CASE STUDY / AUTOMATED VEHICLE ORDERING

The Order Mix the Factory Can Actually Build

Linking demand prediction to constrained, uncertainty-aware decisions about vehicle configurations

EXECUTIVE CONCLUSION

Predicting desirable vehicles is only half the problem. Manufacturing a profitable combination requires portfolio-level optimization.

$64M
incremental profit, project-reported

3 mo.
order horizon

ML + MILP
demand model feeding an order optimizer

Founder-led prior industry project; not delivered by Bit Bros LLC. The $64M amount is project-reported rather than independently audited.

The problem: popular is not always buildable

An automotive manufacturer must choose a feasible collection of models, trims, options, and colors before customer demand is fully known. A prediction model can estimate which configurations buyers will prefer. But independent rankings are not a production plan: multiple highly ranked vehicles may consume the same limited components, violate factory restrictions, or produce the wrong regional mix.

The project combined tree-based demand modeling and mathematical optimization to formulate a roughly three-month order recommendation. Demand estimates informed the economics of candidate configurations, while the optimizer selected a buildable portfolio under manufacturing and supply constraints. The project lead reported approximately $64 million in incremental profit.

THE DECISION TO IMPROVE

Which combinations of vehicle configurations should be ordered over the next three months, given predicted demand, component constraints, plant restrictions, and financial priorities?

1. Convert demand signals into economic choices

A forecast ranking does not resolve shared scarcity

Suppose configuration A earns $1,100 of expected contribution but consumes two units of a scarce option. Configuration B earns $800 and consumes one unit. This month 300 option units are available, and the demand model predicts 120 buyers for A and 200 for B. These figures are synthetic illustrations, not measurements from the project.

	Rank per vehicle	Optimized order
Vehicles of A	120	50
Vehicles of B	60	200
Option units used	300	300
Contribution	$180K	$215K

Ranking by contribution per vehicle orders all the A that demand supports. The optimized order favors B, which earns $800 per option unit against A's $550, and earns 19.4% more from the same supply while also building more vehicles.

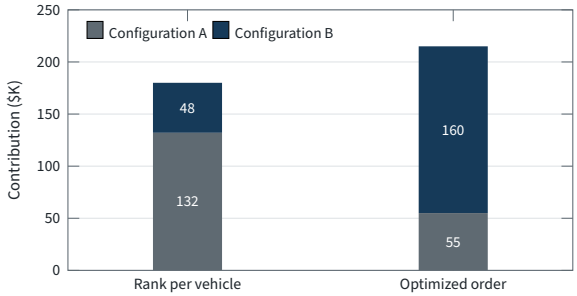

Exhibit 1. Contribution by configuration under each order, from the same 300 option units: $180K against $215K, valued at point forecasts. Synthetic.

With one scarce option, ranking by contribution per option unit would find this answer. Real orders share many options, plant limits, and mix rules at once, so no single ranking holds.

The optimizer must compare whole portfolios. A choice that looks good for one vehicle may eliminate several attractive alternatives elsewhere in the order. This is exactly the type of coupling for which mixed-integer optimization is designed.

Represent decisions explicitly

Symbol	Meaning
$i \in \mathcal{I}$	buildable configuration (model, trim, options, color)
$t \in \mathcal{T}$	order period within the horizon
$k \in \mathcal{K}$	option, component or supplier resource
x_{it}, z_{it}	vehicles ordered (integer); configuration scheduled (binary)
a_{ki}	units of resource k per vehicle of i
C_{kt}, K_{pt}	resource supply; capacity of plant or line p
D_{it}, F_{it}	uncertain demand and its distribution function
m_i, o_i	margin per vehicle sold; cost per vehicle left unsold

A representative ordering model is

$$\max \ \textstyle\sum_{i,t} \pi_{it}(x_{it}) \tag{1}$$

$$\text{s.t.} \ \textstyle\sum_i a_{ki}\, x_{it} \le C_{kt} \qquad \forall k,t \tag{2}$$

$$\textstyle\sum_{i \in \mathcal{P}_p} x_{it} \le K_{pt} \qquad \forall p,t \tag{3}$$

$$L_{gt} \le \textstyle\sum_{i \in \mathcal{G}_g} x_{it} \le U_{gt} \qquad \forall g,t \tag{4}$$

$$\ell_i z_{it} \le x_{it} \le M_i z_{it} \qquad \forall i,t \tag{5}$$

$$x_{it} \in \mathbb{Z}_+, \ z_{it} \in \{0,1\}, \tag{6}$$

where (2) shares scarce options and components, (3) caps plants and lines, (4) encodes approved product and regional mix rules over groups $\mathcal{G}_g$, and (5) enforces minimum build lots. $\pi_{it}(\cdot)$ is the expected contribution of ordering a given quantity, derived in Section 2. Compatibility is enforced by enumerating only buildable configurations. Integrality matters because orders are placed in whole vehicles and minimum lots make small allocations infeasible, not merely suboptimal. This is a teaching formulation; the original proprietary model contained additional operational detail.

When the catalog is too large to enumerate

Options multiply: a few dozen option families can generate millions of buildable configurations. Column generation keeps only the configurations that matter. The restricted master problem is the linear relaxation of (1)–(4) over a working set $\mathcal{I}' \subset \mathcal{I}$, with duals $\mu_{kt} \ge 0$ on (2) and $\nu_{pt} \ge 0$ on (3). A new configuration is worth adding only if its reduced profit is positive,

$$\bar{v}_{jt} = v_{jt} - \textstyle\sum_k \mu_{kt}\, a_{kj} - \nu_{p(j)t} > 0, \tag{7}$$

where v_{jt} is its marginal value at zero volume. The pricing problem, $\max_j \bar{v}_{jt}$ over the compatibility rules, is a small integer program over option choices that never enumerates the catalog. Columns are added until no configuration prices out, and integer orders are recovered by branch-and-price or, in practice, by solving the final MILP over the generated columns.

2. Turn predictions into economics carefully

A linear objective assumes every ordered vehicle sells

Writing the objective as $v_{it}x_{it}$, a predicted contribution times quantity, treats every additional vehicle as equally valuable. It is not: the hundredth unit of a configuration sells only if demand reaches a hundred. With margin m_i per vehicle sold, cost o_i per vehicle left unsold at the end of the horizon (carrying, aging, incentives to clear), and uncertain demand D_i,

$$\pi_i(x) = m_i\, \mathbb{E}[\min(x, D_i)] - o_i\, \mathbb{E}[(x - D_i)^+],$$

The marginal value of the $(x{+}1)$-th vehicle follows from $\min(x{+}1, D) - \min(x, D) = \mathbf{1}\{D > x\}$:

$$\Delta\pi_i(x) = m_i \Pr\{D_i > x\} - o_i \Pr\{D_i \le x\} = m_i - (m_i + o_i) F_i(x). \tag{8}$$

Because F_i is nondecreasing, $\Delta\pi_i(x)$ is nonincreasing, so π_i is concave. Alone, configuration i would be ordered up to the newsvendor quantity, the smallest x with $F_i(x) \geq m_i/(m_i + o_i)$.

Concavity is what keeps the model linear. Split $x_{it} = \sum_s y_{its}$ into segments $0 \leq y_{its} \leq w_s$ with slopes σ_{is} taken from (8), and maximize $\sum_s \sigma_{is} y_{its}$. Since $\sigma_{i1} \geq \sigma_{i2} \geq \cdots$, an optimal solution fills the segments in order without any binary variables forcing it to; a non-concave curve, such as a volume discount, would need the extra binaries. The economics become correct while the MILP stays in the same size class.

The point-forecast plan overstates its own value

Exhibit 2 revisits the example with uncertain demand (negative binomial, coefficient of variation 0.35) and unsold-vehicle costs of $500 for A and $350 for B, all synthetic. The point-forecast optimum still claims $215K, but its expected contribution is about $181K, because 200 units of B equals B's mean demand and many of those vehicles go unsold. The order that maximizes expected contribution, 72 of A and 156 of B, uses the same 300 option units and expects about $190K, 4.6% more. Ranking per vehicle expects only $153K.

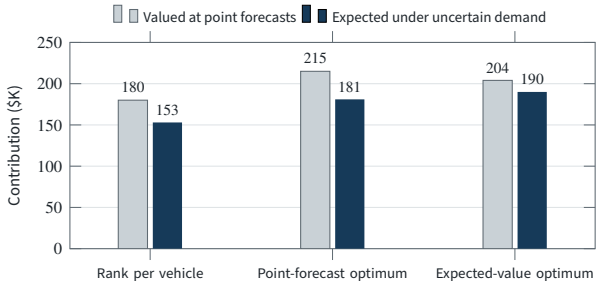

Exhibit 2. The same three orders valued at point forecasts and in expectation under uncertain demand. Synthetic; computed in exhibit/uncertain_order.py.

The shared option has a price, and the configurations solve themselves

Relaxing the shared constraint (2) with a price $\lambda \geq 0$ per option unit decomposes the problem by configuration:

$$L(\lambda) = \lambda C + \textstyle\sum_i \max_{x_i \geq 0} \left[\pi_i(x_i) - \lambda a_i x_i\right]. \tag{9}$$

By (8), each subproblem is a newsvendor with a charge for the option: order the smallest x_i with

$$F_i(x_i) \;\geq\; \frac{m_i - \lambda a_i}{m_i + o_i}. \tag{10}$$

The dual $\min_{\lambda \geq 0} L(\lambda)$ is convex, and $g(\lambda) = \sum_i a_i x_i(\lambda) - C$ is a subgradient, which gives the price update in Algorithm 1. In the Exhibit 2 example the price converges to about $463 per option unit, below the $550 the point-forecast linear program implies, because uncertainty lowers the expected value of the marginal vehicle. Integrality leaves a small duality gap: the subproblems alternate between orders using 299 and 301 units, and a repair that searches feasible orders within one vehicle of the subproblem solution recovers 72 of A and 156 of B, the exhaustive optimum. With hundreds of configurations sharing many options, the same scheme scales linearly and produces the shadow prices supplier negotiations need.

ALGORITHM 1. PRICING A SHARED OPTION BY SUBGRADIENT

$\lambda_0 \leftarrow 0$, step scale $s_0 > 0$
for $n = 1, 2, \ldots, N$ **do**
  for each configuration i **do**
    $x_i \leftarrow \min\{x : F_i(x) \geq (m_i - \lambda_{n-1} a_i)/(m_i + o_i)\}$
  end for
  $g \leftarrow \sum_i a_i x_i - C$ ▷ excess option use
  $\lambda_n \leftarrow \max\{0,\ \lambda_{n-1} + (s_0/n)\, g\}$
end for
Repair: best feasible order within one vehicle of x per configuration
return order x and option price λ_N

Selection on noisy predictions inflates expected value

An optimizer favors configurations whose predicted value is high, and some are high partly because their prediction errors are positive. Smith and Winkler call this the optimizer's curse. In a synthetic illustration where 10 of 50 configurations are chosen on predictions with noise larger than the true spread in value, the selected set's predicted value exceeds its true value by 16% on average. Formally, if $\hat{v}_i = v_i + \varepsilon_i$ with unbiased errors, then $\mathbb{E}[\max_i \hat{v}_i] \geq \max_i \mathbb{E}[\hat{v}_i]$ by Jensen's inequality, since the maximum is convex, so the value of the chosen option is biased upward even when every estimate is unbiased. The Bayesian remedy of Smith and Winkler shrinks each estimate toward the prior mean μ_0,

$$\tilde{v}_i = \mu_0 + \kappa\,(\hat{v}_i - \mu_0), \qquad \kappa = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\varepsilon^2}, \tag{11}$$

where σ_0^2 is the spread of true values and σ_ε^2 the prediction noise. In the illustration $\kappa = 0.31$, and selecting on $\tilde{v}$ cuts the overstatement from 16% to 0.1%. Shrinkage is computed per segment in practice, and the pipeline should be judged on holdout decisions rather than holdout predictions.

Predictive distributions, calibrated

A tree-based demand model typically predicts a central estimate per configuration, region and period. The optimizer needs a distribution. Quantile regression forests, distributional boosting or conformal intervals provide one, and calibration is checked directly: realized demand should fall below the predicted 80th percentile about 80% of the time. Where errors near binding constraints dominate the value, decision-focused training helps. For a decision $w^*(c)$ that minimizes cost $c^\top w$ over a feasible set S with optimal value $z^*(c)$, the SPO+ loss of Elmachtoub and Grigas for a predicted cost vector $\hat{c}$ is

$$\ell_{\text{SPO+}}(\hat{c}, c) = \max_{w \in S}\{c^\top w - 2\hat{c}^\top w\} + 2\hat{c}^\top w^*(c) - z^*(c), \tag{12}$$

a convex surrogate for the regret of acting on $\hat{c}$ when the truth is c. It trains the model on the cost of the decisions it induces rather than on symmetric accuracy, and is worth its added complexity when a few binding constraints determine most of the value.

3. Make the constraint system legible

Decision input	What it contributes
Predicted demand	Distribution of sales for each configuration, region and period
Configuration catalog	Compatible models, trims, colors, and options
Factory limits	Feasible production amounts and build locations
Supplier capacity	Shared scarcity across competing orders
Business policy	Approved regional, product, and mix requirements
Order cadence	How decisions are revised over three months

Inherited rules require scrutiny. A factory limit is physical; a preferred share for a color or trim may represent a commercial objective. Translating every preference into a hard constraint can overconstrain the order and sacrifice economic value. Preferences are better written as elastic constraints, with a penalized deviation variable whose price the rule's owner approves. When the model is infeasible, an irreducible infeasible subsystem names the smallest set of conflicting rules, which distinguishes infeasibility from unfavorable economics.

Price overrides and scarce options explicitly

Two quantities make trade-offs visible. The opportunity cost of a planner override is the drop in objective when the override is fixed and the model is re-solved. The marginal value of one more unit of a scarce option is its shadow price; in a MILP it is estimated by fixing the integer decisions and solving the remaining linear program, or more robustly by re-solving with the capacity changed. In Exhibit 1, B is already ordered to its demand, so one more option unit is worth half a vehicle of A, $550 at the margin, which is the number a supplier negotiation needs.

REFERENCE WORKFLOW

Demand distributions → economics and catalog → plant and supplier limits → portfolio MILP → planner review and approved orders

Re-plan on a rolling horizon without thrashing

A three-month order is revised as information arrives. Each cycle re-solves the horizon, fixes orders inside a frozen fence, and adds a penalty $\lambda \sum_{i,t} |x_{it} - x_{it}^{\text{prev}}|$, linearized with auxiliary variables, so that small forecast revisions do not produce large swings in the order. The penalty weight is a business choice about the value of stability, and it should be set explicitly rather than left to chance.

4. Evaluate the order, not only the prediction

Forecast error matters, but the objective is a good set of orders. Two forecasts can have similar accuracy and produce different decisions near a supplier bottleneck. Conversely, a substantial accuracy improvement may have little value when it leaves the preferred feasible portfolio unchanged.

Counterfactual evaluation is harder than it looks: sales of configurations that were never ordered are unobserved, and sales of those that sold out are censored.

Credible comparisons replay each policy against the demand model's distributions over holdout periods, priced by the same costs, or use regional tests when the operation allows them. A two-by-two design, old or new forecast crossed with ranking or optimizer, separates the value of better prediction from the value of optimization.

Useful operating measures include expected and realized contribution, recommendation acceptance, allocation of scarce components, calibration of the demand distributions, and the stability of the order mix under forecast revisions.

The project lead reported approximately $64 million in incremental profit associated with the earlier automated ordering work. This case study cannot reproduce a full financial reconciliation or isolate the respective contributions of forecast accuracy, optimization, and process redesign. The project figure is consequently presented as reported impact, not a new audited Bit Bros claim.

5. Practical implementation lessons

Start with a narrow slice of the catalog and the resources that actually bind, against a simple forecast-ranking baseline, and add product interactions, regional constraints and governance in stages. Test the model on known cases before building dashboards, validate both the calibration of the predictive distributions and the economic quality of the resulting orders, and record every override with its reason.

TRANSFERABLE LESSON

When several individually attractive recommendations compete for the same resources, the business needs an optimization of the complete decision portfolio, valued under the uncertainty it actually faces.

6. What we would add to a modern version

The same structure supports a stronger decision system. Each addition should be tested against the current order policy rather than assumed to help.

Addition	Question it answers
Scenario-based ordering	Does a sample-average two-stage model, ordering now and allocating after demand is seen, beat the deterministic order?
Order simulation	How do lost sales, aging stock, and late changes play out over the horizon?
Override learning	Which planner changes improve on the model, and which cost money?
Value tracking	What did each order earn against a credible baseline?

These are extensions, not claims about the original project.

Further reading

Elmachtoub, A. N. and Grigas, P. Smart "predict, then optimize." *Management Science* 68(1), 2022.

Smith, J. E. and Winkler, R. L. The optimizer's curse: skepticism and postdecision surprise in decision analysis. *Management Science* 52(3), 2006.

Meinshausen, N. Quantile regression forests. *Journal of Machine Learning Research* 7, 2006.

Shapiro, A., Dentcheva, D. and Ruszczyński, A. *Lectures on Stochastic Programming*, 2nd ed. SIAM, 2014.

Desaulniers, G., Desrosiers, J. and Solomon, M. M. (eds.) *Column Generation*. Springer, 2005.

Fisher, M. L. The Lagrangian relaxation method for solving integer programming problems. *Management Science* 27(1), 1981.

Williams, H. P. *Model Building in Mathematical Programming*, 5th ed. Wiley, 2013.

DeJans Jr., A. *The MILP Optimization Handbook*. Bit Bros, 2025.

About the case and Bit Bros

This paper draws on an automotive ordering project led by **Adam DeJans Jr.** in an earlier industry role. It was not a Bit Bros client engagement. The manufacturer, proprietary constraint data, and financial methodology are not disclosed. The reported profit figure is not represented as an audited realized outcome. Bit Bros LLC helps organizations connect predictive modeling, mathematical optimization, and operating decisions.

Adam DeJans Jr.
Technical Managing Director, Bit Bros LLC
Decision scientist specializing in optimization, simulation, and decisions under uncertainty.

Design the decision. Make the trade-offs explicit.