimization problems are challenging for even modern optimization algorithms because they have large design spaces (billions or more possible combinations) and gradient-based optimization techniques cannot be used because of mixed-integer problem formulations.

The business objective for product line optimization problems is often revenue maximization, but the value of using objectives related to share of preference, profit, and commonality has also been demonstrated [4]. Once a solution has been found, decisions are made about product configuration, price, and production quantities. These outcomes are significant; manufacturers must order parts, design and construct assembly lines, and negotiate for shelf space. As noted by Bertsimas and Misic, product production decisions are both infrequent and require a commitment of manufacturer resources in a way that “cannot be easily reversed or corrected” [9].

There are many sources of uncertainty that, if not considered when solving the optimization problem, can translate to product line solutions with disastrous market performance. As discussed in [9], at least two forms of uncertainty can be associated with the choice model: structural and parameter. Structural uncertainty can be thought of as demand model misspecification [10–12]. Parameter uncertainty is related to the model parameter estimates – including, but not limited to, part-worth values and segment probabilities. Additionally, uncertainty exists when considering competitor product configurations and prices, and the manufacturer's own product attributes and component costs. In this paper, the focus is on the uncertainty in parameter estimates.

Optimization studies utilizing a single set of part-worth coefficient point estimates per respondent (such as the mean of the lower-level posterior distribution in a hierarchical Bayes mixed logit model) benefit from reduced computational cost. However, they neglect how the reported objective function is impacted by parameter uncertainty. Recognizing the potential hazards of using a single set of point estimates when simulating market behavior, especially if used to inform resource allocation decisions, researchers have proposed simulation strategies using draws from the posterior distribution, randomized first choice [13], interval variables, and moment estimation.

Building on these efforts, a robust revenue optimization approach has been introduced by Bertsimas and Misic that maximizes the worst case revenue of the product line under uncertainty. The work in this paper expands on their approach by reformulating the optimization problem as one with multiple objectives. The first objective maximizes overall revenue given a "nominal" model, while the second objective maximizes worst case revenue from an uncertainty set (of models). Realizing that the solution will also drive product inventory and manufacturing decisions, this paper introduces a third objective that considers the variation in choice amongst the products *within the product line*.

The approach presented in this paper is important because it highlights the value forfeited when uncertainty is ignored in product line optimization problems. By reformulating the optimization problem with multiple objectives, a decision maker can develop a richer understanding of the tradeoffs (and risk) associated with different product line solutions. This work also demonstrates the inherent value of quantitative market research models and market simulators throughout the many stages of the design process.

Description of relevant literature

The papers listed in Table 1 provide a representation of how uncertainty has been addressed in recent product design literature. As stated in the previous section, these methods use draws from a posterior distribution, interval variables, or moment estimation.

Camm et al. [14] and Wang et al. [7] use samples from the posterior distribution and introduce post-optimality robustness tests that assess the negative impact of part-worth uncertainty. In [14], individual draws are used so that the deterministic optimization problem can be repeatedly solved. The optimal product configuration was also found using part-worth coefficient point estimates. Resultant solutions were then compared, and the product configuration that maximized first choice share (FCS) when using point estimates aligned with only 23.5% of the random draw solutions. Wang et al. [7] implemented a sample average approximation method using stochastic discrete optimization [15]. Parameter uncertainty was modeled by pulling multiple draws from a respondent's posterior distribution. Each draw was then treated as a separate respondent, and the product line was optimized. Results from this study showed that as the sampling of the posterior distribution increased, the number of optimal products reduced.

Wang and Curry [16], Luo et al. [17], and Besharati et al. [18] defined part-worths using interval variables and investigated the best and worst cases of product utility. Wang and Curry [16] studied robustness in the share-of-choice problem by assuming that individual preferences were bounded, independent, and symmetric. Also, the covariance matrix for individual level part-worths was assumed to have a diagonal form, preventing correlation among product features. Luo et al. [17] and Besharati et al. [18] used segment-level part-worth confidence intervals and calculated the lower and upper bounds of product utility. Both studies only considered the design of a single product (rather than a line) but considered multiple design objectives; namely, maximizing the share of preference using the nominal model, minimizing variation in share of preference, and minimizing the worst case performance. Resende et al. [19] advanced these studies by considering a profit objective and estimated the first and second moments of the objective function by applying the delta method [20]. A closed-form solution was then introduced using a Taylor series expansion when considering a multinomial logit model at a pre-specified risk level.

Table 1. Recent literature considering parameter uncertainty when using market research models in product (line) optimization.

Reference	Method to treat uncertainty in discrete choice methods	Design problem	Design variables	Design objective
Camm et al. [14]	Samples from posterior distribution	A single product	Discrete product attributes	Maximize FCS
Wang and Curry [16]	Manual definition of part-worth intervals	A single product	Discrete product attributes	Maximize FCS
Luo et al. [17]	Interval estimates of part-worths using 95% confidence levels	A single product	Discrete product attributes	Maximize nominal SOP, Minimize SOP variance, Minimize worst-case performance
Besharati et al. [18]	Interval estimates of part-worths using 95% confidence levels	A single product	Discrete product attributes	Maximize nominal SOP, Minimize SOP variance, Maximize engineering design performance
Resende et al. [19]	Moment estimation of market share based on continuous probability function of part-worths	A single product	Continuous product attributes	Maximize profit at specified downside risk tolerance
Wang et al. [7]	Samples from posterior distribution	Product line	Discrete product attributes	Maximize FCS
Bertsimas and Misic [9]	Samples from posterior distribution	Product line	Discrete product attributes	Maximize worst-case expected revenue

FCS: First Choice Share SOP: Share of Preference

The recent publication by Bertsimas and Misic [9] most directly motivates the work in this paper. Product line robustness is explored by formulating an optimization problem that maximizes worst-case expected revenue over an uncertainty set, as shown in Equation 1.

$$\max_{S \subseteq \{1,\ldots,N\}:|S|=P} R(S;\mathcal{M}) \tag{1}$$

In this equation, R is revenue, S is a product line comprised of P products, and $\mathcal{M}$ is a set of choice models that account for parametric and structural uncertainty. Parametric uncertainty is considered for both the hierarchical Bayes mixed logit and latent class multinomial logit models. Structural uncertainty is represented in the latent class model by varying the number of segments.

The worst-case expected revenue for a product line is given by Equation 2, where $\widetilde{m}$ represents the choice model associated with the lowest expected per-customer revenue. Simulation results found that product line solutions that did not account for uncertainty experienced worst case losses as high as 23%. Conversely, a robust solution, using the formulations in Equations 1 and 2 could outperform a nominal solution (where it is assumed that the choice model is known precisely when the product line is optimized) by up to 14%.

$$R(S;\mathcal{M}) = \min_{\widetilde{m} \in \mathcal{M}} R(S;\widetilde{m}) \tag{2}$$

It is also discussed in [9] that the optimization problem given by Equation 1 may be overly conservative; that is, the perceived impact of uncertainty is dependent on how closely the uncertainty set $\mathcal{M}$ describes the consumer population. A constrained optimization problem formulation is presented that maximizes revenue using a nominal choice model while constraining worst-case revenue to a predefined amount, as in Equation 3.

$$\begin{aligned} &\max_{S \subseteq \{1,\ldots,N\}:|S|=P} R(S;m) \\ &subject\ to\colon\ R(S;\mathcal{M}) \geq \underline{R} \end{aligned} \tag{3}$$

This formulation requires accommodating a constraint violation in the fitness function (making the optimization more challenging) and an "educated" approximation of the threshold for worst case revenue, $\underline{R}$. While a weighted-sum objective is also discussed that trades the performance of nominal and worst-case solutions, weighted-sum formulations have noted limitations [21].

Rather than pursue a weighted sum strategy, this paper introduces a multiobjective problem formulation that provides computational savings (in that the Pareto efficient frontier is found in a single optimization run) while allowing the tradeoff between nominal and worst-case revenue to be explored. Additionally, the problem formulations listed in Equations 1-3 model the impact of parameter (and/or structural) choice model uncertainty for the entire line. Changes in revenue represent consumers moving from a product offered by the firm to one that is offered by a competitor (or vice versa).

These works do not consider the ramifications of a choice model that reflects attributes of a product that will be physically manufactured, distributed, and sold. While revenue of the product line is still a driving business objective, the distribution of sales within the product line will dictate the allocation of resources to inventory and manufacturing. It would be expected that uncertainty in the choice model would cause variation in choice amongst the products within the line. A firm looking for a robust product design strategy would also want to minimize the variation in individual product share. Therefore, as part of this work a third objective is introduced that minimizes *the variation in choice amongst the products within the product line*.

Exploring solution performance variation when using samples of the posterior distribution

Previous work presented at the Sawtooth Software conference discussed the advantages of using a multiobjective optimization formulation for product line design problems. Often, however, the simulations driving the optimization use the mean of the posterior distribution from a hierarchical Bayes mixed logit model. This raises a concern when thinking about uncertainty in product line design problems – while the mean of the posterior distribution provides a Pareto efficient frontier, as shown in Figure 1, how large is the "scatter" around each Pareto point when plotting a subset of the draws used to arrive at the posterior mean?

This exploration began by using a multiobjective genetic algorithm (MOGA) to solve a product line design problem with two objectives. Part-worth estimates for 205 respondents were found using Sawtooth Software's CBC/HB module [22]. 800 draws of the lower-level posterior distribution were saved (e.g. 800 draws per respondent) and then used in a market simulator. The modeled objectives were maximizing the average of first choice share (in percent) and the average of profit per respondent obtained by the line (in dollars). It was confirmed that the average of the part-worths across the 800 draws matched the reported mean of the posterior distribution. A first choice rule was used, and the design problem consisted of 5 products, each with 7 configuration variables. The price for each product was set a continuous variables bounded between a lower and upper bound, resulting in a mixed-integer problem formulation of 2 objectives and 40 total design variables.

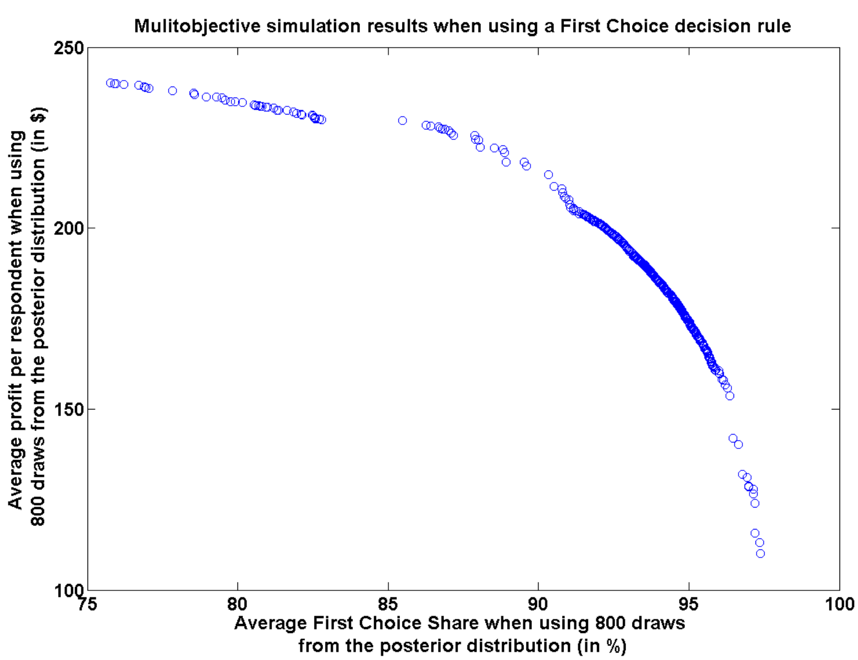

Figure 1. Pareto frontier obtained when using the average of 800 draws per respondent of a HB-ML model. A first choice rule was used to model respondent choice.

The genetic search converged within 300 generations, and 422 non-dominated solutions were identified. Product configurations and prices were recorded for each solution. From these 422 solutions there were 78 unique product line configuration combinations. The remaining solutions were non-unique in that they were priced differently from another product line with a similar content configuration. Four of these solutions were then chosen for further analysis. Two of the solutions were chosen near the extremities of the identified Pareto frontier. The configuration and prices associated with these solutions are shown in Tables 1 and 2. The other two were selected near the "knee" of the Pareto frontier.

Multiple product configurations are needed because customer preferences are heterogeneous and competition exists from the outside good and competitor products that were included in the market simulator. When maximizing a share objective, as shown in Table 1, an optimization algorithm will often drive product prices to their lower bound (for this problem, $52). Because a first choice rule is used, the optimal price for all products does not need to be at this value. Rather, they need to be at a price that does not trigger the change in binary outcome (chosen / not-chosen).

Table 1. Product configuration and pricing when maximizing the objective of average First Choice Share.

Product	Att1	Att2	Att3	Att4	Att5	Att6	Att7	Price	Avg. First Choice Share captured by each product
P1	8	8	3	6	8	6	3	$52	28.75%
P2	8	5	3	4	4	3	4	$52	36.55%
P3	8	8	3	4	6	8	4	$180.50	19.66%
P4	8	5	3	4	6	3	3	$209.03	7.33%
P5	5	8	3	6	8	1	3	$499.07	5.07%

Maximizing the average profit per respondent requires increasing the average price of the product line. As shown in Table 2, the low-end products found in Table 1 have been replaced with products priced around $200. These products will capture a majority of the share within the line, but the solution trades a reduction in market share for increased profit.

Table 2. Product configuration and pricing when maximizing the objective of average profit per respondent.

Product	Att1	Att2	Att3	Att4	Att5	Att6	Att7	Price	Avg. First Choice Share captured by each product
P1	8	5	3	4	2	3	4	$202.03	22.38%
P2	8	5	3	4	2	3	3	$203.13	21.29%
P3	8	8	3	4	5	8	4	$434.10	13.59%
P4	8	8	3	6	8	6	3	$458.87	8.73%
P5	5	8	3	6	8	1	4	$512.88	9.78%

For the four solutions identified, the performance (average First Choice Share, average profit per respondent) of all 800 draws is shown in Figure 2. A 95% confidence interval ellipse is also shown for each solution. Immediate observations from this figure include that the confidence ellipse for the maximum share solution (red) has a smaller major axis than the maximum profit solution (green). The smaller major axis for the maximum share solution is likely due to the lower price of the first two products. These products do not make money, rather they provide a buffer against part-worth variations – they capture a large amount of share and the small variations associated with the draws do not change this outcome. Conversely, the minor axis for the maximum share solution is larger than the maximum profit solution.

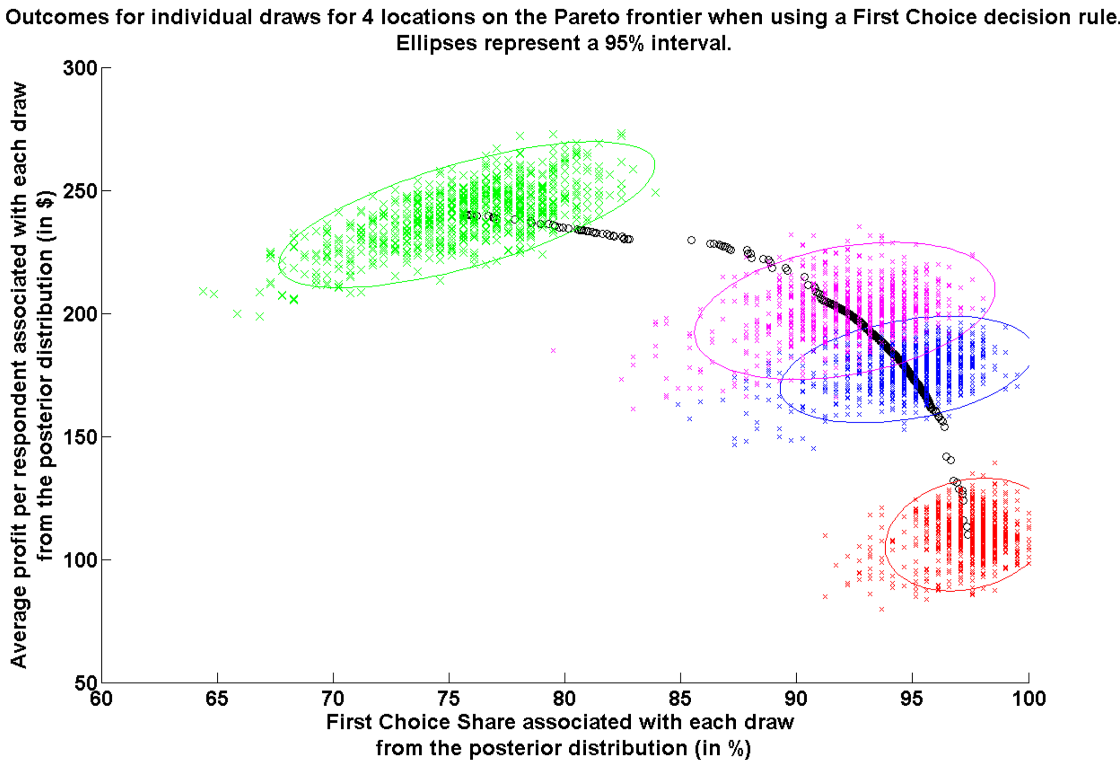

Figure 2. A plot demonstrating the scatter of performance values associated with the 800 draws per respondent from four product line solutions.

The solutions near the knee of the Pareto frontier were chosen because they highlight another challenge presented by uncertainty analysis. Dominance between two designs is no longer determined using a single set of performance values $\{F_1, F_2, \ldots, F_n\}$. Rather, an overlap of confidence intervals opens the possibility for strict dominance to not be maintained. While there has been work on multiobjective optimization algorithms capable of handling uncertainty [23], further studies are needed so that the ramifications for market-driven product design can be better understood.

Creating a multiobjective problem formulation for robust product line design

Parameter uncertainty was shown to have an effect when considering problem formulations driven by two different business objectives; variations of maximizing share and maximizing profit. As previously discussed, Bertsimas and Misic proposed a problem formulation for robust problem line optimization that maximizes the worst case expected revenue given an uncertainty set (Equations 1 and 2). They also discuss how this problem statement could be reformulated to maximize revenue around a nominal choice model subject to maintaining a revenue that is no lower than some predefined amount (Equation 3). Yet, it is challenging to define this amount a priori, and constraints increase the difficulty of creating an effective fitness function.

Such challenges can be overcome by reformulating the constraint-based problem described in Equation 3 as a multiobjective optimization problem. This formulation is constructed around a nominal model (m). For the purpose of this exploration, the mean of the posterior distribution was used as the nominal model because it corresponds to the part-worth values used in previous market simulator implementations. The uncertainty set ($\mathcal{M}$) consisted of the mean of the posterior distribution and the 800 draws (a subset of the total draws) saved from estimating the posterior distribution. The lower bound on product price was also redefined. Price was set at 125% of product cost, plus a constant value that was consistent across all products offered by the manufacturer. The objectives for the optimization were defined as maximizing the revenue per respondent (in dollars) and maximizing the worst case revenue across the uncertainty set, per respondent (in dollars). It should be noted that this is different than the formulation proposed by Bertsimas and Misic who use worst case expected revenue. This formulation for objective F_2 is heavily weighted toward the worst case scenario, and the full formulation is given by Equation 4.

Nominal model = Mean of the posterior distribution

Uncertainty set = 800 draws (per respondent) from the, and the mean of the,
posterior distribution

Product price = 1.25*Product cost + $52 (4)
Number of products = 5 (with 7 configuration variables each)

Use a multiobjective genetic algorithm (MOGA) to solve:
Maximize: F_1 = Revenue per respondent using the nominal model (in $)
Maximize: F_2 = Worst case revenue from uncertainty set, per respondent (in $)

The problem statement given by Equation 4 was optimized using a multiobjective genetic algorithm. Because product price was now a function of configuration cost, the number of unique solutions decreased. As shown in Figure 3, 8 unique product line configurations were identified as Pareto optimal points. Solutions in the upper right corner of the graph are preferred, as they maximize both revenue in the nominal model and the worst case revenue from the uncertainty set. Numerical results for these 8 solutions are presented in Table 3. In this table the maximum revenue per respondent is presented when using the mean of the posterior distribution (the nominal model). The worst case revenue, mean revenue, and the largest revenue, recorded from the 800 draws of the posterior distribution are also presented for each solution.

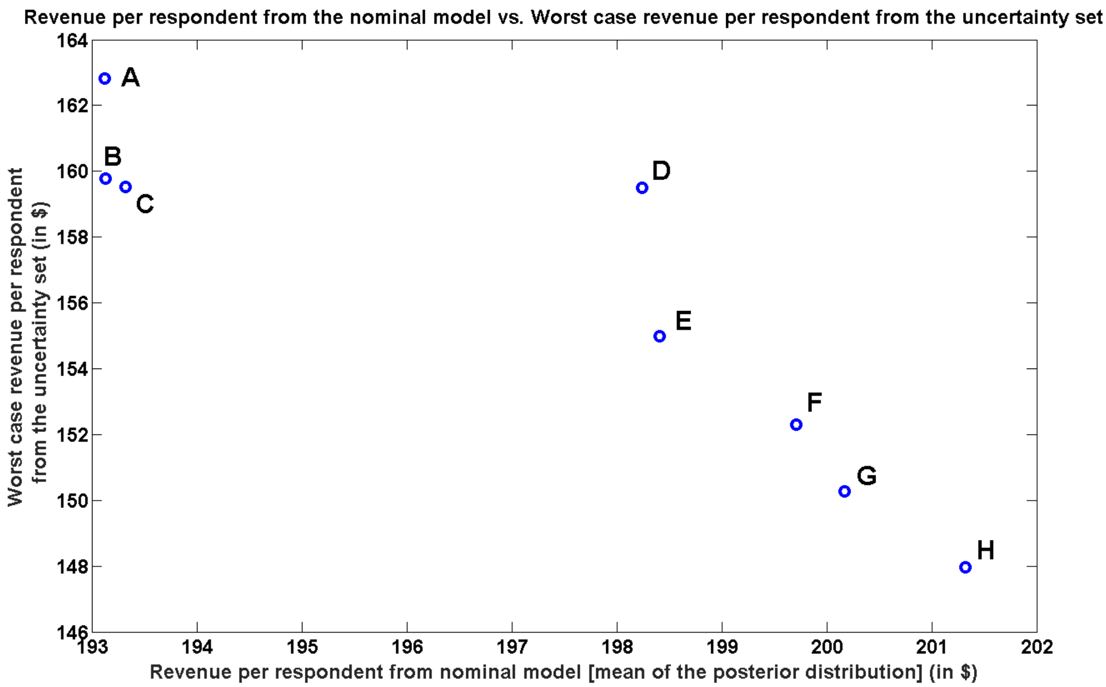

Figure 3. Pareto frontier when maximizing revenue per respondent under the nominal model and maximizing worst case revenue from the uncertainty set.

Table 3. Revenue values for the eight Pareto frontier solutions determined using the nominal model and the uncertainty set.

Solution	Maximum revenue per respondent calculated from mean of posterior distribution	Revenue per respondent calculated from samples of the posterior distribution		
		Minimum (worst case)	Mean	Maximum
A	$193.13	$162.81	$183.30	$208.17
B	$193.14	$159.78	$185.49	$211.19
C	$193.32	$159.52	$185.08	$211.30
D	$198.24	$159.49	$186.50	$211.44
E	$198.41	$154.99	$179.41	$203.64
F	$199.71	$152.29	$180.59	$207.41
G	$200.17	$150.27	$181.98	$208.79
H	$201.32	$147.96	$180.75	$211.15

The samples from the posterior distribution can also be used to create a probability density solution. Two of these distributions are shown in Figure 4. By moving from left to right in Figure 3, the worst case revenue decreases. Figure 4 illustrates that in the presence of parameter uncertainty a solution designed around the criterion of maximizing worst case revenue has opportunities to outperform a solution purely designed for maximum revenue. This creates a scenario where a decision maker must define the level of risk they are willing to adopt. Existing engineering design tools for concept selection provide insight into how such decisions can be made using utility theory and hypothetical alternatives [24].

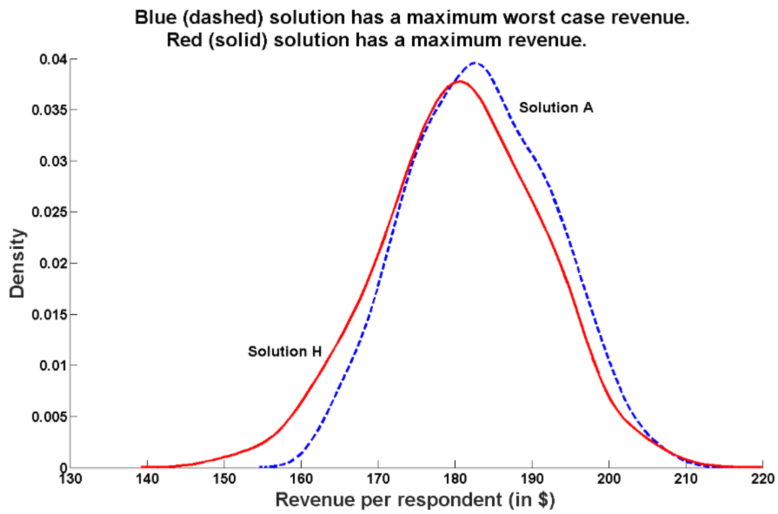

Figure 4. Probability density plot for solutions that maximize revenue (solid) and maximize worst case revenue (dashed).

At the suggestion of Bryan Orme, the revenue over the 800 draws were examined. The concern was that an outlier would make a worst-case revenue objective too aggressive. While all revenue values were found to be within 3.5 standard deviations of the mean, the lack of outliers does not eliminate the significance of this concern. A more effective strategy for this objective may involve defining a worst-case revenue percentile the decision maker is willing to accept. For the eight solutions found in this study, changing from worst case revenue to revenue at the 1^{st} or 5^{th} percentiles can cause solutions to become dominated. Here, Solution B would dominate Solution A, removing A as a Pareto point and preventing it from ever being chosen. The worst case revenue, and the revenue at the 1^{st} and 5^{th} percentiles, for each solution are shown in Table 4.

Table 4. Revenue values for the eight Pareto solutions when considering worst case revenue, 1^{st} percentile revenue, and 5^{th} percentile revenue.

Solution	Worst case revenue	1^{st} percentile revenue	5^{th} percentile revenue
A	$162.81	$163.96	$168.32
B	$159.78	$165.25	$170.04
C	$159.52	$165.26	$169.83
D	$159.49	$164.60	$171.10
E	$154.99	$155.84	$164.65
F	$152.29	$159.27	$165.73
G	$150.27	$159.00	$164.80
H	$147.96	$154.43	$163.06

Concerns about the effect of parameter uncertainty on business objectives aligned with revenue motivated further analysis. If the uncertainty set could be used as a means of determining worst-case revenue, then the uncertainty set could also be used to explore variations of first choice share ***within a product line***. Uncertainty analysis conducted this way provides a product line perspective that has not been discussed in the literature, and is discussed in the next section.

Exploring the effect of parameter uncertainty from a product share perspective

Additional motivation for exploring the effect of parameter uncertainty from a product share perspective is shown in Figure 5. Consider a scenario where the manufacturer has decided to offer three products. These products compete in a market against three competitor products and an outside good. Now, consider a single respondent making a choice in this market. Their selection is modeled using a first choice decision rule. For a given draw from the posterior distribution (let us call it Draw A), results from the market simulator indicate that the respondent selects the first of the three products offered by the manufacturer. Since a first choice rule is being used, choice is fully assigned to a single product.

The uncertainty set discussed in the previous section was comprised of a set of draws from the posterior distribution. If another draw is considered (we will call this one Draw B), the results from the market simulator indicate that the ***same respondent*** has now chosen the second of the firm's three offerings. From the share perspective of a product line, nothing has changed; the effect of uncertainty would be unobservable. Yet, from a product manufacturing and component inventory perspective, the change in respondent choice is significant. While the first choice share for this respondent at the product line level remains consistent at 100%, the deviation of share within the product line is also 100% (going from the first offering to the second offering).

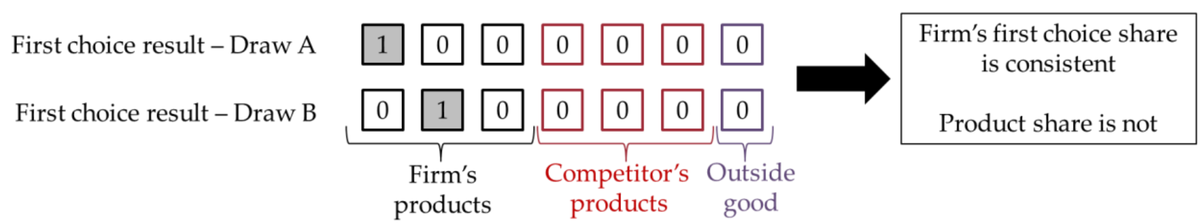

Figure 5. Representative example demonstrating how parameter uncertainty can be unobservable at the product line level, while having significant impact at the product share level.

This led to a question that motivated the second half of this work – how big of an issue is parameter uncertainty when making resource and production allocation decisions? As an initial exploration, the variability in product share was examined for Solution A from Figure 3. Observations for First Choice Share within the product line were taken over the 800 draws from the posterior distribution. A summary of these observations are reported in Table 5. Reported values include the First Choice Share of each product in the line from the nominal model (the mean of the posterior distribution) and the mean, standard deviation, minimum, and maximum values of first choice share distribution from the 800 draws.

The significance of parameter uncertainty when making configuration and pricing decisions is demonstrated by the misalignment of within line share distribution between the nominal model and the uncertainty set. First, there is a difference in the mean values of First Choice Share between the nominal model and the uncertainty set. Perhaps more important is the range between minimum and maximum percent distribution of First Choice Share. The density plot for Product 1's First Choice Share is shown for the uncertainty set in Figure 6. In the next section, a problem formulation strategy designed to reduce the width of this distribution is introduced.

Table 5. Comparison of First Choice Share for products within a product line solution. Results for both the nominal model and the uncertainty set are reported.

Product within the line	Percent distribution of within line First Choice Share (in %)				
	Model: Mean of the posterior distribution	Model: Uncertainty set			
		μ	σ	Min	Max
Product 1	13.01	16.90	4.54	5.34	31.03
Product 2	9.59	12.25	3.45	2.92	21.43
Product 3	12.33	15.04	4.23	5.17	27.52
Product 4	43.15	34.77	4.97	20.97	45.65
Product 5	21.92	21.04	4.59	9.72	32.79

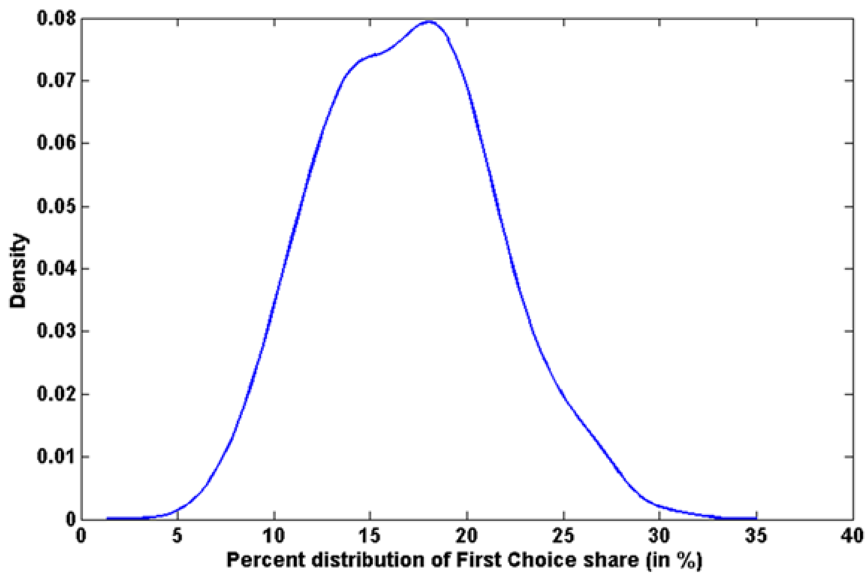

Figure 6. Density plot of First Choice Share distribution for Product 1 using the uncertainty set.

Building variation in product share from model uncertainty into the problem formulation

We build on the results presented in the previous section by further exploring how Respondent #1's product selection changes over the 800 draws from the posterior distribution. A bar chart showing choice rule outcomes is shown in Figure 7. Using the nominal model of the mean of the posterior distribution, the choice rule results in a selection of Product 1 from the firm. For approximately 300 of the 800 draws, this choice rule result is also obtained. Less than

100 of the draws divert share from the firm to the competitor products or the outside good. Rather, over half of the draws from the posterior distribution maintain firm-level share while diverting product share, with most of the share going to Products 4 and 5.

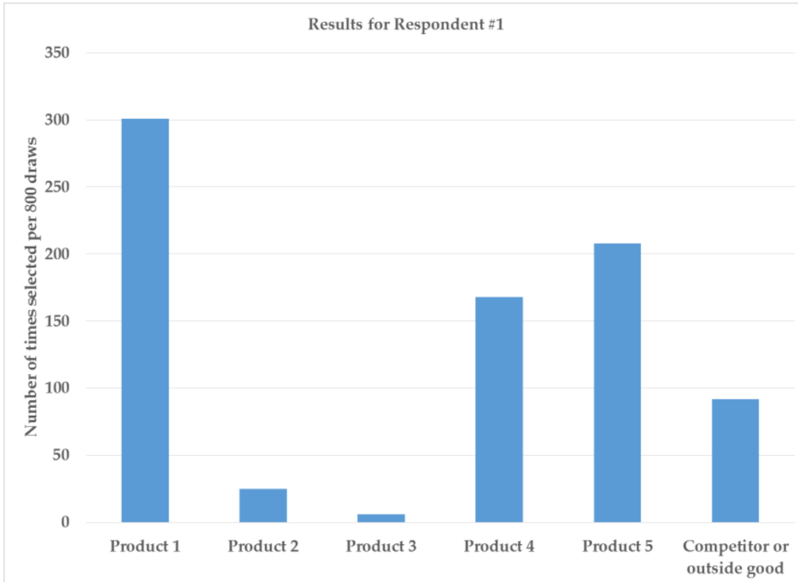

Figure 7. First choice rule results for Respondent #1 using the uncertainty set. Under the nominal model this respondent's first choice was Product 1.

These results support the argument that a robust product line solution should be one that captures maximum market share with minimal variability, while ***also*** minimizing the variation in product share. A first thought was to develop a metric that quantified choice consistency so that it could be incorporated into a multiobjective problem formulation. This metric, as shown in Equation 5, calculates the average number of choice inconsistencies per respondent. Here, N is the number of respondents who chose one of the firm's products using the first choice rule and the mean of the posterior distribution. R is the number of draws (for this problem 800).

As shown in Figure 8, the mean of the posterior distribution is used as the "truth" for each respondent. The results of the first choice rule using these point-estimates are then compared against the market simulator results for all 800 draws. For cases where the choice rule outcome between the nominal model and one of the draws of the uncertainty set align, the indicator function is 0. When the outcome of the choice rule between the nominal and uncertainty model are different, the indicator function is 1.

$$ACI \equiv P\left[\boldsymbol{c} \notin \Omega_{PL,D}\right] = \frac{1}{N}\sum_{n=1}^{N}\sum_{r=1}^{R} I_{\mathrm{CF}}^{n,r}\left(\boldsymbol{c}_{\boldsymbol{n},\boldsymbol{r}}\right) \tag{5}$$

The introduction of Equation 5 allows for the formulation of a three-objective optimization problem. This problem, shown in Equation 6, builds on the previous formulation with the additional goal of minimizing the average number of choice inconsistencies per respondent. 318 unique solutions were found using this optimization problem. As might be expected, the design solutions found to be most robust from choice inconsistencies performed poorly on revenue objectives. A two-dimensional scatterplot is shown in Figure 9 that illustrates the tradeoff between revenue per respondent and average choice inconsistencies per respondent. Solutions that reduce choice inconsistencies lead to product line solutions that generate minimal revenue.

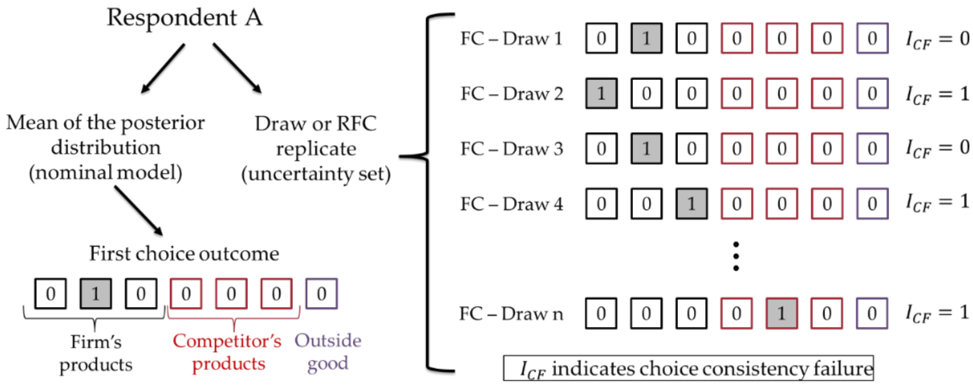

Figure 8. Representation of choice consistency failure when considering multiple draws. The first choice decision under the nominal model is used as the reference.

Nominal model = Mean of the posterior distribution

Uncertainty set = 800 draws (per respondent) from the, and the mean of the, posterior distribution

Product price = 1.25*Product cost + $52
Number of products = 5 (with 7 configuration variables each)

Use a multiobjective genetic algorithm (MOGA) to solve:
Maximize: F_1 = Revenue per respondent (in $)
Maximize: F_2 = Worst case revenue from uncertainty set, per respondent (in $)
Minimize: F_3 = Variation in choice when considering the uncertainty set

(6)

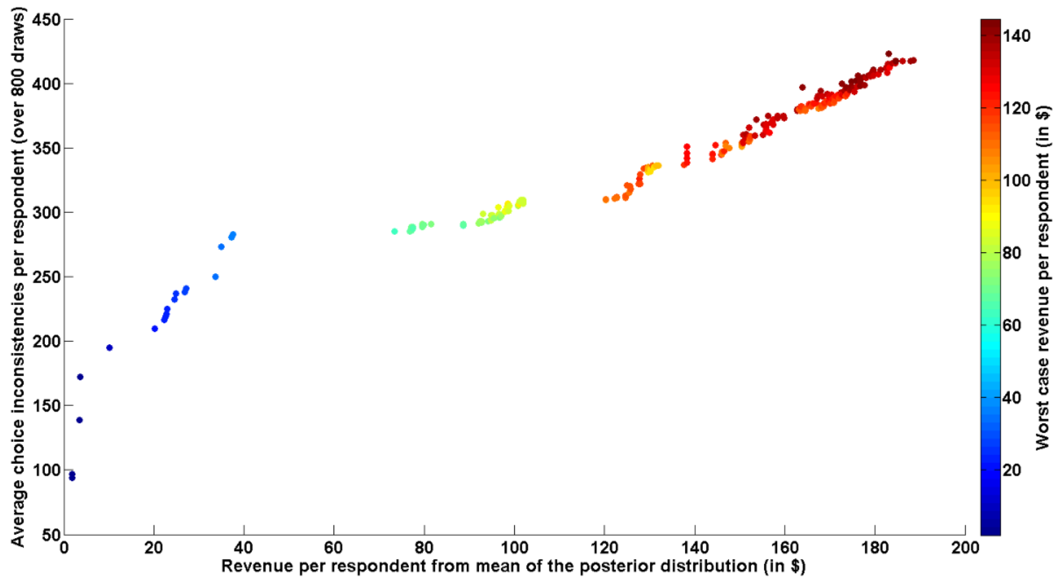

Figure 9. Scatterplot of the optimal designs generated for the three-objective problem formulation. As average choice inconsistencies decrease, so too does the revenue generated by the line.

Reformulating the optimization problem for market-level behavior

The solutions shown in Figure 9 are based on the concept that choice inconsistencies should be minimized for each respondent. Yet, this formulation may not reflect the actual goal. The thought of reformulating the third objective was inspired by Chris Chapman's paper at the 2013 Sawtooth Software conference [25]. In this paper, Chris discussed how the results of a market simulator were not intended to focus on the behavior of an individual respondent, but the overall response of the market as a whole.

This led to a realization; choice inconsistencies at the respondent level could cancel each other out, but this was not accounted for in Equation 5. Rather, this outcome was being penalized twice. The choice inconsistency formulation shown in Equation 5 was then replaced with a variation in First Choice Share (FCS) calculation shown in Equation 7. Here, n represents the number of products being developed by the manufacturer. The first choice share is calculated for each product using the nominal model. The difference between the FCS from the nominal model and the mean FCS obtained from the uncertainty set is then determined. This result is squared and multiplied by a weighting factor w_i. Weighting factors are bounded between 0 and 1, and the sum of the weighting factors must equal 1. While the weighting factors for most problems may be equal, the weight in FCS deviation can be increased for a particular product when it has configuration parameters specific to it. For example, a single product in the line may use a unique engine type, or a particular material, that could not be used on other products in the line if manufacturing numbers are adjusted. Equation 7 is then used in the reformulated three-objective optimization problem shown in Equation 8.

$$\sqrt{\sum_{i=1}^{n} w_i\left(FCS(i)_{nominal} - \mu(FCS(i))_{uncertainty\,set}\right)^2}$$

$$\sum w_i = 1 \tag{7}$$

$$0 \leq w_i \leq 1$$

Nominal model = Mean of the posterior distribution

Uncertainty set = 800 draws (per respondent) from the, and the mean of the, posterior distribution

Product price = 1.25*Product cost + $52
Number of products = 5 (with 7 configuration variables each) (8)

Use a multiobjective genetic algorithm (MOGA) to solve:
Maximize: F_1 = Revenue per respondent (in $)
Maximize: F_2 = Worst case revenue from uncertainty set, per respondent (in $)
Minimize: F_3 = Variation in First Choice Share distribution (in %)

Solving this optimization problem allows for the simultaneous consideration of business and manufacturing tradeoffs in the presence of parameter uncertainty. As shown in Figure 10, a decision maker can identify that the product line solution with the maximum worst case revenue also has one of the highest variations in first choice share distribution. Multi-attribute decision making tools can be used when selecting a final solution from this set of non-dominated product lines. A solution has been identified in Figure 10 that reduces the variation in first choice share distributions to under 4% for the product line that lies near the efficient frontier for the tradeoff between average revenue per respondent (from the nominal model) and the average worst case revenue per respondent (from the uncertainty set).

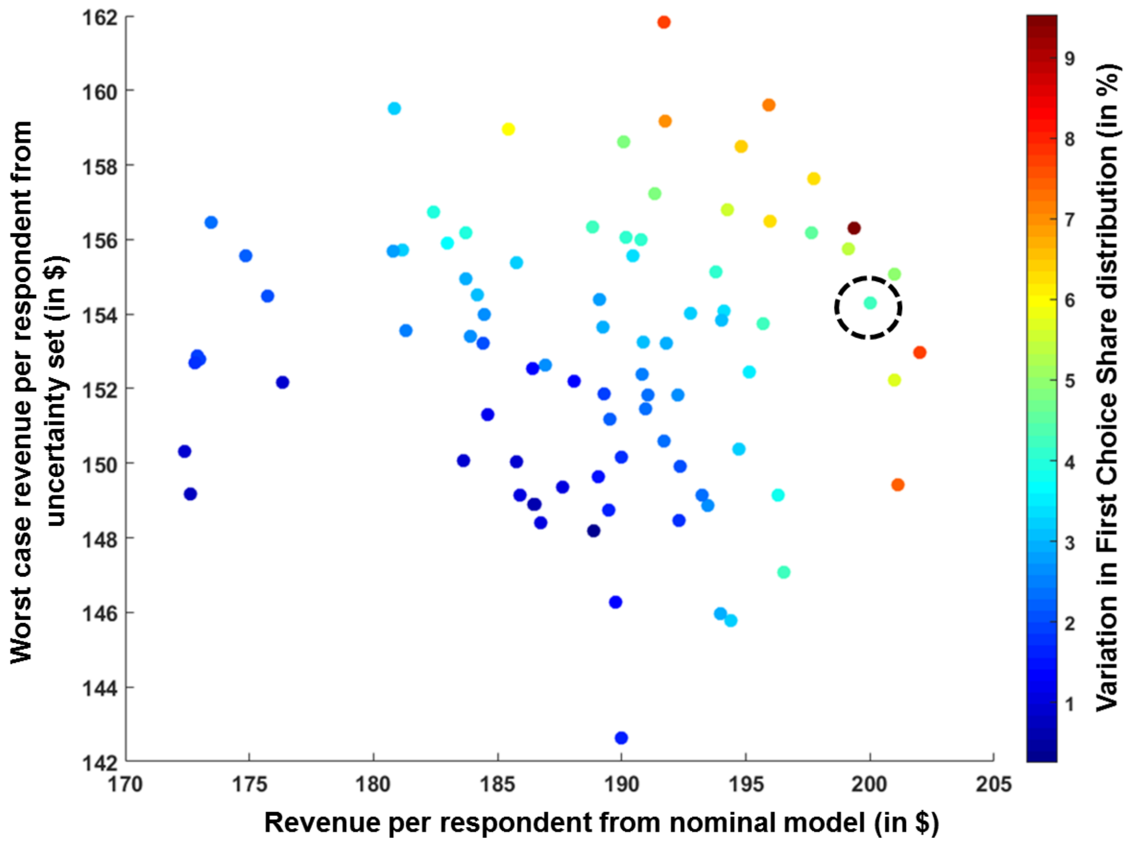

Figure 10. Scatterplot of the reformulated three-objective problem. Here, the colorbar represents the variation in First Choice Share distribution of products within the line.

The configurations for this product line are shown in Table 6. Some commonality is observed in this solution – all products use configuration level 3 for the third attribute, only levels 5 and 8 are used for the second attribute, etc. Commonality reduces the concern associated with variation in first choice share caused by parameter uncertainty. However, for the sixth product attribute, four different configuration levels compose the product line solution. As part of future work, the weighting term in Equation 7 could be scaled as a function of the number of unique components used in a particular product. This would enforce greater consistency for those components that must be purchased for only a single product. Developing a greater understanding of how variation in first choice share distribution impacts inventory and supply chain decisions could reduce negative business outcomes caused by parameter uncertainty.

Table 6. Product line configuration profile for the solution identified in Figure 10.

Product	Att1	Att2	Att3	Att4	Att5	Att6	Att7
P1	4	8	3	6	7	1	4
P2	5	5	3	4	2	3	4
P3	8	5	3	6	5	4	3
P4	8	8	3	3	2	3	4
P5	8	8	3	4	5	8	3

Conclusions and future work

Market simulators created from estimates of customer preference are powerful tools for exploring market response to new product offerings. When heterogeneity is represented using a hierarchical Bayes mixed logit model, the most basic market simulators will use part-worth values for each respondent associated with the mean of the lower-level posterior distribution. This choice is made because it reduces computational complexity, allowing for faster simulations and reduced cost when optimizing a product line. However, a failure to account for uncertainty can undermine these computational advantages and result in product configuration and pricing decisions that forfeit value.

The simulations presented in this paper demonstrate the importance of accounting for uncertainty when conducting market simulations. Parameter uncertainty was addressed by considering 800 draws saved from the lower-level posterior distribution of a hierarchical Bayes mixed logit model. For a multiobjective optimization problem of share versus profit, four locations of the Pareto frontier were explored. As more weight is placed on the profit objective, the major axis of the confidence ellipse grew (meaning more scatter in the predicted share of the product line over the 800 draws). Conversely, the minor axis of the confidence ellipse shrunk, leading to less scatter on the profit objective. Comments have been made at previous Sawtooth Software conferences about the possibility of an overstated variance for uncertainty when simulating from draws obtained from the lower-level posterior distribution. Simulating from the upper-level model may provide a more accurate uncertainty representation, though lower-level models have been shown to reflect respondent heterogeneity for product line problems. Exploring how information from both the upper- and lower-level models can be used is an opportunity for future work.

Attention then turned to an optimization problem that maximized the revenue generated by a product line. Here, the mean of the posterior distribution was used as the basis for a nominal model. This mean value from the lower-level HB model and the 800 draws from the posterior distribution were combined to create an uncertainty set of models. A multiobjective optimization problem was formulated that maximized revenue under the nominal model while simultaneously maximizing worst case revenue from the uncertainty set. By looking at worst case revenue from an entire set of models, an effective "lower bound" for revenue could be determined for each product line solution.

There are interesting challenges raised by this problem formulation. The problem formulation given by Equation 4 considers a worst case scenario. Outliers may drive the optimization result, and the decision to define a percentile threshold may be more effective. As the paper's discussant at the conference, Mark Beltramo noted that maximizing worst-case revenue amounts to maximizing revenue from a small quantile of the distribution, corresponding to a decision maker that is extremely risk averse. Because optimizing the nominal model provided a biased estimate for maximum revenue compared to the average revenue over the uncertainty set, he continued by proposing a problem formulation where the first objective maximized average revenue per respondent over the uncertainty set (expected return), while the second objective maximized average revenue per respondent minus the k^{th} quantile over the draws (risk). Since robust product line design is similar in concept to balancing risk and return in a stock portfolio [26], he suggested that the 5^{th} percentile may be considered for k to align with common practice in finance. Finally, the composition of the uncertainty set could be further explored. All draws that were saved from the posterior distribution are considered to have equal value, though it may be true that some are closer to resembling true market behavior.

A significant contribution of this paper is the introduction of a third objective that aligns with the resource and production allocation decisions discussed by Bertsimas and Misic. The introduction of a third objective began with a measure of choice inconsistency within a product line that was measured at the respondent level. However, individual respondent choice may not be the appropriate concern. This third objective was reformulated to minimize the variation in First Choice Share for each product in the line. The two measures used were the within-line First Choice Share from the nominal model and the average within-line First Choice Share from the uncertainty set. By adding a third objective, solutions can be found that balance revenue uncertainty at the product line level while providing insight into the extent that the choice of individual products varies. This information is significant because it can be used to identify how parameter uncertainty within the market simulator might impact component ordering and production decisions that need to be made by a manufacturer. The discussant noted that expected holding costs increase with variance of demand, and a formulation of the third objective that minimized a weighted sum of the variances of the individual product choice shares could be used. Such considerations have been unexplored using market simulators, and this formulation presents significant opportunities for the use of quantitative market research models when designing products that are physically produced.

While some opportunities for future work have already been discussed, it is important to note that the results presented in this paper only consider model parameter uncertainty. Structural uncertainty has been discussed in the literature, though it was not addressed in this paper. Further, the incorporation of uncertainty into market simulators can include uncertainty related to the product attributes used by the firm, component costs, and the attributes and prices of competitor products. Incorporating these additional uncertainties will increase the computational expense of a simulation. Yet, the potential hazard of ignoring uncertainty in market simulators can lead to configuration and pricing decisions that forfeit value and can result in resource allocation decisions that cannot be easily reversed or corrected.

Acknowledgements

I would like to thank Bryan Orme for his insightful questions and comments during presentation preparation. I would also like to thank Mark Beltramo for the discussion he provided at the Sawtooth Software conference and for his insights about future work in this area.

References

[1] Chapman, C., and Alford, J., 2011, "Product portfolio evaluation using choice modeling and genetic algorithms," Proc. 2010 Sawtooth Softw. Conf.

[2] Turner, C., Foster, G., Ferguson, S., Donndelinger, J., and Beltramo, M., 2012, "Creating targeted initial populations for genetic product searches," 2012 Sawtooth Softw. Conf.

[3] Foster, G., Ferguson, S., and Donndelinger, J., 2014, "Demonstrating the Need and Value of a Multiobjective Product Search," 2013 Sawtooth Software Conference, October 14-18, Dana Point, CA.

[4] Ferguson, S. M., 2015, "Climbing the Content Ladder: How Product Platforms and Commonality Metrics Lead to Intuitive Product Strategies," 2015 Sawtooth Software Conference, Orlando, FL.

[5] Tsafarakis, S., Marinakis, Y., and Matsatsinis, N., 2011, "Particle swarm optimization for optimal product line design," Int. J. Res. Mark., **28**(1), pp. 13–22.

[6] Michalek, J. J., Ceryan, O., Papalambros, P. Y., and Koren, Y., 2006, "Balancing

Marketing and Manufacturing Objectives in Product Line Design," J. Mech. Des., **128**(6), p. 1196.

[7] Wang, X. F., Camm, J. D., and Curry, D. J., 2009, "A branch-and-price approach to the share-of-choice product line design problem," Manage. Sci., **55**(10), pp. 1718–1728.

[8] Belloni, a., Freund, R., Selove, M., and Simester, D., 2008, "Optimizing Product Line Designs: Efficient Methods and Comparisons," Manage. Sci., **54**(9), pp. 1544–1552.

[9] Bertsimas, D., and Mišić, V. V, 2017, "Robust Product Line Design," Oper. Res., **65**(1).

[10] Gilbride, T. J., and Lenk, P. J., 2010, "Posterior Predictive Model Checking: An Application to Multivariate Normal Heterogeneity," J. Mark. Res., **47**(5), pp. 896–909.

[11] Abramson, C., Andrews, R. L., Currim, I. S., and Jones, M., 2000, "Parameter Bias from Unobserved Effects in the Multinomial Logit Model of Consumer Choice," J. Mark. Res., **37**(4), pp. 410–426.

[12] Montgomery, A. L., and Bradlow, E. T., 1999, "Why Analyst Overconfidence About the Functional Form of Demand Models Can Lead to Overpricing," Mark. Sci., **18**(4), pp. 569–583.

[13] Orme, B. K., and Huber, J., 2000, "Improving the value of conjoint simulations," Mark. Res., **12**(4), pp. 12–20.

[14] Camm, J. D., Cochran, J. J., Curry, D. J., and Kannan, S., 2006, "Conjoint Optimization: An Exact Branch-and-Bound Algorithm for the Share-of-Choice Problem," Manage. Sci., **52**(3), pp. 435–447.

[15] Kleywegt, A. J., Shapiro, A., and Homem-de-Mello, T., 2002, "The Sample Average Approximation Method for Stochastic Discrete Optimization," SIAM J. Optim., **12**(2), pp. 479–502.

[16] Wang, X. (Jocelyn), and Curry, D. J., 2012, "A robust approach to the share-of-choice product design problem," Omega, **40**(6), pp. 818–826.

[17] Luo, L., Kannan, P. K., Besharati, B., and Azarm, S., 2005, "Design of Robust New Products under Variability: Marketing Meets Design*," J. Prod. Innov. Manag., **22**(2), pp. 177–192.

[18] Besharati, B., Luo, L., Azarm, S., and Kannan, P. K., 2006, "Multi-Objective Single Product Robust Optimization: An Integrated Design and Marketing Approach," J. Mech. Des., **128**(4), p. 884.

[19] Resende, C. B., Grace Heckmann, C., and Michalek, J. J., 2012, "Robust Design for Profit Maximization With Aversion to Downside Risk From Parametric Uncertainty in Consumer Choice Models," J. Mech. Des., **134**(10), p. 100901.

[20] Ver Hoef, J. M., 2012, "Who Invented the Delta Method?," Am. Stat., **66**(2), pp. 124–127.

[21] Das, I., and Dennis, J. E., 1997, "A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems," Struct. Optim., **14**(1), pp. 63–69.

[22] Sawtooth Software inc., 2014, "Sawtooth Software CBC/HB 5.5.3."

[23] Chapman, J. L., Lu, L., and Anderson-Cook, C. M., 2014, "Incorporating response variability and estimation uncertainty into Pareto front optimization," Comput. Ind. Eng., **76**, pp. 253–267.

[24] See, T.-K., Gurnani, A., and Lewis, K., 2004, "Multi-Attribute Decision Making Using Hypothetical Equivalents and Inequivalents," J. Mech. Des., **126**(6), p. 950.

[25] Chapman, C., 2013, "9 Things Clients Get Wrong about Conjoint Analysis," Proceedings of the 2013 Sawtooth Software Confernece, B. Orme, ed., Dana Point, CA.

[26] Markowitz, H. M., 1952, “Portfolio selection,” J. Finance, **7**(60), pp. 77–91.