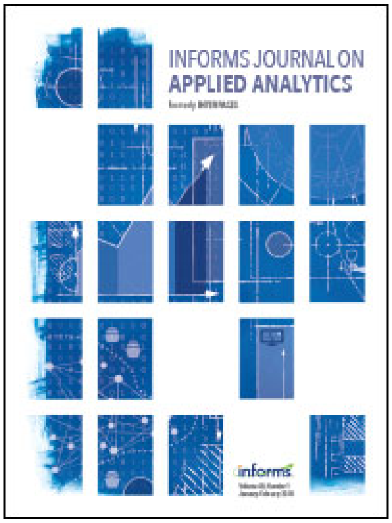


INFORMS Journal on Applied Analytics

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Uncertainty Quantification in Vehicle Content Optimization for General Motors

Eunhye Song, Peiling Wu-Smith, Barry L. Nelson

informs
http://pubsonline.informs.org/journal/inte
INFORMS JOURNAL ON APPLIED ANALYTICS
Vol. 50, No. 4, July–August 2020, pp. 225–238
ISSN 0092-2102 (print), ISSN 1526-551X (online)

Uncertainty Quantification in Vehicle Content Optimization for General Motors

Eunhye Song,[a] Peiling Wu-Smith,[b] Barry L. Nelson[c]

[a] The Pennsylvania State University, University Park, Pennsylvania 16801; [b] General Motors, Warren, Michigan 48092; [c] Northwestern University, Evanston, Illinois 60201

Contact: eus358@psu.edu, https://orcid.org/0000-0002-5171-0614 (ES); peiling.wu-smith@gm.com, https://orcid.org/0000-0002-5801-3395 (PW-S); nelsonb@northwestern.edu, https://orcid.org/0000-0002-1325-2624 (BLN)

Received: January 7, 2019
Revised: November 14, 2019; April 1, 2020
Accepted: April 9, 2020
Published Online in Articles in Advance: July 10, 2020

https://doi.org/10.1287/inte.2020.1041

Abstract. A vehicle content portfolio refers to a complete set of combinations of vehicle features offered while satisfying certain restrictions for the vehicle model. Vehicle Content Optimization (VCO) is a simulation-based decision support system at General Motors (GM) that helps to optimize a vehicle content portfolio to improve GM's business performance and customers' satisfaction. VCO has been applied to most major vehicle models at GM. VCO consists of several steps that demand intensive computing power, thus requiring trade-offs between the estimation error of the simulated performance measures and the computation time. Given VCO's substantial influence on GM's content decisions, questions were raised regarding the business risk caused by uncertainty in the simulation results. This paper shows how we successfully established an uncertainty quantification procedure for VCO that can be applied to any vehicle model at GM. With this capability, GM can not only quantify the overall uncertainty in its performance measure estimates but also identify the largest source of uncertainty and reduce it by allocating more targeted simulation effort. Moreover, we identified several opportunities to improve the efficiency of VCO by reducing its computational overhead, some of which were adopted in the development of the next generation of VCO.

History: This paper was refereed.
Funding: This work was funded by the General Motors Operations Research Department.

Keywords: vehicle market simulation • design of experiments • discrete choice model • uncertainty quantification • sensitivity analysis

Introduction

The three key players in the automotive sales market—automakers, car dealers, and customers—inherently focus on different decisions with different measures of success. Automakers decide what vehicles to produce and what combinations of features to offer and how to price them, which directly affect what car dealers can order to sell to customers and what customers can potentially purchase from dealerships. Automakers typically measure success in terms of market share, revenue, and profit. Car dealers, who sell new vehicles at the retail level based on a contract with an automaker, decide what vehicles to stock on their lots and at what price to sell to each customer. In the United States, franchise laws require new cars be sold only by dealerships. Although direct order from automakers is available to customers, the majority of vehicles are sold through dealer stock. Hence, how well dealers stock their lots plays an important role in customers' satisfaction so that they can find what they want. Dealers measure their success in terms of the profitability of their businesses, which includes both the profit on current sales and the prospect of future sales. Customers make decisions on which vehicle to purchase given what is available in the market and how much they are willing to spend. Customers' satisfaction with a vehicle purchase impacts dealer profitability. Both customer satisfaction and dealer profitability are critical to the long-run growth of profit and market share for automakers.

At General Motors (GM), the *content* of each vehicle refers to the features the vehicle has to offer, such as car seats, engines, sound systems, etc. The *content portfolio* of a vehicle model is a complete set of combinations of features offered by GM while satisfying certain restrictions for the vehicle model. Because dealers stock their lots by ordering vehicles with different content prescribed by the content portfolio, it is important to reflect customers' preferences and willingness to pay when deciding a content portfolio.

To optimize the content portfolios of vehicle models, GM developed Vehicle Content Optimization (VCO), which is now a standard part of GM's vehicle development process (Wu-Smith et al. 2014). VCO starts with a conjoint-based market research study to quantify customers' preferences for features and their willingness to pay for them. At the heart of VCO is a market simulator that simulates the dynamics of customer choice, such as what features customers want, how much they are willing to pay, how likely they are to find what they want on the dealers' lots at that price, and how they make purchase decisions. This market simulator enables GM's product teams to evaluate different content portfolios and make informed decisions based on estimated performance measures such as profit, sales volume, and market share.

VCO is a leap forward from the ad hoc decision-making process in the past; however, it requires intensive computing power owing to the stochastic and combinatorial nature of the problem. This has forced GM to trade off the accuracy and precision of the estimated performance measures with the computation time. Simultaneously, the simulation results reflect inherent natural variability in the real-world vehicle market, which is confounded with the estimation error as a result of the finite computation budget. Moreover, VCO consists of several steps, each requiring a substantial computation budget. However, these budgets were prespecified without explicitly quantifying how they affect the overall uncertainty in performance measure estimates. Because VCO is used in all major vehicle models' content optimization decisions, uncertainty in VCO poses potential business risk to GM. Currently, GM achieves, on average, more than a billion dollars in annual revenue from each vehicle model. According to their market report, GM sold 758,000 vehicles in the United States in the second quarter of 2018, and its average transaction price was $35,500, which amounts to 27 billion dollars in revenue (General Motors 2018). Given the scale of GM's sales, even a 1% error in the estimated sales volume translates to millions of dollars' difference in revenue.

The purpose of our study was to establish an uncertainty quantification procedure for VCO that works for any GM vehicle. In particular, we aimed to provide the VCO team actionable guidance on how to efficiently reduce uncertainty in the performance measure estimates by recommending targeted computational effort. To this end, GM would be able to either reduce the estimation error of its key performance measures to the level that is not harmful to its business decisions or make more conservative decisions to hedge business risks accounting for quantified uncertainty.

If a computer model is linear in inputs, then uncertainty in the model output can be directly explained by its derivative with respect to the inputs. However, when the model is nonlinear in inputs, as in VCO's case, the derivative at a particular set of inputs only delivers local sensitivity information. An alternative approach is *global sensitivity analysis* designed to analyze uncertainty in a highly nonlinear computer model. Global sensitivity analysis measures how the model's output varies when its inputs change according to their distributions and evaluates the effect of each input's contribution to the overall uncertainty in the model. Several global sensitivity indices have been proposed in the literature. Sobol′ (1993) applies functional analysis of variance to decompose variance in the model output and attribute it to each subset of inputs. Homma and Saltelli (1996) define the first-order and total effects of each input from Sobol′ indices to measure the contribution of an individual input alone and together with other inputs, respectively. These measures are extremely popular in a wide range of applications including nuclear safety assessment (Saltelli and Tarantola 2002), chemical experiment planning (Saltelli et al. 2005), land-use policy assessment (Ligmann-Zielinska et al. 2014), flood simulation (Pianosi et al. 2016), and fire spread analysis (Song et al. 2016).

As summarized by Iooss and Lemaître (2015), index-based global sensitivity analysis is most suitable for a computer model with a small to moderate number of inputs owing to its computational cost because it involves sampling multiple sets of inputs and running computer experiments for each set. Although VCO has a relatively small number of inputs, its nested structure is a challenge to uncertainty quantification. Namely, some input parameters of the VCO's market simulator are first estimated via calibrating the simulation outputs with historical data, which involves solving a highly nonlinear optimization problem. Thus, "sampling multiple sets of input parameters" requires running multiple calibrations, which is infeasible for GM because of high computational cost. Instead, we take a hybrid local-and-global sensitivity analysis approach; we analyze local sensitivity of the simulation output to calibration but compute global sensitivity indices with respect to other inputs. This significantly reduces the computational cost of full global sensitivity analysis and thereby enables GM to quantify uncertainty in VCO within an affordable time frame. Prior to introducing our approach, we present an example of content portfolio optimization at GM and highlight its challenges to emphasize the importance of VCO to GM.

Content Portfolio Optimization at GM

Figure 1 displays an example of a partial content portfolio for the model year 2018 (MY18) Chevrolet Cruze—namely, what trim levels to be offered, what vehicle features to be offered as standard or as optional on each trim level, and what optional features to be bundled as option packages. There are five trim levels for the MY18 Chevrolet Cruze: L, LS, LT, Diesel, and Premier, as indicated by the five column names. Various vehicle feature categories are shown by rows,

Figure 1. A Portion of the Trim Levels and Corresponding Features of MY18 Chevrolet Cruze

SPECIFICATIONS

● STANDARD ○ AVAILABLE – NOT AVAILABLE

	L	LS	LT	DIESEL	PREMIER
EXTERIOR					
Door handles: Body-color	●	●	●	●	–
Body-color with chrome strip	–	–	–	–	●
Foglamps	–	–	○[1]	○[1]	○[1]
Glass: Solar absorbing	●	●	●	●	●
Headlamps with automatic exterior lamp control and delay: Halogen composite	●	●	–	–	–
Halogen	–	–	●	●	●
IntelliBeam high beam assist	–	–	–	–	○[2]
Power-adjustable manual-folding body-color mirrors	●	●	–	–	–
Heated	–	–	●	●	–
Heated with turn signal indicators	–	–	–	–	●
ENGINE/CHASSIS					
Engine: 1.4L DOHC turbocharged 4-cylinder with Continuously Variable Valve Timing, 153 hp and 177 lb.-ft. of torque	●	●	●	–	●
1.6L Turbo-Diesel DOHC 4-cylinder with 153 hp and 177 lb.-ft. of torque	–	–	–	●	–
Steering: Power, electric-assisted, rack-mounted	●	●	●	●	●
Suspension: Front, independent MacPherson struts	●	●	●	●	●
Rear, compound crank	●	●	●	●	–
Rear, compound crank with Z-link	–	–	–	○	●
Transmission: 6-speed manual	●	●	○[1]	○	–
6-speed automatic; electronically controlled with overdrive and stop/start technology	–	○	○	–	●
9-speed automatic; electronically controlled with overdrive and stop/start technology	–	–	–	○	–
SAFETY & SECURITY					
10 air bags:[3] Driver and front passenger frontal, includes Passenger Sensing System; front and rear seat-mounted side-impact; roof rail-mounted head-curtain for front and rear outboard seating positions; driver and front passenger knee	●	●	●	●	●
SAFETY & SECURITY (CONTINUED)					
Brakes: 4-wheel antilock, 4-wheel disc; Duralife brake rotors with corrosion-resistant technology	●	●	●	●	●
OnStar Guidance Plan[4] with limited service trial, including Automatic Crash Response, Stolen Vehicle Assistance, Roadside Assistance, Turn-by-Turn Navigation, Advanced Diagnostics and more (trial excludes Hands-Free Calling Minutes)	●	●	●	●	●
Teen Driver technology	●	●	●	●	●
Tire Pressure Monitoring System (excludes spare tire)	●	●	●	●	●
INTERIOR					
Climate control: Single-zone, manual	●	●	●	●	●
Single-zone, automatic	–	–	–	–	○[5]
Console: Center, floor-mounted	●	–	–	–	–
Floor-mounted with armrest (sliding armrest on manual transmission models)	–	●	●	●	●
Cruise control: Electronic with Set and Resume Speed; steering wheel-mounted controls	–	–	●	●	●
Driver Information Center: Monochromatic display	●	●	●	●	●
4.2-inch diagonal color display	–	–	○[6,7]	○[6,7]	○[6]
Floor mats: Carpeted, front and rear	–	●	●	●	●
Rear vision camera	●	●	●	●	●
Remote Keyless Entry	●	●	●	●	●
Remote vehicle starter system	–	–	○[7,8]	○[7]	●
Steering column: Tilt and telescopic	●	●	●	●	●
Steering wheel: Mounted audio and phone controls	–	–	●	●	●
Heated, leather-wrapped	–	–	–	○[9]	●
Sunroof: Power, tilt/sliding	–	–	○[7,10]	○[7,11]	○[12]
Wireless charging[13] for devices	–	–	–	–	○[5]
SEATING					
Front buckets with reclining seatbacks and adjustable head restraints	●	●	●	●	●
Rear folding seat with adjustable head restraints for outboard seating positions	●	●	–	–	–
Rear 60/40 split-folding seat with adjustable head restraints for outboard seating positions and center fold-down armrest with two cup holders	–	–	●	●	●
Heated seats: Front	–	–	○[7]	●	●
Rear outboard	–	–	–	–	○[5]
Driver-seat adjuster: 6-way manual	●	●	●	–	–
8-way power	–	–	○[7]	●	●
Front passenger seat adjuster: 2-way manual	●	●	●	●	–
4-way manual	–	–	–	○[9]	●
Seat trim: Cloth	●	●	●	●	–
Leather appointments	–	–	–	○[9]	●
ENTERTAINMENT					
Chevrolet MyLink[14] Radio with 7-inch diagonal color touch-screen display, AM/FM stereo, Bluetooth audio streaming[15] for select phones, voice-activated technology, and Android Auto™[16] and Apple CarPlay™[17] compatibility	●	●	●	●	●
Chevrolet MyLink[14] Radio with 8-inch diagonal color touch-screen display, includes features listed above; also includes the ability to browse, select and install apps on the MyLink system with Shop[18]	–	–	○[6]	○[6]	–
Chevrolet MyLink[14] Radio with Navigation and 8-inch diagonal color touch-screen display, includes features listed above plus enhanced navigation and SiriusXM NavTraffic[19] with 3-month trial subscription	–	–	–	–	○[6,12]
Bose® premium 9-speaker sound system (Hatchback includes 8-speaker system)	–	–	○[10]	○[6]	○[6]
SiriusXM Satellite Radio[19] All Access Package with 3-month trial subscription	–	–	●	●	●

such as "Exterior," "Engine/Chassis," "Safety & Security," "Interior," etc. Typically, each vehicle feature category has several feature subcategories. For example, the "Engine/Chassis" category has four subcategories: "Engine," "Steering," "Suspension," and "Transmission." A typical vehicle model has more than 100 subcategories under consideration when making content decisions. Within a feature subcategory, there are several *attributes*. An attribute denoted by a solid dot is standard, a circle is optional, and a dash means not available on a given trim level. The superscripts reflect certain restrictions that the product team has imposed on the offering. For instance, fog lamps are offered in a vehicle only if it is also equipped with a rear spoiler and a sport body kit.

This example highlights that there are combinatorially many possible ways to construct a content portfolio. In the past, GM's product teams relied heavily on their experience and judgment by looking back at historical sales (much like using the rearview mirror) or looking around at what competitors offered (much like using the side-view mirrors). What was missing is the forward-looking windshield view of what customers want for their next new vehicle. Based on the feedback from GM internal interviews with subject matter experts (SMEs), it was suggested that GM should build what customers want, not just mimic what competitors do. SMEs also revealed the challenges that arise because there are competing objectives within a product team as a result of different business functions (e.g., marketing, engineering, finance) being evaluated based on different business objectives. Overall, there was a lack of analytical support for SMEs to understand customers' preferences for vehicle features and their impact on customers' purchase decisions, to consider the impact of competition and dealer stocking behaviors, and to balance the competing objectives of maximizing profit and market share and minimizing build complexity (e.g., the number of vehicle configurations). When poor content decisions are made, they present both a risk of losing customers and a risk of giving away content at prices below what customers would have been willing to pay. Both risks have the potential to result in lost profit and market share opportunities.

VCO provides quantitative performance measures of profit, sales volume, and market share that are data driven and transparent and that speak more comprehensively and objectively than expert opinions and "guestimates." For example, one of the midsized SUV product teams had an internal debate about whether to offer an optional driver power seat on the middle trim. Notice that when a feature is added on a trim as optional, it doubles the build complexity of that trim. The Engineering team proposed removing that option in an attempt to reduce build complexity. However, the Marketing team was very concerned about removing it because they strongly believed some middle-trim customers would want to have a power seat for the driver. If the option was removed, those customers might not want to pay extra money to move up to the next trim; instead, they might choose to go to competitors and become a loss to GM. The simulation results from the VCO analysis showed that when removing the optional driver power seat from the middle trim, the build complexity would be slightly reduced as expected; however, both the estimated market share and profit would be significantly decreased. In the end, the cross-functional product team assessed the trade-offs and decided to keep the driver power seat as optional on the middle trim. This is one of many testimonies to how critical VCO analysis is for GM in vehicle contenting decisions.

Vehicle Content Optimization Process

Figure 2 illustrates the key steps of VCO and how outputs from one step are used as inputs for the next step. This provides an overview of how uncertainty from different sources is propagated to the performance measures computed at the end of the VCO process. Our uncertainty quantification focuses on the simulation analysis step. However, some inputs to the simulation analysis step are outputs from previous steps of VCO. Hence, it is critical for understanding of our approach to describe each step of VCO in detail.

Conjoint Study

Input: Conjoint Design, Respondent Group; Output: Survey Data. VCO starts from a conjoint study (Orme 2006), a survey technique used to assess consumers' preferences for attributes of a product or service. Prior to the study, GM's marketing team carefully designs the questionnaire referred to as a "conjoint design" to effectively study the respondents' preferences for content features of the target vehicle model. The respondents are chosen so that the group represents the target market segment of the vehicle model. Of course, a larger group is preferred; however, recruiting qualified respondents is expensive, and a study with a larger respondent group tends to take longer. In practice, the size of the group is chosen to balance several cost factors and is taken as given in our work.

In each survey question, several combinations of vehicle features and their corresponding prices are presented. Table 1 shows an example of the conjoint study questions used for the Chevrolet Cruze, whose partial content portfolio is presented in Figure 1. Each respondent is asked to select his or her most preferred combination and whether he or she would consider the selected vehicle for purchase. Typically, each respondent receives 20–30 such questions. Asking all possible combinations of attributes in Figure 1 is impossible, and only a subset of feature attributes is presented each time. These questions are carefully designed to study the respondents' trade-offs between feature preferences and price sensitivity when they make a purchase decision. As a result, the survey data help GM quantify consumer preferences for both features and price.

Choice Model Estimation via Markov Chain Monte Carlo

Input: Survey Data; Output: β Sample Library. The survey results from the conjoint study are used to estimate the distribution of the respondents' utilities for the feature attributes. Here, utility is a numerical value assigned to each vehicle feature attribute to represent the respondent's relative preference. At GM, the vector of all respondents' utilities for all feature-attribute combinations is referred to as *beta* (β). GM takes a Bayesian approach to estimate the distribution of β (Train 2003). That is, a prior distribution on β is assumed first, and then its posterior distribution is updated conditional on the conjoint survey data. The survey data only inform the respondents' choices, not their utilities for features. What connects their choices and utilities is a *choice model*; GM adopts a logit model (see the appendix) to represent the probability of each respondent selecting

Figure 2. The Vehicle Content Optimization Process at GM

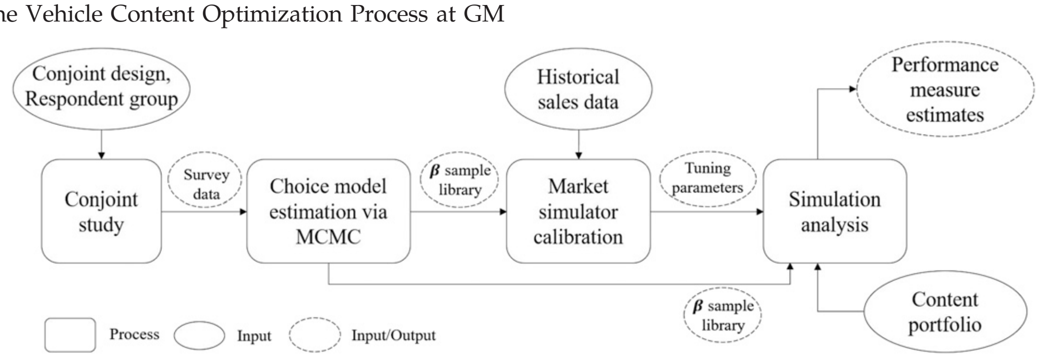


Table 1. An Example of the Conjoint Study Question for the Chevrolet Cruze

Car 1	Car 2	Car 3	Car 4
Cruise control	No cruise control	No cruise control	Cruise control
No fog lamps	No fog lamps	Fog lamps	Fog lamps
1.4L DOHC turbocharged 4-cylinder with 153 hp and 177 lb.-ft. of torque	1.6L DOHC turbocharged 4-cylinder with 153 hp and 177 lb.-ft. of torque	1.6L DOHC turbocharged 4-cylinder with 153 hp and 177 lb.-ft. of torque	1.4L DOHC turbocharged 4-cylinder with 153 hp and 177 lb.-ft. of torque
Single-zone, automatic climate control	Single-zone, automatic climate control	Single-zone, automatic climate control	Single-zone, automatic climate control
Door handles: body color with chrome strip	Door handles: body color	Door handles: body color with chrome strip	Door handles: body color
No remote vehicle starter system	No remote vehicle starter system	Remote vehicle starter system	Remote vehicle starter system
6-speed manual	6-speed automatic	9-speed automatic	6-speed automatic
Driver information center: monochromatic display	Driver information center: 4.2-inch diagonal color display	Driver information center: 4.2-inch diagonal color display	Driver information center: 4.2-inch diagonal color display
Heated, leather-wrapped steering wheel: mounted audio and phone controls	Steering wheel: mounted audio and phone controls	Steering wheel: mounted audio and phone controls	Heated, leather-wrapped steering wheel: mounted audio and phone controls
Sunroof: power, tilt/sliding	Sunroof: power, tilt/sliding	No sunroof	No sunroof
Wireless charging for devices	No wireless charging for devices	No wireless charging for devices	No wireless charging for devices
Price: $17,200	Price: $20,500	Price: $19,900	Price: $21,800

Notes. The respondent was asked to select one of four cars that best fits his or her needs and an extra question: "If the vehicle you selected above were a real vehicle and the rest of its features were acceptable to you, would you consider purchasing this vehicle?" (yes/no). DOHC, dual overhead camshaft; hp, horsepower.

an alternative among candidates. The logit model lets us write the likelihood function of $\boldsymbol{\beta}$ given the observed choices in the survey data. Combining the likelihood function with the prior distribution of $\boldsymbol{\beta}$, the posterior distribution of $\boldsymbol{\beta}$ is obtained. Although it is difficult to derive the closed-form expression for the posterior distribution, we may obtain samples from it via Markov chain Monte Carlo (MCMC). Given the high dimension of $\boldsymbol{\beta}$, it typically takes a long time for the MCMC procedure to converge to its posterior distribution. Once converged, a finite number of $\boldsymbol{\beta}$ draws is generated from the MCMC simulation and saved. Any successive $\boldsymbol{\beta}$ s sampled via MCMC are highly correlated. Therefore, in between two saved $\boldsymbol{\beta}$ draws, a large number of $\boldsymbol{\beta}$ s are discarded to obtain an approximately independently and identically distributed (i.i.d.) set of $\boldsymbol{\beta}$ s, which makes this procedure computationally expensive. We refer to the collection of saved $\boldsymbol{\beta}$ draws as the $\boldsymbol{\beta}$ *sample library*.

In the subsequent VCO steps, the posterior distribution of $\boldsymbol{\beta}$ is approximated by the empirical distribution of $\boldsymbol{\beta}$ s in the $\boldsymbol{\beta}$ sample library—that is, the $\boldsymbol{\beta}$ draws in the library represent all possible realizations of $\boldsymbol{\beta}$ given by the posterior distribution, and we sample from the library with replacement instead of sampling from $\boldsymbol{\beta}$'s posterior. Of course, the larger the $\boldsymbol{\beta}$ sample library is, the better it approximates the posterior distribution of $\boldsymbol{\beta}$. However, it takes a long time to generate approximately i.i.d. $\boldsymbol{\beta}$ draws because of the correlation issue, and the high dimension of $\boldsymbol{\beta}$ makes storing a large number of $\boldsymbol{\beta}$ draws cumbersome. Typically, a few hundred $\boldsymbol{\beta}$ draws are saved to the library.

Market Simulator Calibration

Input: $\boldsymbol{\beta}$ Sample Library and Historical Sales Data; Output: Tuning Parameters. The market simulator, which is a core component of VCO, simulates the dynamics of customer choice. For the simulator to produce a meaningful output (e.g., market share, sales volume), *calibration* is crucial. The survey data from the conjoint study represent the *stated* preferences of respondents. Thus, running the simulation with $\boldsymbol{\beta}$ draws means we rely on the stated preferences. On the other hand, we can also observe the *revealed* preferences from historical sales data that detail which vehicles were purchased by customers at what prices and which other vehicles (of the same model) were on the dealers' lots at the time of the purchases. In the conjoint analysis literature, it is well understood that there typically is a gap between respondents' stated willingness to pay in the survey and their actual spending (Louviere et al. 2000). To bridge this gap, VCO calibrates a set of simulation parameters so that the simulated sales outcome matches the historical sales data. The resulting calibrated simulation parameters, $\boldsymbol{\theta}$, are called "tuning parameters" in VCO. Calibrating the simulator involves optimizing a loss function that represents the discrepancy between the historical sales data and the simulated results. The loss function is nonconvex in $\boldsymbol{\theta}$ and can only be evaluated (with stochastic error) by the market simulator. Thus, there is no guarantee that a local minimum is the global minimum. To avoid being stuck at a local minimum, there is extra randomness induced in the optimization process controlled by a pseudorandom number seed to make the algorithm randomly search less explored feasible regions of $\boldsymbol{\theta}$. This class of simulation optimization algorithm is referred to as *random search* (Andradóttir 2015); thus, we refer to the pseudorandom number seed that controls the search as the "random search seed." VCO combines several random search algorithms to increase the effectiveness of calibration.

In fact, the calibration step is a major bottleneck of VCO as each loss function evaluation requires a single simulation run. As a result, it may take orders of magnitude longer than a simulation run (depending on the complexity of the content portfolio of the simulated vehicle model) until the loss function converges to a reasonable range. The execution time of a single simulation run is roughly linearly increasing in the number of $\boldsymbol{\beta}$ draws saved in the $\boldsymbol{\beta}$ sample library. To alleviate the computational burden, VCO takes a subsample of $\boldsymbol{\beta}$ s from the library to represent the posterior distribution of $\boldsymbol{\beta}$ instead of using the entire library. We refer to this subsample as the *calibration subsample of* $\boldsymbol{\beta}$. Because there is no guarantee the loss function converges to 0 as calibration continues, a time budget is set for calibration, and the best tuning parameter values are retained at termination. This means that the calibration may end prematurely before even reaching a local minimum. In fact, owing to the highly nonconvex nature of the loss function, it is unlikely that the calibration procedure reaches the global optimum given finite calibration time. The resulting $\boldsymbol{\theta}$ from a calibration procedure is a function of the calibration time as well as random search seed. Such dependence combined with the nonconvexity of the loss function makes $\boldsymbol{\theta}$ random, and ultimately, it becomes a source of uncertainty in the simulated performance measure estimates of VCO.

Simulation Analysis

Input: β Sample Library, Tuning Parameters θ, and Content Portfolio; Output: Performance Measure Estimates. In the final step of VCO, the calibrated market simulator is used to evaluate different content portfolios to identify the ones that yield high performance measures (e.g., sales volume, profit, market share). We focus on the sales volume here because profit and market share can be directly computed from the sales volume.

The simulator consists of two main parts. The first part simulates the purchase behavior of customers in the market segment. Within each simulation run, each customer forms a candidate set of vehicles from which they make a purchase decision (including no purchase). This candidate set includes different combinations of vehicle content a customer would see on dealer lots constructed by sampling from all possible content combinations prescribed by the content portfolio (e.g., Figure 1) with a probability distribution representing the likelihood of each combination being stocked at a dealership. The details of the behavioral model that generates the customers' candidate vehicle sets and the dealerships' stocking likelihood are GM proprietary information and cannot be disclosed.

The second part of the simulator computes the sales volume of the simulated customers based on their purchase behavior. Fundamentally, the simulator uses a logit choice model similar to the one presented in the appendix to predict each individual customer's purchase decision. The difference is that Cars 1–4, the alternatives presented in the conjoint question, are replaced by the customer's candidate vehicle set plus a no purchase option. Once the probability of purchasing a GM vehicle is computed, it can be regarded as the expected "volume" of the GM vehicle this customer purchases. Adding these expected purchases from all simulated customers and scaling to the entire U.S. market segment, the total sales volume is obtained. Note that the $\boldsymbol{\beta}$ draws we obtained from the second step play an important role in computing the sales volume as they are fed into the logit model.

There are three computational compromises made in the simulation step. First, as in the calibration step, a subsample of the $\boldsymbol{\beta}$ sample library, which we refer to as the *simulation subsample of* $\boldsymbol{\beta}$, is selected to represent the market preferences. Again, this is to save the computational cost of each simulation run. Typically, GM used the same subsample for both calibration and simulation. However, we distinguish them because we ended up recommending using different sets of $\boldsymbol{\beta}$ draws for calibration and simulation as a result of this study.

Second, when each simulated customer's candidate vehicle set is formed, it is difficult to consider all possible combinations of feature attributes in the content portfolio because of its combinatorial nature. As a result, the simulator first subsamples a smaller set of combinations and uses them as a stand-in for the content portfolio to make the simulation more efficient. This subset is referred to as the *product library*. Clearly, sampling a product library introduces an extra layer of uncertainty. Finally, when each customer's candidate vehicle set is constructed, the simulator samples it from an approximate distribution of the candidate vehicle sets instead of the exact distribution because of computational complexity. There is no analytical expression for the candidate vehicle set distribution; instead, it is defined implicitly by a GM proprietary behavioral model inside the simulator. Thus, the simulator takes an approach to pregenerate a large number of candidate vehicle sets at the beginning of the simulation and sample from them every time a random customer is simulated. Such pregeneration is controlled by a pseudorandom number seed, which we refer to as the *vehicle set generation seed*; thus, the sampled candidate vehicle sets depend on the vehicle set generation seed, which becomes another source of uncertainty.

In addition to these three computational compromises, uncertainty is propagated from the calibration step because tuning parameters $\boldsymbol{\theta}$ obtained from the calibration step are fed into the simulation step. Such compounding effects make uncertainty quantification of VCO challenging.

Sources of Uncertainty in the VCO Process

We summarize the sources of uncertainty in the VCO process here. There is a distinction between uncertainty caused by the natural variability that arises in the real-world vehicle market and the estimation error introduced by finite computational effort or data.

1. Natural variability
   - Customers' individual preferences within the market segment
   - Dealerships' individual variability in stocking behavior
   - Customers' individual variability in purchase behavior
2. Estimation error
   - Nondegenerate posterior distribution of $\boldsymbol{\beta}$ reflecting imperfect learning of a respondent's preferences
   - *Error introduced by representing the posterior distribution of* $\boldsymbol{\beta}$ *with the* $\boldsymbol{\beta}$ *sample library, calibration subsample, and simulation subsample*
   - *Randomness in tuning parameters* $\boldsymbol{\theta}$ *caused by nonconvexity of the loss function, finite calibration time, and use of random search controlled by the random search seed*
   - *Error caused by sampling a product library from a content portfolio*
   - *Error caused by approximating the candidate vehicle set with an empirical distribution of pregenerated vehicle sets whose randomness is controlled by the vehicle set generation seed*

Identifying the sources of natural variability is important as it allows GM to measure their business risk as a result of market randomness; however, it cannot be reduced by expending more computation effort. On the other hand, estimation error as a result of sampling $\boldsymbol{\beta}$ s and the product library, randomness in tuning parameters, and approximation error of the candidate vehicle set distribution can be reduced by increasing the computational effort. We narrowed our focus to these four sources of uncertainty (in italics) to quantify how much impact they have on the overall uncertainty of the performance measure estimates from the VCO.

That said, not all estimation error can be reduced, given the scope of our study. For example, perfectly learning each respondent's preferences is impossible given a finite number of conjoint questions. Such uncertainty is reflected in the posterior distribution of the $\boldsymbol{\beta}$ fitted from the conjoint data. In other words, without running a new conjoint study with additional questions, we cannot reduce this uncertainty further, which was outside the scope of our analysis.

Mathematical Representation of Input-Output Relationships

After identifying the target sources of uncertainty, we established a mathematical model that represents the input-output relationships for each step of VCO. See the appendix for the full description of the model. This helped us rigorously and unambiguously identify how the different sources of uncertainty in VCO propagate to the performance measure estimates. Here, we summarize major insights obtained from mathematical modeling.

First, the set of local optima of the loss function for calibration is conditional on the calibration subsample of $\boldsymbol{\beta}$. If the random search algorithm used in calibration converges to a local optimum as the calibration time increases without a bound, then $\boldsymbol{\theta}$ will converge to one of the local optima in the set; the particular local optimum it converges to depends on the initial solution as well as the random search seed. In other words, even with infinite calibration time, the variability of $\boldsymbol{\theta}$ is nonzero. Nonetheless, we may postulate that the variability of $\boldsymbol{\theta}$ is a decreasing function of the calibration time.

Second, we hypothesized that the variability of $\boldsymbol{\theta}$ is larger when the calibration subsample is smaller. Intuitively, if we used an infinitely large set of $\boldsymbol{\beta}$ for calibration, then it would perfectly represent the posterior distribution of $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ would no longer depend on the calibration subsample. For this reason, one of GM's main concerns was how large the calibration subsample should be so that its impact on the variability of $\boldsymbol{\theta}$ is small while obtaining a good calibration result in a reasonable length of time. We later present a preliminary study that indicates that the variability of $\boldsymbol{\theta}$ is less sensitive to the calibration subsample size than to the random search seed.

Finally, the total sales volume of a content portfolio is estimated from the sum of sales of each content combination in the product library. To compute the expected total sales volume, the simulator averages the total sales volume across all $\boldsymbol{\beta}$ draws in the simulation subsample of $\boldsymbol{\beta}$. Thus, the variance of the estimated expected sales volume is inversely proportional to the simulation subsample size. Because simulation time increases linearly in the simulation subsample size, there is a trade-off between computational cost and output uncertainty.

Preliminary Study

Although the mathematical representation provides a framework for uncertainty quantification, we had little knowledge about the computational cost of each VCO step, which is crucial to consider for providing an actionable solution for GM given their limited time budget. Also, we wanted to have a deeper understanding of the calibration procedure because it is the bottleneck of VCO and affects the subsequent simulations as uncertainty in $\boldsymbol{\theta}$ is propagated to the performance measure estimates.

To answer these questions, we conducted a series of experiments using one of GM's past vehicle models. The particular model was chosen because it had a relatively small content portfolio and modest conjoint study data, which made MCMC, calibration, and simulation analysis faster than other vehicle models, and therefore, we could run many otherwise computationally demanding exploratory experiments. Because the insights we gained from this preliminary study were crucial for shaping our approach to the problem, we highlight its major findings in the following subsections before we introduce our methodology.

Local Optimality of Calibration Causes Large Variability in Performance Measure Estimates

As discussed earlier, we conjectured that the variability of $\boldsymbol{\theta}$ is a decreasing function of the calibration subsample size. The VCO team was concerned that using a smaller calibration subsample may cause the tuning parameters to be biased. Therefore, we performed calibrations with multiple calibration subsamples with four different sizes, 25, 50, 75, and 100, while setting the calibration time large enough so that the loss function evaluated at termination was small for all sizes. Different random search seeds were used for all calibration runs. For each tuning parameter, we examined the trend in the sample means and the marginal[1] sample variances of the tuning parameters as a function of the calibration subsample size but did not find a statistically significant trend in either. The observed insensitivity of the sample means of the tuning parameters with respect to the size of the calibration subsample assured the VCO team that the calibration subsample need not be too large for the purpose of reducing the bias in $\boldsymbol{\theta}$. Nevertheless, the tuning parameters had large marginal variances, which was a result of the nonconvexity of the loss function as well as randomness in drawing a calibration subsample.

To separate these two effects, we fixed the calibration subsample size and performed two sets of experiments. For the first set, we used the same calibration subsample and ran n calibrations using different random search seeds. For the second set, we drew n calibration subsamples from the $\boldsymbol{\beta}$ sample library and paired each subsample with a different random search seed. From the marginal variances of the resulting tuning parameters, we learned that the second set of experiments yields slightly more variable tuning parameters; however, the differences were statistically insignificant for most of the tuning parameters.

This result was quite a surprise to the VCO team, as they had expected drawing a calibration subsample to be a significant source of variability for $\boldsymbol{\theta}$. Because the ultimate goal was to quantify uncertainty in the estimated performance measures, we performed two sets of simulation experiments to measure variability in the expected sales volume estimate. Two sets of size n simulations were run, where the first (second) set was run using n tuning parameters obtained from the first (second) set of calibrations. In each run, the simulation subsample was chosen to be identical to the calibration subsample. The first set of simulations estimates variability caused purely by the random search seed, whereas the second estimates variability caused by both random search seed and the calibration subsample. Both variances were estimated to be large, but the former was less than 50% of the latter. This shows that (1) the variability in the tuning parameter caused by multiple local optima of the loss function indeed induces large variability in the performance measure estimate, and (2) using a finite sample of $\boldsymbol{\beta}$ s causes large variability in the simulation results, although not so much in the tuning parameters.

Notice that in these experiments, the same $\boldsymbol{\beta}$ subsample was used for each calibration-simulation pair, which was consistent with the VCO team's protocol. However, from our calibration experiments, we learned that $\boldsymbol{\theta}$ is less sensitive to the particular choice of calibration subsample or its size, which means the calibration subsample need not be too large. On the other hand, the variance of estimated expected sales volume is inversely proportional to the size of the simulation subsample. Thus, *we recommended GM use a simulation subsample larger than the calibration subsample.*

Calibration Time Matters

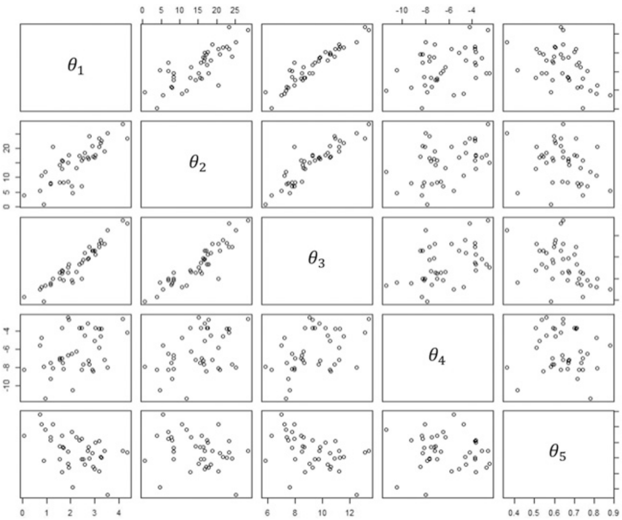
For the preliminary study, we made multiple calibration runs to examine the variability of $\boldsymbol{\theta}$. In practice, however, the VCO team has limited time for analysis. Given a time budget, there is a trade-off between running a single long calibration and running multiple short calibrations to estimate how much variability in the performance measure estimate is caused by the random search seed. Although it is clear that local optimality of $\boldsymbol{\theta}$ is a significant source of uncertainty from the above-mentioned experiments, the tuning parameters from a short calibration procedure may be far from any local optimum of the loss function; hence, the resulting performance measure estimate may be biased. Moreover, if we perform multiple calibrations, then we would probably want to pick the $\boldsymbol{\theta}$ that has the smallest loss function value as the single "best" tuning parameter instead of using them all. Therefore, we decided to draw a single calibration subsample and perform a long calibration run so that a good set of tuning parameter values can be obtained.

A question remains, however: How long is long enough? The VCO team typically set an absolute bound on the loss function as a stopping criterion; for example, the loss function value drops below 1. However, such a cutoff is somewhat arbitrary and may not be attainable for some vehicle models, as there is no guarantee that the loss function is less than 1 at the global minimum. Instead, we devised an iterative calibration method that increments calibration time until the estimated performance measure is no longer sensitive to the change in the tuning parameters, as discussed later.

Dimensionality of Calibration Can Be Reduced

While analyzing the joint distribution of the tuning parameters obtained from the calibration experiments, we discovered dependence among tuning parameters. Figure 3 shows pairwise scatterplots of the five elements of $\boldsymbol{\theta}$, $\theta_1 - \theta_5$.[2] Clearly, there are strong linear relationships among the first three parameters, whereas other pairs do not exhibit statistically significant relationships. Such trends were consistent across different calibration subsamples of different sizes. This implies that at least one of θ_1, θ_2, and θ_3 can be removed to reduce the number of parameters for the calibration procedure. The benefit of dimensionality reduction is twofold: it speeds up the calibration optimization procedure by reducing the number of decision variables, and it may reduce uncertainty in the estimated performance measures when we have fewer tuning parameters. Adopting our recommendation, GM carefully examined redundancy of tuning parameters in its model and reduced its dimension in the next generation of VCO.

Figure 3. Dependence Among Five Tuning Parameters, θ_1, θ_2, …, θ_5.



Note. There are strong linear relationships among θ_1, θ_2, and θ_3.

Uncertainty Quantification with First-Order Effects

The *first-order effect* is a global sensitivity index that quantifies how much of the simulation output variance is caused by each random input in isolation when there are multiple inputs. When there are k variable inputs to a simulator, the first-order effect of the ith input is defined as the variance of the conditional mean of the simulation output given the ith input. Because the conditional mean averages out all other inputs' variability, the first-order effect captures the portion of the variance of the simulation output solely from the ith input's variability.

For VCO, the first-order effect can be used to measure the impact of four inputs—the tuning parameters $\boldsymbol{\theta}$, the simulation subsample of $\boldsymbol{\beta}$, the product library sampled from the content portfolio, and the vehicle set generation seed that controls the randomness in approximating the candidate vehicle set distribution—on the performance measure estimates. Among these four inputs, $\boldsymbol{\theta}$ is the most expensive to "sample," as it involves running multiple calibrations, whereas drawing a simulation subsample of $\boldsymbol{\beta}$, sampling a product library, or choosing a different vehicle set generation seed is cheap. As learned from the preliminary study, running multiple calibrations given a time budget was not desirable. Therefore, we decided to separate out $\boldsymbol{\theta}$ from the other inputs and perform *global sensitivity* analysis with respect to the remaining three inputs and analyze the *local sensitivity* of the performance measure estimates with respect to $\boldsymbol{\theta}$.

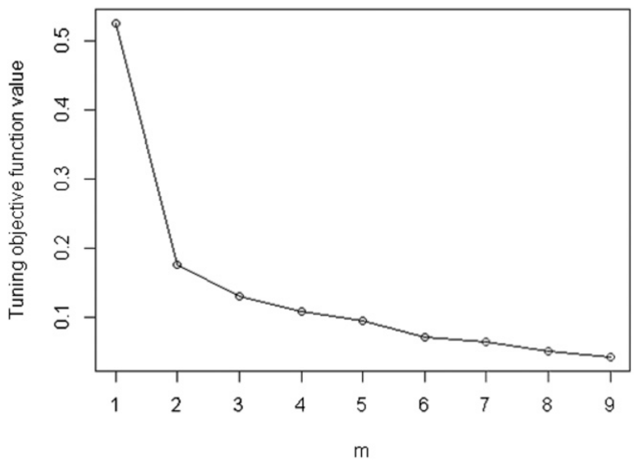
In the appendix, we show a mathematical representation of the first-order effects with respect to the three inputs. A naive estimation approach is to first estimate the inner conditional expectation by running N_1 simulations, and then estimate the variance of N_2 conditional expectation estimates. However, this approach requires $3N_1N_2$ simulations to compute all three first-order effects, where N_1 and N_2 should be reasonably large. Instead, we follow a more efficient estimation method by Saltelli (2002) that computes all three first-order effects from $5N$ simulations for some reasonably large N. From these runs, we can also estimate the overall variance of the sales volume caused by all three inputs. See the appendix for details.

If the estimated overall variance is small anyway, then GM would not be too concerned about the estimated first-order effects. When it is large, however, GM can compare the relative magnitudes of the first-order effects to investigate which source of uncertainty causes the largest uncertainty in the simulation output. If the first-order effect of the simulation subsample of $\boldsymbol{\beta}$ is the largest, then the variance of the expected sales volume estimate can be decreased by increasing the size of the simulation subsample (recall that the variance is inversely proportional to the subsample size). If the first-order effect of the product library is the largest, then increasing the product library size is effective so that the content portfolio is better represented by each sampled product library. In the preliminary study, we empirically confirmed that while fixing other inputs, the variance of the estimate of the expected sales volume decreases proportionally to the product library size as long as the product library is reasonably small relative to the content portfolio. If the first-order effect of the vehicle set generation seed is the largest, then we can increase the number of the candidate vehicle sets the simulator internally creates to obtain a better approximation of the candidate vehicle set distribution. However, based on the experiments, this first-order effect tends to be negligible compared with the other two sources of uncertainty.

Iterative Calibration and Local Sensitivity to Tuning Parameters

As noted earlier, a longer calibration run is preferable to repeating multiple short runs. The remaining question was how to choose the calibration time. We approached this problem by segmenting the calibration time into smaller intervals with length t. At the end of each time t interval, a stopping criterion is applied to decide whether to continue or not. Figure 4 shows an example of the trajectory of the loss function value at the end of each time t interval. Notice that the trajectory is a nonincreasing function of time because,

Figure 4. An Example of the Trajectory of the Objective (Loss) Function Value at the End of Each Time t Interval



by design, the optimization procedure for calibration never takes a step toward an increasing direction.

There are several possible stopping criteria. For instance, one could stop at the end of an interval if reduction in the loss function value in that interval is less than a threshold. Because this stopping criterion is based on the relative improvement in the loss function during each interval and the loss function is nonnegative, the criterion is expected be satisfied after some finite number of intervals. Either reduction in the loss function value or the loss function value itself will fall below the threshold after some iterations; in the latter case, there is no need to further calibrate. On the other hand, GM's original criterion to stop when the loss function falls below 1 may not be satisfied for the given vehicle model even if the calibration is run indefinitely.

Although our suggestion is an improvement, this stopping criterion is based on reduction in the loss function during an interval. A more important question is, how much difference in the *performance measure estimates* is there if we continue calibrating for another interval? Suppose calibration is terminated at the end of the mth iteration, and $\boldsymbol{\theta}_m$ is the vector of tuning parameters to be used for simulations. We can perform simulations using both $\boldsymbol{\theta}_{m-1}$ and $\boldsymbol{\theta}_m$ and compare the performance measure estimates. If this difference is small, then the performance measure is likely to change little even if we continue calibrating for another time t interval and update $\boldsymbol{\theta}_m$ to $\boldsymbol{\theta}_{m+1}$. Because running simulations at an additional set of tuning parameters is much cheaper than running an additional iteration of calibration, the computational cost for such local sensitivity check is negligible.

Quantifying Bias as a Result of Product Library Sampling

In addition to variability, sampling the product library induces bias in the expected sales volume estimate. This bias is always negative because sales volume as a result of combinations of features not included in the product library is missed. Unlike variance, such bias cannot be "averaged away" by running more simulations. Our solution is to correct the bias directly by estimating it as a function of product library size.

We conjectured that the bias can be effectively approximated by a quadratic function of the reciprocal of the product library size if all other inputs are fixed. The coefficients of the quadratic function can be estimated via least squares regression by running simulations with product libraries of different sizes; see the appendix for details.

Applications

In this section, we demonstrate the proposed uncertainty quantification procedure on the same vehicle model we used for the preliminary study. For iterative calibration, we used a calibration subsample of 50 $\boldsymbol{\beta}$ draws and 10-minute intervals. For the stopping criterion, we used the threshold of 0.01. We also set the minimum calibration time to 90 minutes—that is, the calibration can be stopped using the stopping criterion only after this time. Figure 4 represents the loss function of the calibration procedure at the end of the mth interval for $m = 1, 2, \ldots, 9$. Because the reduction in the loss function at the ninth interval is $0.009 < 0.01$, the calibration stopped after 90 minutes ($m = 9$), and $\boldsymbol{\theta}_9$ was used for uncertainty quantification, whereas $\boldsymbol{\theta}_8$ was saved for local sensitivity analysis with respect to the tuning parameters. Notice that with GM's original criterion, we would have stopped calibrating after 10 minutes, as the loss function value fell below 1, which is clearly a premature termination in this case.

We drew a simulation subsample of size 50 and sampled 306 of 9,216 configurations in the content portfolio to include in the product library. To compute the first-order effect, we chose $N = 10$, which amounts to total 50 simulation runs. For the performance measure, we computed the expected market segment share of GM, which is a simple function of the sales volume (i.e., the ratio of the sales volume to the total market segment size). The mean share estimate and its standard error (squared root of its variance) were 22.8% and 0.19%, respectively. The latter quantifies uncertainty in the share estimate caused by all three input sources. The estimated normalized first-order effects of the simulation subsample of $\boldsymbol{\beta}$, product library, and vehicle set generation seed are 0.964, 0.032, and 0.003, respectively. These indicate that most of the uncertainty in the share estimate is caused by drawing the simulation subsample of $\boldsymbol{\beta}$, and if GM desires to reduce it, increasing the size of the subsample would be the most effective way.

To examine the sensitivity of the mean share estimate to the tuning parameters, we ran simulations using $\boldsymbol{\theta}_8$ while fixing other inputs exactly the same as those for $\boldsymbol{\theta}_9$. These simulations are much faster to run than calibrating for an additional 10 minutes. The estimated share and its standard error are 23.4% and 0.19%, respectively. To see whether the two share estimates are statistically significantly different, we set up a hypothesis test with the null being that they are equal. The test statistic for the two-sample t-test is $|23.4 - 22.8|/\sqrt{0.19^2 + 0.19^2} = 2.233$ resulting in a p value of 0.0315 (degrees of freedom = 38); hence, we concluded that longer calibration is required by choosing a stopping criterion threshold smaller than 0.01. This was a valuable guideline for GM; if we stopped calibrating using GM's original stopping criterion without checking the sensitivity of the simulation outputs, we would not have ended up with the same recommendation.

Finally, the bias correction scheme was also tested. For this particular content portfolio, the number of all possible configurations (9,216) was relatively small. To fit the quadratic model to estimate the bias via regression, we chose two additional product library sizes, 612 and 918, and we ran simulations to estimate the shares given all other inputs were fixed. The estimated shares for these two product library sizes were 23.26% and 23.53%, respectively. After estimating the coefficients of the quadratic model via least squares, the bias-corrected estimate of share was obtained as 24.16%. Compared with the share estimate with $p = 918$, we closed the bias by 24.16% − 23.53% = 0.63%. This is significant bias reduction, considering it represents GM's U.S. market segment share, and the corresponding revenues are vastly different.

Conclusions of the Study

The VCO process supports one of GM's major business decisions—what content portfolios to offer to the market, which directly impacts their market share and profit as well as consumer satisfaction. However, because of the VCO's expensive computation requirements, several computational compromises were made without explicitly quantifying uncertainty in the estimated performance measures that such compromises would induce, which ultimately may pose business risk. The goal of this research was to develop a method to systematically quantify uncertainty in the performance measure estimates and diagnose which sources of uncertainty have the largest influence on the overall uncertainty so that targeted computational effort can be expended.

We identified four major sources of uncertainty that can be reduced by increasing computational effort. Of these sources, we separated calibration from the rest and used the first-order effects to quantify the contribution of each source of uncertainty to the overall uncertainty in the performance measure estimates. Then, the sensitivity to the tuning parameters is measured locally to decide whether further calibration effort is needed. We also proposed a model to estimate the bias from finite product library sampling, which can make GM's performance measure estimate more accurate without increasing simulation effort much.

Lessons Learned

We share several valuable lessons learned from this research that can be appreciated by others who study uncertainty quantification of complex systems simulation.

- Establishing the mathematical representation of input-output relationship of the VCO process helped us identify the sources of uncertainty we can and cannot measure and thereby let us focus on the measurable sources of uncertainty. This also provided GM a "big picture" description of VCO that is free of the complicated details of each step.
- A thorough preliminary study was a key to our success; we better understood VCO and its associated computational costs and were able to devise an uncertainty quantification procedure that is computationally efficient and makes targeted suggestions to reduce uncertainty most effectively. Had we tried to develop the procedure in the abstract, we would certainly not have created the same procedure.
- Computational constraints matter in practice; GM wanted an actionable solution to the uncertainty quantification problem. There are more theoretically elegant approaches available in the literature, but they often require significantly more extensive computation than what GM could afford. For this reason, GM sacrificed uncertainty quantification for faster content decision making. Our recommendations were first tested for their computational feasibility. To this end, we devised the hybrid local-and-global sensitivity analysis for VCO, which requires a marginal increase in computational cost beyond GM's current practice.
- Developing procedures that do not require customization for each vehicle model was our aim, so that they can be applied by any analyst at GM and produce objective results. Uncertainty quantification, iterative calibration, and bias reduction procedures are all designed to be applicable to any GM vehicle models without having to customize the parameters of the procedures to each model.

Acknowledgments

The authors thank Dr. Jonathan Owen for his executive support for the research collaboration as well as Dr. Philip Keenan, Dr. George Beaujon, and Dr. Rebecca Stockbridge for their assistance in past work review and providing test data.

Appendix

This appendix provides mathematical notation for key quantities and the details of our uncertainty quantification approach.

A.1. Vector Representation of β and Logit Model

Let $\beta_{r\ell}$ represent the rth respondent's utility for the ℓth vehicle feature for $r = 1,2,\ldots,R$ and $\ell = 1,2,\ldots,L$, where R and L are the number of respondents and the number of available feature attributes for this vehicle model, respectively. In particular, the Lth attribute is the price of the vehicle. Then, we may define $\boldsymbol{\beta}$ as anRL-dimensional vector: $\boldsymbol{\beta} = \{\beta_{r\ell}\}_{r=1,2,\ldots,R,\ell=1,2,\ldots,L}$.

GM adopts a logit model to represent the probability of each respondent selecting an alternative among candidates. For instance, the probability of the rth respondent selecting Car 1 out of four candidates in Figure 1 is modeled as

$$\frac{\exp\left(\sum_{\ell\in\text{Car1}}\beta_{r\ell}\right)}{\exp\left(\sum_{\ell\in\text{Car1}}\beta_{r\ell}\right)+\exp\left(\sum_{\ell\in\text{Car2}}\beta_{r\ell}\right)+\exp\left(\sum_{\ell\in\text{Car3}}\beta_{r\ell}\right)+\exp\left(\sum_{\ell\in\text{Car4}}\beta_{r\ell}\right)}.$$

A.2. Mathematical Representation of Key Inputs and Outputs

In the following, we denote the calibration subsample and the simulation subsample by $\mathcal{B}^c$ and $\mathcal{B}^s$, respectively. The pseudorandom number seeds for random search and candidate vehicle set generation are denoted by ω and γ, respectively. The sampled product library is represented by π, and $p = |\pi|$ is the number of unique combinations of features in the product library.

Calibration of the simulator yields a vector of tuning parameters, $\boldsymbol{\theta}$, which depends on calibration subsample $\mathcal{B}^c$, the calibration time T, and random search seed ω. Thus, $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathcal{B}^c, T, \omega)$. Note that $\boldsymbol{\theta}$ is a random vector because it is a function of ω as well as $\mathcal{B}^c$.

In the simulation analysis step, $\boldsymbol{\theta}$ becomes an input to the market simulator that generates realizations of sales volume, market share, and revenue for a given content portfolio. Let $\boldsymbol{\beta}_d \in \mathcal{B}^s$ be the dth draw in $\mathcal{B}^s$ and $\pi_\ell \in \pi$ the ℓth content combination in π for $\ell = 1,2,\ldots,p$. GM's sales volume when the customers' utilities are represented by $\boldsymbol{\beta}_d$ is computed as

$$Y_d(\boldsymbol{\theta},\boldsymbol{\beta}_d,\pi,\gamma) = \sum_{\ell=1}^{p} Y_{d\ell}(\boldsymbol{\theta},\boldsymbol{\beta}_d,\pi_\ell,\gamma), \tag{A.1}$$

where $Y_{d\ell}(\boldsymbol{\theta},\boldsymbol{\beta}_d,\pi_\ell,\gamma)$ is the simulated sales volume of the ℓth content combination. In other words, (A.1) is the sum of sales of all p content combinations in π. The expected sales volume is then estimated across all $D = |\mathscr{B}^s|$ draws by

$$\bar{Y}(\boldsymbol{\theta}, \mathscr{B}^s, \pi, \gamma) = \sum_{d=1}^{D} Y_d(\boldsymbol{\theta}, \boldsymbol{\beta}_d, \pi, \gamma)/D.$$

Thus, it is clear that the simulation time increases linearly in D with all else equal. On the other hand, the variance of $\bar{Y}(\boldsymbol{\theta}, \mathscr{B}^s, \pi, \gamma)$, conditional on $\boldsymbol{\theta}$, is proportional to $1/D$; that is, the standard error of the expected sales volume decreases as $1/\sqrt{D}$.

A.3. First-Order Effects for VCO

For simulation output $Y(X_1, X_2, \ldots, X_k)$ with k random inputs $X_1, X_2, \ldots, X_k$, the first-order effect of input X_1 is defined as

$$\mathrm{V}(\mathrm{E}[Y(X_1, X_2, \ldots, X_k)|X_1]). \tag{A.2}$$

Note that the inner expectation averages out the randomness caused by $X_2, X_3, \ldots, X_k$; therefore, (A.2) measures the portion of uncertainty in Y that is solely attributed to X_1.

In the context of VCO, we estimated the first-order effects of three sources of uncertainty conditional on $\boldsymbol{\theta}$: $\mathscr{B}^s$ sampling, product library sampling (π), and candidate vehicle set sampling (γ):

$$F_{\mathscr{B}} = \mathrm{V}(\mathrm{E}[\bar{Y}(\mathscr{B}^s, \boldsymbol{\theta}, \pi, \gamma)|\mathscr{B}^s]),$$
$$F_{\pi} = \mathrm{V}(\mathrm{E}[\bar{Y}(\mathscr{B}^s, \boldsymbol{\theta}, \pi, \gamma)|\pi]),$$
$$F_{\gamma} = \mathrm{V}(\mathrm{E}[\bar{Y}(\mathscr{B}^s, \boldsymbol{\theta}, \pi, \gamma)|\gamma]).$$

We adopted an efficient computation method by Saltelli (2002) that estimates all inputs' first-order effects at the same time using $(k+2)N$ simulation runs, where k represents the number of inputs (e.g., $k = 3$, as in our case), and $N \geq 2$. First, we sample $2N$ sets of $\{\mathscr{B}^s, \pi, \gamma\}$ and form the following two $N \times 3$ matrices:

$$A = \begin{pmatrix} \mathscr{B}_1^s, & \pi_1 & \gamma_1 \\ \mathscr{B}_2^s, & \pi_2 & \gamma_2 \\ \vdots, & \vdots & \vdots \\ \mathscr{B}_N^s, & \pi_N & \gamma_N \end{pmatrix}, B = \begin{pmatrix} \mathscr{B}_{N+1}^s, & \pi_{N+1} & \gamma_{N+1} \\ \mathscr{B}_{N+2}^s, & \pi_{N+2} & \gamma_{N+2} \\ \vdots, & \vdots & \vdots \\ \mathscr{B}_{2N}^s, & \pi_{2N} & \gamma_{2N} \end{pmatrix}.$$

Combining the columns of A and B, we define the following three additional matrices:

$$C_{\mathscr{B}} = \begin{pmatrix} \mathscr{B}_1^s, & \pi_{N+1} & \gamma_{N+1} \\ \mathscr{B}_2^s, & \pi_{N+2} & \gamma_{N+2} \\ \vdots, & \vdots & \vdots \\ \mathscr{B}_N^s, & \pi_{2N} & \gamma_{2N} \end{pmatrix}, C_{\pi} = \begin{pmatrix} \mathscr{B}_{N+1}^s, & \pi_1 & \gamma_{N+1} \\ \mathscr{B}_{N+2}^s, & \pi_2 & \gamma_{N+2} \\ \vdots, & \vdots & \vdots \\ \mathscr{B}_{2N}^s, & \pi_N & \gamma_{2N} \end{pmatrix},$$
$$C_{\gamma} = \begin{pmatrix} \mathscr{B}_{N+1}^s, & \pi_{N+1} & \gamma_1 \\ \mathscr{B}_{N+2}^s, & \pi_{N+2} & \gamma_2 \\ \vdots, & \vdots & \vdots \\ \mathscr{B}_{2N}^s, & \pi_{2N} & \gamma_N \end{pmatrix}.$$

For each row of these five matrices, we perform one simulation run given $\boldsymbol{\theta}$, which results in $5N$ simulation runs in total. For instance, if we choose $N = 10$, then the overall simulation budget would be 50 simulation runs. Let $\bar{Y}^{M,i}$ represent the expected sales volume estimate from the ith row's inputs in matrix M. Then we can compute the estimates of three first-order effects as follows:

$$\hat{F}_{\mathscr{B}} = \sum_{i=1}^{N} \bar{Y}^{A,i}\bar{Y}^{C_{\mathscr{B}},i}/N - \left(\sum_{i=1}^{N} (\bar{Y}^{A,i} + \bar{Y}^{B,i})/2N\right)^2,$$
$$\hat{F}_{\pi} = \sum_{i=1}^{N} \bar{Y}^{A,i}\bar{Y}^{C_{\pi},i}/N - \left(\sum_{i=1}^{N} (\bar{Y}^{A,i} + \bar{Y}^{B,i})/2N\right)^2,$$
$$\hat{F}_{\gamma} = \sum_{i=1}^{N} \bar{Y}^{A,i}\bar{Y}^{C_{\gamma},i}/N - \left(\sum_{i=1}^{N} (\bar{Y}^{A,i} + \bar{Y}^{B,i})/2N\right)^2.$$

By pooling the simulation results from A and B, we can also compute the mean sales volume estimate, $\bar{\bar{Y}} = \sum_{i=1}^{2N} \bar{Y}^i/(2N)$, where $\bar{Y}^1, \bar{Y}^2, \ldots, \bar{Y}^{2N}$ are the sales volumes computed from each of the $2N$ rows of A and B. The variance of $\bar{\bar{Y}}$ can be also estimated as

$$\hat{\mathrm{V}}(\bar{\bar{Y}}) = \hat{\mathrm{V}}(\bar{Y}(\boldsymbol{\theta}, \mathscr{B}^s, \pi, \gamma))/2N = \sum_{i=1}^{2N} (\bar{Y}^i - \bar{\bar{Y}})^2/(2N(2N-1)).$$

A.4. Regression Model for Bias

The expected sales volume conditional on all other inputs but π can be approximated by a quadratic function of $1/p$:

$$\mathrm{E}[\bar{Y}(\boldsymbol{\theta}, \mathscr{B}^s, \pi(p), \gamma)|\boldsymbol{\theta}, \mathscr{B}^s, \gamma] \approx c_0 + c_1/p + c_2/p^2, \tag{A.3}$$

where c_0, c_1, and c_2 are unknown quantities depending on $\boldsymbol{\theta}$, $\mathscr{B}^s$, and γ, and $c_1 < 0$; the latter condition is postulated from the expectation that the (negative) bias becomes closer to 0 for large p. Notice that we explicitly indicate the size of the product library as $\pi(p)$ in (A.3). We use $\pi(P)$ to represent the set of all possible combinations of features prescribed by the content portfolio, where P is the size of the set. For a typical content portfolio, $P \gg p$. Under model (A.3), the bias as a result of sampling $\pi(p)$ conditional on other inputs can be written as

$$\mathrm{E}[\bar{Y}(\boldsymbol{\theta}, \mathscr{B}^s, \pi(p), \gamma)|\boldsymbol{\theta}, \mathscr{B}^s, \gamma] - \mathrm{E}[\bar{Y}(\boldsymbol{\theta}, \mathscr{B}^s, \pi(P), \gamma)|\boldsymbol{\theta}, \mathscr{B}^s, \gamma]$$
$$= c_1(1/p - 1/P) + c_2(1/p^2 - 1/P^2).$$

In other words, the bias-corrected estimator of the sales volume when $p < P$ is $\bar{Y}(\boldsymbol{\theta}, \mathscr{B}^s, \pi(p), \gamma) - c_1(1/p - 1/P) - c_2(1/p^2 - 1/P^2)$.

To estimate the constants c_0, c_1, and c_2 in model (A.3), we pick $k \geq 3$ different product library sizes $p_1 < p_2 < \cdots < p_k$ and sample a product library of each size: $\pi(p_1), \pi(p_2), \cdots, \pi(p_k)$. After running simulations with these product libraries while fixing other inputs, $\bar{Y}(\boldsymbol{\theta},\ \mathscr{B}^s, \pi(p_1), \gamma)$, $\bar{Y}(\boldsymbol{\theta}\mathscr{B}^s, \pi(p_2), \gamma), \ldots,$ $\bar{Y}(\boldsymbol{\theta},\ \mathscr{B}^s, \pi(p_k), \gamma)$ can be used to fit model (A.3) to estimate the constants via least squares regression.

Endnotes

[1] Recall that $\boldsymbol{\theta}$ is a vector of tuning parameters.

[2] We do not disclose what θ_1–θ_5 are owing to GM proprietary information.

References

Andradóttir S (2015) A review of random search methods. Fu MC, ed. *Handbook of Simulation Optimization* (Springer, New York), 277–292.

General Motors (2018) GM gains half a point of market share as Chevrolet and GMC second quarter deliveries soar. Press release, General Motors, Detroit, MI. Accessed June 22, 2020, https://media.gm.com/media/us/en/gm/news.detail.html/content/Pages/news/us/en/2018/jul/0702-gmsales.html.

Homma T, Saltelli A (1996) Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engrg. Systems Safety* 52(1):1–17.

Iooss B, Lemaître P (2015) A review on global sensitivity analysis methods. Dellino G, Meloni C, eds. *Uncertainty Management in Simulation-Optimization of Complex Systems* (Springe``r, Boston), 101–122.

Ligmann-Zielinska A, Kramer DB, Spence Cheruvelil K, Soranno PA (2014) Using uncertainty and sensitivity analyses in socioecological agent-based models to improve their analytical performance and policy relevance. *PLoS One* 9(10):e109779.

Louviere JJ, Hensher DA, Swait JD (2000) *Stated Choice Methods: Analysis and Application* (Cambridge University Press, New York).

Orme BK (2006) *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research* (Research Publishers, Madison, WI).

Pianosi F, Beven K, Freer J, Hall JW, Rougier J, Stephenson DB, Wagener T (2016) Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environ. Model. Software* 79(May):214–232.

Saltelli A (2002) Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Comm.* 145(2):280–297.

Saltelli A, Tarantola S (2002) On the relative importance of input factors in mathematical models. *J. Amer. Statist. Assoc.* 97(459): 702–709.

Saltelli A, Ratto M, Tarantola S, Campolongo F (2005) Sensitivity analysis for chemical models. *Chemical Rev.* 105(7):2811–2828.

Sobol' IM (1993) Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Experiment.* 1(4):407–414.

Song E, Nelson BL, Staum J (2016) Shapley effects for global sensitivity analysis: Theory and computation. *SIAM/ASA J. Uncertainty Quantification* 4(1):1060–1083.

Train K (2003) *Discrete Choice Methods with Simulation* (Cambridge University Press, New York).

Wu-Smith P, Beaujon G, Donndelinger J, Owen J, Costy T (2014) Customer choice analysis tool and process for vehicle content and pricing optimization. Technical report, General Motors, Warren, MI.

Verification Letter

Jonathan H. Owen, PhD, CAP®, Director of Operations Research, Chief Scientist for Operations Research and Advanced Analytics, General Motors Global Research & Development, General Motors Company, Warren, MI 48092-2031, writes:

"I am pleased to verify our use of the method and analysis capability presented in 'Uncertainty Quantification in Vehicle Content Optimization for General Motors' by Song, Wu-Smith, and Nelson.

"Vehicle Content Optimization (VCO) supports General Motors' vehicle content and packaging decisions as an important part of the vehicle development process; however, no uncertainty was explicitly considered. The uncertainty quantification and reduction capability described in the paper allows us to conduct vehicle content optimization analyses that are more robust to model uncertainty by systematically quantifying uncertainty in the performance measure estimates and diagnosing which sources have the highest influence on the overall uncertainty. In addition, the project identified several opportunities to improve the efficiency of VCO by reducing its model redundancy and computational overhead."

Eunhye Song is the Harold and Inge Marcus Early Career Assistant Professor in the Department of Industrial and Manufacturing Engineering at the Pennsylvania State University. Her research expertise spans simulation design of experiments, uncertainty and risk quantification, and optimization via simulation. She is particularly interested in finding a robust solution in the presence of simulation model risk. Her website can be found at http://eunhyesong.info.

Peiling Wu-Smith is a group manager, program lead, and GM Technical Fellow in the Advanced Analytics COE of General Motors Chief Data and Analytics Office, focusing on applying advanced analytics to GM's enterprise-wide business. She received her PhD in industrial engineering and operations research at Lehigh University and is recognized as a member of the GM team awarded the INFORMS Prize for its "sustained track record of innovative and impactful applied operations research and advanced analytics."

Barry L. Nelson is the Walter P. Murphy Professor of the Department of Industrial Engineering and Management Sciences at Northwestern University and a distinguished visiting scholar at Lancaster University in England. His research focus is on the design and analysis of computer simulation experiments. He has published numerous papers and three books, including *Foundations and Methods of Stochastic Simulation: A First Course* (Springer, 2013). He is a fellow of INFORMS and IISE.