AUTOMOTIVE CASE STUDY / PRODUCTION ALLOCATION

The Vehicle Mix Is an Economic Decision

How mathematical optimization helped an automotive planning team allocate scarce capacity across North America

EXECUTIVE CONCLUSION

A better production mix can unlock material value without requiring the planner to predict every future outcome perfectly.

Vehicle-production planning is a portfolio decision under binding resource constraints. Once those constraints are explicit, an optimizer can compare feasible mixes across products and periods and show where scarce capacity creates the most value. This note describes automotive work led by Adam DeJans Jr. in an earlier industry role: the formulation, how scarcity is priced, how competing objectives, plan stability and infeasibility are handled, and how such a plan should be validated.

40,000
additional vehicles/year

$390M
realized profit, attributed to the work

24 mo.
planning horizon

200
vehicle types modeled (approx.)

Figures as reported by the project. The $390M was reported as realized incremental profit attributed to this work; Bit Bros has not independently audited it. Work was performed in an earlier industry role, not as a Bit Bros client engagement.

The planning problem

A North American automotive manufacturer must continually determine which vehicle types to produce, where, and when. Every feasible plan consumes assembly capacity, shared parts, labor, and supplier resources. Customer demand differs across products and periods, as do economics and environmental consequences. A vehicle mix that looks attractive in isolation may block a more valuable combination elsewhere in the network.

The project addressed a long-horizon supply-allocation problem spanning approximately 24 months and around 200 vehicle types, translating forecasts and operating constraints into a consistent production mix while accounting for profit, volume, and greenhouse-gas considerations.

THE DECISION

How should the business allocate limited production resources across vehicle types, plants and future periods so that the resulting mix is feasible, economically attractive, and aligned with enterprise priorities?

1. Price the scarcity before allocating it

Capacity is valuable because of what it could produce instead

An unconstrained plan can chase expected demand. A constrained plan must rank competing uses of scarce resources, and the economically relevant comparison is the contribution of one use against the contribution forgone elsewhere. Neither a demand ranking nor a unit-margin ranking identifies the best feasible mix.

A two-product illustration

Suppose a plant has 1,000 hours of a bottleneck operation per month. Product A contributes $3,000 per unit and needs 2 bottleneck hours; product B contributes $2,000 and needs 1 hour. Monthly demand is 400 units of A and 700 of B. The numbers are invented to show the mechanism.

	Rank by unit margin	Optimized mix
Units of A	400	150
Units of B	200	700
Bottleneck hours used	1,000	1,000
Contribution	$1.60M	$1.85M

Ranking by unit margin builds all the A that demand allows and fills the remaining hours with B. The optimized mix does the opposite, because B earns $2,000 per bottleneck hour against A's $1,500. The same capacity produces 15.6% more contribution, and more vehicles.

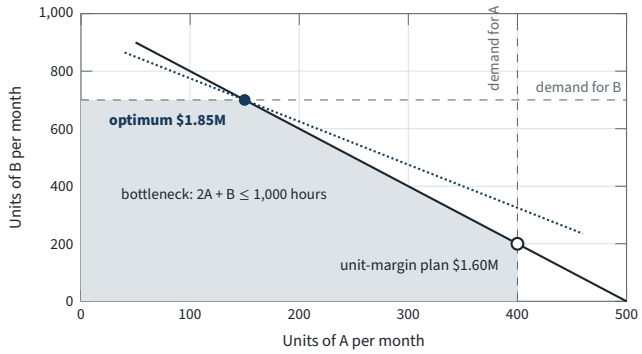

Exhibit 1. Feasible mixes (shaded) for the two-product illustration. The dotted line is every mix worth $1.85M; no feasible mix lies above it.

Shadow prices put a number on scarcity

The linear-programming dual of the bottleneck constraint states what one more hour is worth. At the optimum above it is $1,500: B already meets its demand, so a marginal hour builds half a unit of A. The dual on B's demand bound is $500, the premium B earns per hour over A. Exhibit 2 shows the value of an hour as capacity changes. It is $2,000 while capacity is short of B's needs, $1,500 while A absorbs it, and zero once both demands are met. These prices answer the questions planners are actually asked: what an overtime shift is worth, and what a requested exception costs the rest of the plan.

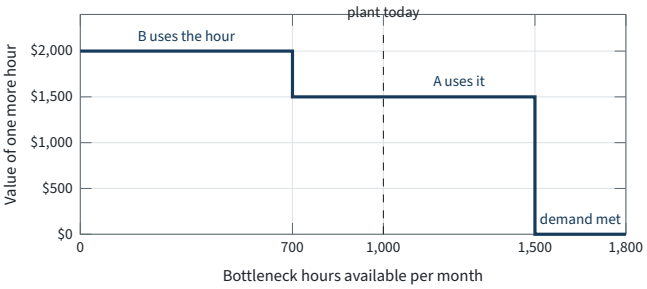

Exhibit 2. Shadow price of bottleneck capacity in the two-product illustration, from parametric linear programming. Synthetic data; computed by the script in the note's exhibit folder.

Two cautions apply in practice. Duals of a mixed-integer program are not defined directly; they are read from the linear relaxation with integer decisions fixed, or, more robustly, by re-solving with the resource changed. And a dual is valid only over the range in which the binding constraints stay the same, which is why Exhibit 2 is a staircase, not a single number.

The duality behind the price

For the linear core $\max\{m^\top x : Ax \le b,\ 0 \le x \le u\}$, the dual is

$$\min_{y,\mu \ge 0} \ b^\top y + u^\top \mu \quad \text{s.t.} \quad A^\top y + \mu \ge m. \tag{1}$$

Strong duality makes the two optimal values equal, and complementary slackness ties them together: $y_k\,(b_k - a_k^\top x) = 0$ for every resource, and $x_j\,(a_j^\top y + \mu_j -$

$m_j) = 0$ for every product. A resource with slack has a zero price, and every product actually built earns exactly what its resources are worth, so its reduced cost $m_j - a_j^\top y - \mu_j$ is zero.

In the illustration the dual is min $1000y + 400\mu_A + 700\mu_B$ subject to $2y + \mu_A \geq 3000$ and $y + \mu_B \geq 2000$. Its solution, $y = 1{,}500$, $\mu_A = 0$, $\mu_B = 500$, has value \$1.85M, equal to the primal. A's demand bound has slack, so $\mu_A = 0$; both products are built, so both dual constraints hold with equality. The price of an hour is therefore read off A, the product at the margin.

The optimal value as a function of capacity, $V(C)$, is piecewise linear and concave, as the value function of any linear program is in its right-hand side, and the dual price is its slope:

$$V(C) = \begin{cases} 2{,}000\,C, & C \leq 700, \\ 1.4\text{M} + 1{,}500\,(C - 700), & 700 \leq C \leq 1{,}500, \\ 2.6\text{M}, & C \geq 1{,}500. \end{cases}$$

Concavity is the formal statement of diminishing returns to capacity: each additional hour is worth no more than the last, which is what makes marginal capacity decisions safe to take one step at a time.

2. Write the network down as one model

Notation

Symbol	Meaning
i, p, r, t, k	Vehicle type, plant, market, period, shared resource
$\mathcal{E}$	Eligible (type, plant) pairs: what can physically be built where
m_{irt}, h_i	Contribution per unit sold; holding cost per unit-period
a_{ikp}, C_{kpt}	Resource k used per unit of i at p; capacity available
L_{irt}, U_{irt}	Lower and upper allocation bounds (demand, mix rules)
e_i, $\bar{E}_t$	Emissions per unit; fleet-average target
ℓ_{ip}, $\bar{C}_{ip}$	Minimum run quantity; tight upper bound on production
x_{ipt}, y_{irt}, I_{it}	Production, allocation to markets, end inventory (≥ 0)
$z_{ipt} \in \{0, 1\}$	Whether type i runs at plant p in period t

The allocation program

$$\max \ \textstyle\sum_{i,r,t} m_{irt}\, y_{irt} - \sum_{i,t} h_i\, I_{it} \tag{2}$$

$$\text{s.t.} \ \textstyle\sum_{i:(i,p)\in\mathcal{E}} a_{ikp}\, x_{ipt} \leq C_{kpt} \qquad \forall k, p, t \tag{3}$$

$$I_{it} = I_{i,t-1} + \textstyle\sum_p x_{ipt} - \sum_r y_{irt} \qquad \forall i, t \tag{4}$$

$$L_{irt} \leq y_{irt} \leq U_{irt} \qquad \forall i, r, t \tag{5}$$

$$\textstyle\sum_{i,r} (e_i - \bar{E}_t)\, y_{irt} \leq 0 \qquad \forall t \tag{6}$$

$$\ell_{ip}\, z_{ipt} \leq x_{ipt} \leq \bar{C}_{ip}\, z_{ipt} \qquad \forall (i,p) \in \mathcal{E},\ t \tag{7}$$

Variables x_{ipt} exist only for $(i, p) \in \mathcal{E}$, so eligibility is enforced by omission, which keeps the model smaller than any constraint could. The bound $\bar{C}_{ip} = \min_k C_{kpt}/a_{ikp}$ in (7) is the tightest valid one, which matters for the strength of the linear relaxation; a generic large constant would weaken it. Integer variables appear only where the business is genuinely discrete, such as minimum runs, batch sizes or low-volume types. This is a schematic, not the proprietary production formulation.

Ratio targets become linear constraints

A fleet-average target, $\sum e_i y_i / \sum y_i \leq \bar{E}$, looks nonlinear, but multiplying through by the positive volume $\sum y_i$ gives the linear constraint (6), so it fits inside the same program. Its dual prices the target. Adding a 125 g/km average to the two-product illustration, with A at 180 and B at 120, cuts A from 150 units to 63 and costs \$261,000 of monthly contribution: the price of the target, stated in the unit leadership manages.

Constraint family	Representation	Hard or soft
Plant capacity	Build limits by facility and period	Hard
Shared components	Parts consumed jointly across types	Hard
Eligibility	Variables exist only where buildable	Hard
Demand and mix	Bounds and share rules by market	Soft, priced
Multi-period balance	Inventory links periods	Hard
Environmental limits	Fleet-average and product impacts	Policy

The project's governing rule was that constraints which were physics stayed hard, while preferences became explicit trade-offs. That classification is itself a modeling decision with owners.

3. Objectives, stability and infeasibility

Competing objectives need an explicit method

Profit, volume, demand coverage and emissions conflict. A weighted sum is quick but requires commensurate units and hides the trade-off inside the weights. Two alternatives make the trade-off governable. A *lexicographic* method optimizes the first priority, fixes it within an approved tolerance, then optimizes the next. An *ε-constraint* method optimizes one objective while bounding the others. Formally, with objectives $f_1, \ldots, f_q$ to maximize over the feasible set X,

$$P(\varepsilon): \quad \max_{x \in X} f_1(x) \quad \text{s.t.} \quad f_j(x) \geq \varepsilon_j, \ \ j = 2, \ldots, q. \tag{8}$$

A plan x^* is *efficient* (Pareto optimal) if no feasible x is at least as good on every objective and strictly better on one. Optimal solutions of $P(\varepsilon)$ are weakly efficient, and efficient when unique, and sweeping ε recovers efficient plans that a weighted sum cannot reach when the feasible set is integer or non-convex. The lexicographic method is the special case in which earlier objectives are fixed within a tolerance τ: solve $z_1^* = \max f_1$, then maximize f_2 subject to $f_1(x) \geq (1 - \tau) z_1^*$, and so on.

Applied to profit and the fleet-emissions target in the illustration, the frontier is:

Fleet-average cap (g/km)	Contribution	Units of A	Cost per g/km tightened
131 or looser	\$1.85M	150	not binding
130	\$1.82M	140	\$51K
125	\$1.59M	63.6	\$46K
120	\$1.40M	0	\$38K

Linear-programming values; the unconstrained mix averages 130.6 g/km, so the cap binds only below that, and the cost column is the contribution lost per g/km from the row above. Leadership chooses a point on this curve rather than a set of weights it cannot interpret.

Stability is a constraint the planners feel

Long-horizon plans are re-solved as inputs change, and degenerate or near-optimal alternatives can make a plan swing for no business reason. A change penalty against the last approved plan x^0 prevents that:

$$x_{ipt} - x^0_{ipt} = u_{ipt} - v_{ipt}, \qquad u, v \geq 0,$$

with $\lambda \sum (u+v)$ subtracted from the objective. At any optimum $\min(u_{ipt}, v_{ipt}) = 0$: if both were positive, lowering each by δ keeps the difference, stays feasible and improves the objective by $2\lambda\delta$. Hence $u + v = |x - x^0|$ exactly, and the penalty is a linear model of the absolute change. The weight λ is a price on disruption and should be set, like any other price, by people who own the consequence. Near-term periods are often frozen entirely by a time fence, $x_{ipt} = x^0_{ipt}$ for $t \leq \tau$.

Infeasibility must be explained, not just reported

With hundreds of rules, an infeasible instance is common and "infeasible" is useless to a planner. Two techniques make it actionable. An *irreducible infeasible subsystem* (IIS) identifies a minimal set of conflicting constraints. *Elastic* constraints add penalized slack and minimize the weighted violation,

$$\min_{x,s} \ \textstyle\sum_k w_k s_k \quad \text{s.t.} \quad Ax - s \leq b, \ \ s \geq 0, \tag{9}$$

with weights w_k ordered by priority, so the model returns the least damaging violation and names the rules it bends. An IIS is found by a deletion filter: drop one constraint at a time and re-test feasibility; if the rest is still infeasible, the dropped constraint is not needed to explain the conflict and is discarded permanently. What remains after one pass is irreducible, at the cost of one feasibility solve per constraint. The site record lists explained infeasibilities among the workflow features that kept the plan reviewable.

Inputs need versions and owners

Every plan is a function of forecasts, costs, capacities and rules, and a plan that cannot be reproduced cannot be defended. The project versioned its forecasts and assumptions. A sound implementation stores each solve's inputs, model version and outputs together, assigns every input an owner (finance for margins, plants for capacities, product planning for mix rules), and logs which inputs changed between plans, so a swing can be traced to its cause.

Solve it inside the planning cycle

Scale arithmetic matters early. Two hundred types over 24 months is 4,800 type-periods before plants and markets multiply them, so production and allocation

variables quickly number in the tens of thousands. Most of the structure is capacity and flow balance, whose linear relaxation is tight, so integer variables should be confined to the genuinely discrete rules. Four practices keep solve times inside the cycle: warm-start from the last approved plan; aggregate distant periods (monthly near term, quarterly beyond) on a rolling horizon; build only eligible variables; and keep a pool of near-optimal solutions, so planners can choose among plans of almost equal value instead of overriding a single answer.

PROCEDURE 1. ROLLING-HORIZON REPLANNING

Require: last approved plan x^0, time fence τ, versioned inputs for each cycle
for each planning cycle **do**
  load versioned forecasts, capacities, margins and rules
  keep monthly periods near term; aggregate later periods to quarters
  fix $x_{ipt} = x^0_{ipt}$ for $t \le \tau$ ▷ time fence
  solve (2)–(7) with change penalty, warm-started at x^0
  if infeasible **then**
    solve the elastic model; report the bent rules and an IIS
  end if
  evaluate the plan on scenarios; report duals of binding resources
  on planner approval, set $x^0 \leftarrow x^*$
end for

4. Uncertainty and validation

Compare scenarios before approving a plan

A plan optimized on point forecasts is optimal for one future. Scenario comparison, part of the project's planner workflow, tests the recommended mix across plausible demand, supply and policy conditions before approval. A fuller treatment optimizes a two-stage stochastic program. The build plan x is decided now; demand ξ^ω is revealed in scenario ω with probability p_ω; sales s^ω then adapt:

$$\max_{x \in X} \sum_\omega p_\omega \, Q(x, \xi^\omega), \tag{10}$$

$$Q(x, \xi) = \max_{0 \le s \le \min(x, \xi)} m^\top s - c^\top (x - s),$$

where c is the loss on an unsold unit. Written with one copy x^ω per scenario, the *nonanticipativity* constraints $x^\omega = x$ for all ω state that the plan cannot depend on information not yet known. Two quantities measure whether this is worth doing. With RP the optimal value above, EEV the expected value of the plan optimized for mean demand, and WS the expected value if demand were known before planning,

$$\text{VSS} = \text{RP} - \text{EEV} \ge 0, \qquad \text{EVPI} = \text{WS} - \text{RP} \ge 0.$$

In the illustration, with demand for A of 250, 400 or 550 and for B of 450, 700 or 950, all nine combinations equally likely, and an unsold vehicle losing $1,500 (A) or $1,000 (B), the mean-demand plan (150 A, 700 B) earns $1.60M in expectation. The stochastic plan builds 250 A and 500 B and earns $1.70M: a VSS of $100K, or 6%. The last units of B in the mean plan sell only when demand is high, so their expected value per hour falls below that of more A. The EVPI is $142K, an upper bound on what better demand information could be worth to this decision.

Validate the plan and the value separately

Plan validity is checked by replaying the model on historical cycles and auditing every recommendation against plant feasibility. Value is a counterfactual and needs a defined baseline: the project reported its results against the relevant planning baseline. A sound evaluation states the baseline, holds data and constraints fixed across the comparison, and reports a range rather than a point.

Check	Question it answers
Feasibility audit	Can every plant actually build the recommended mix?
Historical replay	Would the model have produced sensible plans in past cycles?
Scenario stress	How does the plan perform when demand or supply moves?
Stability	How much does the plan change for a small input change?
Value	What did the plan earn against a stated baseline, under the same constraints?

REFERENCE DECISION FLOW

Versioned demand and economics → constraints with owners → allocation MILP → scenarios, duals and planner review → approved build mix

5. What the project reported

The project reported a production-allocation improvement of approximately 40,000 additional vehicles per year and approximately $390 million in realized incremental profit attributed to this work, measured against the relevant planning baseline. These are the figures as the project reported them; this paper does not add an independent audit or a public reconciliation to the underlying financial statements.

Where the value comes from

The mechanism is a better feasible product mix. When a constrained resource is allocated using a consistent objective, production moves to combinations with stronger enterprise value while respecting the network. Even with fixed capacity, economic output changes because the allocation of that capacity changes. Secondary benefits, such as faster scenario testing, are unquantified here.

6. Lessons that transfer beyond automotive

Report shadow prices, not only plans

The value of an extra hour, part or shift is the most useful number an allocation model produces, and planners rarely see it.

Classify rules before turning them into code

Physical constraints must hold; economic preferences belong in the objective or an explicit trade-off; historical conventions deserve scrutiny. Every preference coded as a hard constraint protects yesterday's choices at the expense of tomorrow's value.

A USEFUL EXECUTIVE DIAGNOSTIC

If one more hour of your scarcest operation became available next month, does anyone know what it is worth, and which product it should build?

7. What a modern version would add

Beyond the scenario and two-stage treatment above, a modern version would simulate repeated replanning to test the policy rather than a single plan, keep an auditable record of overrides and approved trade-offs, and measure realized performance against what the model predicted. None of these is a claim about the original project; each earns its place only if it changes the preferred mix enough to justify its cost.

Further reading

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

Pochet, Y. and Wolsey, L. A. *Production Planning by Mixed Integer Programming*. Springer, 2006.

Ehrgott, M. *Multicriteria Optimization*, 2nd ed. Springer, 2005.

Chinneck, J. W. *Feasibility and Infeasibility in Optimization*. Springer, 2008.

Birge, J. R. and Louveaux, F. *Introduction to Stochastic Programming*, 2nd ed. Springer, 2011.

Chankong, V. and Haimes, Y. Y. *Multiobjective Decision Making: Theory and Methodology*. North-Holland, 1983.

Bertsimas, D. and Tsitsiklis, J. N. *Introduction to Linear Optimization*. Athena Scientific, 1997.

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

About the case and Bit Bros

This paper draws on automotive supply-allocation work led by **Adam DeJans Jr.** in an earlier senior industry role. It was not a Bit Bros client engagement. The manufacturer's identity, proprietary model, client data, and individual operational constraints are not disclosed. The mathematical examples are intentionally simplified to explain the decision architecture; they are not represented as production source code. Reported project metrics have not been independently verified for this publication. Bit Bros applies optimization, simulation, and decision science to operating problems, building systems that turn forecasts, constraints, and economics into repeatable decisions.

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

Better decisions are built, not merely calculated.