Data-Driven Fleet Lifecycle Management: Integrating Interpretable Machine Learning and Stochastic Optimization

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Lanette Marie Phillips

May 2026

© by Lanette Marie Phillips, 2026
All Rights Reserved.

For my parents and family, whose patience, love, and example instilled in me the values that guide this work.

And for my mentors, colleagues, and friends, whose steady support and thoughtful counsel made this achievement possible.

Acknowledgements

I would like to express my deepest appreciation to my **dissertation advisor and committee chair, Dr. Yu**, for his guidance, patience, and invaluable insight throughout this research. His thoughtful mentorship has shaped both the direction of this dissertation and my broader understanding of decision analytics and public-sector optimization.

I also wish to thank the members of my **dissertation committee—Dr. He, Dr. Ostrowski, and Dr. Schelble**—for their constructive feedback and continued support in refining the scope and clarity of this work. Their expertise and perspective greatly strengthened this study.

I am sincerely grateful to my colleagues and mentors in the **Department of Industrial and Systems Engineering at the University of Tennessee, Knoxville**, whose collaboration, technical discussions, and encouragement enriched this research and my experience as a doctoral student.

Finally, I extend my heartfelt thanks to my family and friends for their steady encouragement and understanding throughout this journey. Their support made this work—and its completion—possible.

"The great decisions are not made by numbers alone."

— Richard Bellman

Abstract

Public agencies manage large vehicle fleets under strict budget constraints, service-level requirements, and public accountability. Replacement and procurement decisions are often governed by fixed age or mileage rules that simplify administration but fail to reflect heterogeneous usage patterns, nonlinear cost behavior, and uncertainty in maintenance, downtime, and operational exposure. This dissertation develops an integrated analytical framework that combines stochastic optimization, interpretable machine learning, empirical lifecycle analysis, and decision-focused integration to support transparent and economically defensible fleet planning in the public sector.

The dissertation first formulates fleet replacement and capacity planning as a two-stage stochastic program capturing the commitment–recourse trade-off between owned capacity and adaptive recourse decisions under uncertain demand and operating conditions. Classical decomposition methods—Benders Decomposition, Lagrangian Relaxation, and Progressive Hedging—are evaluated to demonstrate tractability, policy consistency, and complementary forms of transparency.

It then develops a heterogeneity-aware, interpretable machine-learning framework for vehicle-year operating cost prediction using variance-stabilized gradient boosting models trained on log-transformed cost. Under VIN-safe and agency-blocked validation that prevents temporal and organizational leakage, log-target XGBoost achieves strong predictive accuracy and calibration in the presence of heavy-tailed maintenance expenditures. SHAP-based explanations decompose predicted cost

on the log scale, revealing nonlinear age effects, ownership–utilization interactions, and agency-level structural variation. Counterfactual simulations show no stable lease–own crossover within the empirically supported 4–7 year policy window once censoring and compositional effects are controlled.

Building on these results, the dissertation empirically estimates class-specific diminishing-returns replacement ages through nonparametric smoothing of sixteen years of statewide fleet data. The analysis identifies the ages at which lifetime-average annual operating cost begins to increase for major vehicle classes, demonstrating substantial heterogeneity and highlighting the inefficiency of uniform replacement cycles.

Finally, the framework integrates prediction and optimization through a hybrid Decision–Coupled Learning approach in which predictive cost structure, empirically estimated lifecycle behavior, and uncertainty distributions parameterize the stochastic optimization model. Model performance is evaluated based on downstream economic decision quality—measured through expected cost, risk exposure, and policy regret—rather than statistical error alone, producing replacement policies that are uncertainty-aware, economically grounded, and operationally stable.

Together, these components establish a transparent, reproducible, and uncertainty-aware decision infrastructure for public-sector fleet management, linking observed lifecycle cost behavior, interpretable prediction, and prescriptive optimization into a unified analytical system.

Table of Contents

1 Introduction and Motivation **1**

1.1 Problem Definition and Research Objectives 2

1.1.1 Formal Decision Structure of the Fleet Replacement Problem . 4

1.1.2 Dissertation Contributions 6

1.2 Data and Case Setting (Overview) 6

1.3 Institutional Context and Limitations of Current Replacement Practices 7

1.3.1 Budget and Capital Structure 7

1.3.2 Procurement Rules and Administrative Simplicity 8

1.3.3 Operational Heterogeneity and Censoring Effects 8

1.3.4 Risk, Uncertainty, and Policy Communication 9

1.4 High-Level Research Gaps 9

1.5 Key Findings and Impact 13

1.6 Dissertation Organization 13

1.6.1 Scope and Boundaries 14

2 Literature Review and Theoretical Positioning **15**

2.1 Overview 16

2.2 Fleet Management and Lifecycle Cost Modeling 17

2.3 Stochastic Programming and Decomposition Methods 18

2.4 Machine Learning for Cost Modeling and Interpretability 22

2.5 Decision-Focused and Hybrid Methods 23

2.6 Empirical Context and Data Challenges . . . 23
2.7 Comparative Limitations of Existing Approaches . . . 24
2.7.1 Deterministic Lifecycle Cost Models . . . 24
2.7.2 Pure Stochastic Optimization Models . . . 24
2.7.3 Pure Predictive Machine Learning Models . . . 25
2.7.4 Decision-Focused and Prescriptive Learning . . . 25
2.7.5 Implications for Public-Sector Fleet Management . . . 25
2.8 Synthesis and Research Positioning . . . 26

3 Stochastic Optimization for Fleet Planning under Uncertainty 28
3.1 Structural Derivation of the Stochastic Fleet Model . . . 29
3.1.1 Commitment and Recourse Structure . . . 29
3.1.2 Sources of Uncertainty . . . 30
3.1.3 Why Deterministic Models Are Insufficient . . . 30
3.1.4 Resulting Two-Stage Structure . . . 31
3.1.5 Implications for Decomposition . . . 31
3.2 Two-Stage Stochastic Programming Preliminaries . . . 31
3.3 Model . . . 33
3.3.1 Sets and Indices . . . 34
3.3.2 Parameters . . . 34
3.3.3 Calibration Values Used in Experiments . . . 34
3.3.4 Decision Variables . . . 35
3.3.5 Two-Stage Stochastic Program . . . 36
3.4 Economic Interpretation of Dual Variables and Recourse Sensitivity . 37
3.4.1 Shadow Prices of Capacity and Demand . . . 37
3.4.2 Recourse Function as a Value-of-Flexibility Measure . . . 37
3.4.3 Interpretation of Benders Cuts . . . 38
3.4.4 Lagrangian Multipliers and Coordination Penalties . . . 38
3.4.5 Policy-Relevant Sensitivity Insights . . . 39

3.5 Positioning within Stochastic Fleet and Capacity Planning 39
3.6 Methods 40
3.7 Computational Results 43
3.8 Managerial Insights 46
3.9 Discussion and Conclusion 47

4 Machine Learning and Explainable Cost Modeling 50
4.1 Data and Feature Preparation 51
4.1.1 Fleet Dataset Overview 51
4.2 Data Pipeline and Preprocessing 52
4.2.1 Features 53
4.2.2 Target Variable 54
4.2.3 Preprocessing and Validation Design 54
4.2.4 Validation Strategy and Leakage Prevention 55
4.3 Why Tree-Based Ensembles? Methodological Rationale 55
4.4 Predictive Modeling Layer 56
4.5 Model Validation and Benchmarking 57
4.6 Explainability via SHAP 60
4.6.1 Global Insights 61
4.6.2 Counterfactual and Age–Ownership Effects 63
4.6.3 Partial Dependence: Joint Effects 65
4.7 Summary and Integration with Optimization 66

5 Empirical Diminishing Returns in Fleet Lifecycle Costs 68
5.1 Background and Motivation 69
5.2 Diminishing-Returns Age Framework 69
5.3 Data and Descriptive Analysis 70
5.3.1 Data Sources and Scope 70
5.3.2 Descriptive Cost Patterns 71
5.4 Methods 71

5.4.1 Annualized Operating Cost . . . 71
5.4.2 Vehicle Age Construction . . . 71
5.4.3 Nonparametric Smoothing . . . 72
5.4.4 Tipping-Age Identification . . . 72
5.5 Results . . . 72
5.5.1 Cross-Class Comparison . . . 72
5.5.2 Sedans . . . 74
5.5.3 SUVs . . . 74
5.5.4 Pickup Trucks . . . 75
5.5.5 Cross-Class Interpretation . . . 75
5.6 Policy Implications and Managerial Use . . . 75
5.7 Connection to Optimization and Hybrid Models . . . 75
5.8 Limitations and Future Work . . . 76
5.9 Conclusion . . . 76

6 Hybrid Predict–Then–Optimize Framework 78
6.1 Conceptual Overview . . . 78
6.2 Data and Feature Preparation . . . 80
6.2.1 Fleet Dataset Overview . . . 80
6.3 Data Pipeline and Preprocessing . . . 81
6.4 Predictive Modeling Layer . . . 83
6.5 Stochastic Optimization Model . . . 84
6.6 Hybrid Decision-Focused Integration . . . 85
6.6.1 Positioning of the Hybrid Framework within Decision-Focused Learning . . . 87
6.7 Evaluation Framework . . . 88
6.7.1 Economic Implications of Decision-Coupled Learning . . . 89
6.8 Sensitivity and Robustness Analyses . . . 90

6.8.1 Policy Sensitivity to Capital Cost Assumptions 91
6.9 Outcomes and Delivered Artifacts 91
6.10 Summary and Integration of the Hybrid Framework 93

7 Integration and Conclusion **94**
7.1 Overview 94
7.2 Synthesis of Contributions 95
7.2.1 Stochastic Optimization and Decomposition 95
7.2.2 Interpretable Machine Learning for Cost Modeling 96
7.2.3 Empirical Identification of Diminishing Returns and Economic Life 96
7.2.4 Hybrid Predict–Then–Optimize Integration 97
7.3 Managerial and Policy Implications 97
7.4 Conceptual and Methodological Integration 98
7.5 Scope and Limitations 99
7.6 Future Research Directions 100
7.7 Concluding Remarks 100

Bibliography **102**

Vita **112**

List of Tables

2.1 Selected studies on lifecycle cost modeling and fleet replacement. . . . 19

2.2 Summary of literature streams, research gaps, and dissertation contributions. . . . 21

3.1 Excess owned capacity (owned fleet minus scenario demand) under the optimal expansion. . . . 43

3.2 Comparison of decomposition methods on the test instance. . . . 45

4.1 Key features retained for predictive and optimization modeling. . . . 54

4.2 Leakage-safe hold-out performance (original dollar scale). . . . 58

5.1 Empirical diminishing-returns results and data coverage by vehicle class. 73

6.1 Fleet composition by ownership and vehicle category. . . . 81

6.2 Key features retained for predictive and optimization modeling. . . . 82

List of Figures

1.1 Conceptual roadmap showing the integration of stochastic optimization, interpretable machine learning, and hybrid predictive–prescriptive modeling across the dissertation's three studies. 3

2.1 Conceptual positioning of the dissertation within four major research streams: lifecycle cost and fleet management, stochastic programming and decomposition, machine learning and interpretability, and decision-focused hybrid methods. 17

3.1 Structure of the two-stage fleet model. First-stage ownership decisions are made before demand is realized; scenario-specific recourse actions ensure feasibility under each demand outcome. 36

3.2 Benders convergence: lower and upper bounds tighten monotonically and meet at $5,770,750 after five iterations. 43

3.3 Lagrangian Relaxation convergence: the dual bound matches the optimal objective on this instance after one multiplier update. 44

3.4 Progressive Hedging convergence: scenario decisions align within eight iterations as nonanticipativity residuals fall below the stopping tolerance. 44

3.5 Comparison of the three decomposition methods on the calibrated fleet instance. All methods identify the same optimal expansion plan and expected cost. 45

4.1 Distribution of asset age by ownership type. Leased vehicles are concentrated in early service years due to contractual turnover and programmatic replacement policies, whereas owned vehicles extend to substantially higher ages. Vertical dashed lines indicate the 4–7 year policy window commonly associated with mid-life replacement discussions. 52
4.2 Machine learning workflow for lifecycle cost modeling, from raw data to predictive models and explainable outputs. The pipeline enforces leakage-safe validation and supports direct integration with the stochastic optimization model. 53
4.3 Observed versus predicted annual cost on the leakage-free hold-out set using the log-target XGBoost model. Predictions are evaluated on the original dollar scale after back-transformation. 58
4.4 Calibration curve by prediction decile for the hold-out set. The log-target XGBoost model remains close to the ideal diagonal across most cost bins, with modest underprediction in the extreme upper tail. . . 59
4.5 Residuals versus asset age on the hold-out set for the log-target XGBoost model. LOWESS smoothing (solid line) highlights lifecycle-dependent dispersion patterns. 60
4.6 Global SHAP feature importance for the log-target XGBoost model. Bars represent mean absolute SHAP value on the log-cost scale, indicating average contribution magnitude to predicted annual cost. . 61
4.7 SHAP dependence plot for Asset Age under the log-target XGBoost model. Points represent individual observations; the smoothed curve highlights the nonlinear marginal contribution of age to predicted log annual cost. 62

4.8 Observed annual cost patterns (left) and SHAP-implied model contributions (right) by asset age. The right panel reflects SHAP values from the log-target XGBoost model; additive contributions on the log scale correspond to multiplicative effects in dollar terms. Both panels show no lease–own crossover in the 4–7 year range once empirical support constraints are accounted for, countering naive interpretations based on raw averages. 63

4.9 Counterfactual ownership–lease cost differential across asset age. Negative values indicate a cost advantage for ownership. The shaded region reflects empirically supported age overlap in the data, and crossover analysis is restricted to the 4–7 year policy window corresponding to practical replacement decision horizons. 64

4.10 Mean and 95% confidence interval of cost difference by age. Predictive uncertainty increases with vehicle age, reinforcing the importance of uncertainty-aware replacement planning. 65

4.11 Partial dependence of annual cost on asset age and ownership type. Costs rise nonlinearly with age and exhibit distinct ownership effects across the lifecycle. 66

5.1 Smoothed lifetime-average operating cost per year by vehicle age for sedans, SUVs, and pickup trucks using LOWESS. Dashed vertical lines indicate class-specific diminishing-returns tipping ages. The locations of these tipping points vary substantially across vehicle classes, indicating that uniform replacement cycles are economically inefficient. 73

6.1 Conceptual architecture of the hybrid predict–then–optimize framework linking ML prediction, uncertainty embedding, and policy optimization. 79

Nomenclature

α_k	Step size at iteration k (LR)
$\bar{C}_i$	Lifetime-average annual operating cost for vehicle i
$\bar{x}^{(k)}$	Consensus first-stage solution at iteration k (LR/PH)
$\boldsymbol{\theta}$	Parameter vector of the machine-learning model
Δx^*	Optimal first-stage expansion vector (by vehicle class)
$\frac{d\hat{C}(a)}{da}$	Slope of the smoothed lifetime-average cost curve with respect to age (used to detect tipping)
$\hat{C}(a)$	LOWESS-smoothed lifetime-average operating cost per year at age a
κ	Risk adjustment factor applied to cost variance
λ_s	Lagrange multiplier associated with scenario s nonanticipativity constraint
$\lambda_s^{(k)}$	Value of multiplier λ_s at iteration k
$\mathbb{E}[\cdot]$	Expectation operator
$\mathbb{Z}_+$	Set of nonnegative integers
$\mathcal{L}_{\mathrm{DCL}}$	Decision–Coupled Learning loss function

$\mathcal{O}$	Optimization operator representing the two-stage fleet model
$\mathcal{X}_v$	Feasible decision set for vehicle class v
Regret($\boldsymbol{\theta}$)	Policy regret induced by model parameters $\boldsymbol{\theta}$
$\pi(\cdot)$	Replacement policy induced by the stochastic optimization model
π_s	Dual variables for scenario-s recourse constraints (Benders multipliers)
π_{oracle}	Oracle replacement policy computed using realized costs
ρ	Progressive Hedging penalty parameter
τ	Empirical diminishing-returns tipping age
$\tilde{\xi}$	Random vector of uncertain parameters
ε	Stopping tolerance
ξ	A realization of $\tilde{\xi}$
ξ_s	Scenario realization for scenario s
A	First-stage constraint matrix in the compact stochastic formulation
a	Age (years) used as a continuous evaluation grid for the smoothed cost curve
A_i	Observed service age (years) for vehicle i
B	Recourse (second-stage) constraint matrix in the compact stochastic formulation
c	First-stage cost coefficient vector

c_v^{idle}	Idle carrying cost per unit of vehicle type v
C_i	Cumulative observed operating and maintenance cost for vehicle i
$d_{\cdot,s}$	Demand vector across vehicle classes under scenario s
$f_s(x_s, y_s)$	Scenario-s objective function
$f_{\boldsymbol{\theta}}$	Machine-learning prediction function
$h(\xi)$	Right-hand side vector in recourse constraints
k	Iteration counter
LB, UB	Lower and upper bounds on the optimal objective value
$LB^{(k)}$	Best lower bound at iteration k
n_1, n_2	Dimensions of the integer and continuous components of x
$q(\xi)$	Scenario-dependent recourse cost vector
$Q(x, \xi)$	Second-stage (recourse) optimal value given first-stage x and realization ξ
q_s, T_s, W_s, h_s	Scenario-specific data for scenario s
$T(\xi)$	Technology matrix coupling first- and second-stage decisions
$W(\xi)$	Recourse constraint matrix
$w_{v,s}$	Idle owned units of vehicle type v in scenario s
X	Feasible set for first-stage decisions
x	First-stage decision vector (here-and-now)
$x^{(k)}$	First-stage solution at iteration k (BD)

$x_s^{(k)}$	Scenario-s first-stage solution at iteration k (LR/PH)
x_v^0	Initial owned fleet of type v
y	Second-stage recourse decision vector (generic two-stage formulation)
$Y(\xi)$	Feasible set for recourse decisions under realization ξ
y_s	Recourse decision vector in scenario s
$y_{v,s}$	Emergency procurement for vehicle type v in scenario s
$z_M^{(k)}$	Master-problem objective value at iteration k

Chapter 1

Introduction and Motivation

Public agencies and municipal fleet operators manage heterogeneous vehicle assets under tight budgets, service-level requirements, and rising uncertainty in maintenance, supply chains, and residual values. Replacement decisions that appear simple—"when should we retire and replace this vehicle?"—are complicated by non-linear aging, downtime risks, and volatile capital markets. Traditional analytics often follow a *predict–then–optimize* pipeline: first fit a model to annual operating costs, then plug predictions into a life-cycle cost (LCC) or equivalent annual cost (EAC) optimizer. However, statistical accuracy does not necessarily translate into better decisions when the objective is economic rather than predictive (Bertsimas and Kallus, 2020; Elmachtoub and Grigas, 2022). This disconnect is especially important for high-value, long-lived public assets.

In public-sector settings—where transparency, fairness, and defensibility are essential—this disconnect poses a practical barrier to adopting advanced analytics.

This dissertation develops a unified analytical framework that integrates stochastic optimization, interpretable machine learning, and predictive–prescriptive modeling to improve fleet replacement planning under uncertainty while preserving interpretability and institutional defensibility. The work is organized around three complementary

methodological components corresponding to optimization, prediction, and their integration.

Building on recent advances in decision-focused learning and structured prediction (Elmachtoub and Grigas, 2022; Ban and Rudin, 2019), the dissertation explores how predictive models can be aligned with downstream economic objectives rather than trained solely for statistical accuracy. In particular, predictive outputs are shaped to reflect the equivalent annual cost (EAC) surface faced by fleet managers, with interpretable risk and downtime penalties that yield interior-optimal replacement ages when economically justified. This integrated perspective enables analytically grounded, policy-relevant recommendations that connect lifecycle cost estimation directly to replacement and procurement decisions.

1.1 Problem Definition and Research Objectives

Public-sector fleet managers face a fundamentally multi-layered decision problem under uncertainty. Agencies must determine when to retire vehicles and how much capacity to own while navigating nonlinear aging costs, downtime risk, capital budget constraints, procurement delays, and heterogeneous usage patterns. Replacement decisions must balance expected annual cost (EAC), operational reliability, safety considerations, and fiscal accountability in the presence of uncertain future maintenance, utilization, and salvage values.

The core challenge is not merely predicting operating cost, but making economically defensible replacement and procurement decisions when future conditions are uncertain. Traditional rule-based policies (e.g., retire sedans at 120,000 miles or 10 years) provide administrative simplicity but ignore heterogeneity in duty cycles and agency mission, leading to premature retirements, prolonged use of inefficient assets, and inconsistent policy justification.

The central research question is therefore:

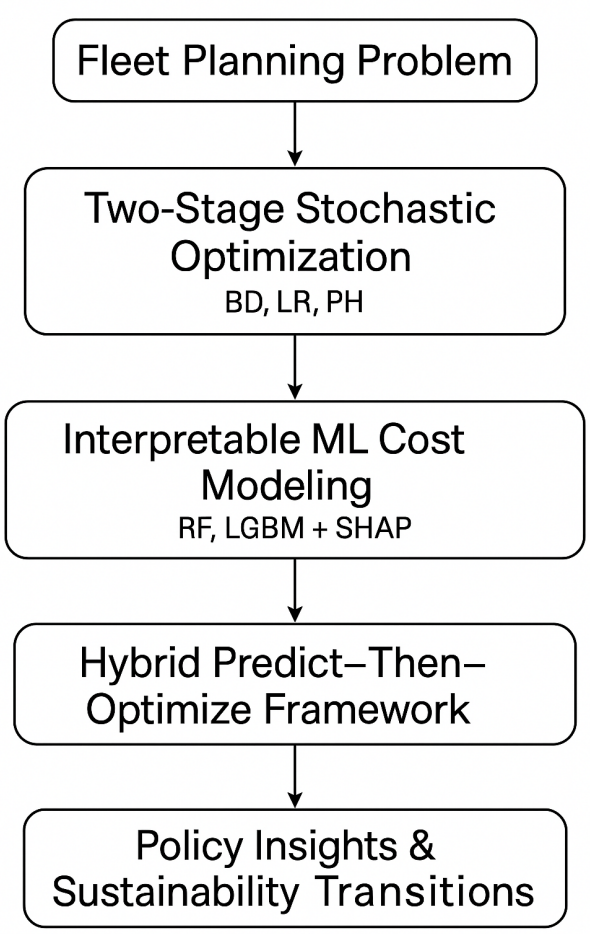


Figure 1.1: Conceptual roadmap showing the integration of stochastic optimization, interpretable machine learning, and hybrid predictive–prescriptive modeling across the dissertation's three studies.

> **How can data-driven, interpretable, and uncertainty-aware models improve fleet replacement decisions for public agencies while remaining transparent, defensible, and computationally tractable?**

To address this problem, the dissertation develops an integrated analytical framework built on three complementary methodological components:

1. **Stochastic Optimization.** Formulate fleet planning as a two-stage stochastic program capturing uncertainty in operating cost, downtime, demand, and salvage

outcomes, and evaluate classical decomposition methods—Benders Decomposition (BD), Lagrangian Relaxation (LR), and Progressive Hedging (PH)—for scalability, interpretability, and policy traceability.

2. **Interpretable Machine Learning.** Develop predictive models for lifecycle cost and downtime using tree-based ensembles with SHAP-based explanations to quantify nonlinear age effects, utilization heterogeneity, organizational variation, and ownership impacts while preventing temporal and cross-agency information leakage.
3. **Decision-Focused Integration.** Construct a Decision–Coupled Learning (DCL) framework that embeds economic objectives directly into predictive evaluation, aligning model development with downstream decision quality rather than statistical accuracy alone.

Together, these components transform fleet replacement from a heuristic, rule-based process into a transparent, uncertainty-aware, and economically grounded decision system suitable for public-sector governance environments.

1.1.1 Formal Decision Structure of the Fleet Replacement Problem

At its core, the public-sector fleet replacement problem is a structured economic decision problem under uncertainty. Let $v \in V$ denote vehicle classes and t represent discrete planning periods. The decision-maker must determine whether to retain, replace, lease, or expand capacity for each class, subject to budgetary, operational, and policy constraints.

The objective is to minimize total expected lifecycle cost over the planning horizon, typically expressed in equivalent annual cost (EAC) terms. Formally, the planner seeks to choose decisions x that minimize

$$\min_{x \in \mathcal{X}} \ \mathbb{E}\big[C(x, \xi)\big],$$

where $C(x, \xi)$ denotes total lifecycle cost under uncertain future outcomes ξ, including maintenance cost escalation, downtime events, demand fluctuations, salvage value realization, and capital cost variation.

The feasible decision set $\mathcal{X}$ reflects several structural features:

- **Budget constraints**, limiting capital expenditures in each fiscal period;
- **Service constraints**, ensuring sufficient vehicle availability to meet agency demand;
- **Indivisibility constraints**, since many fleet assets are discrete units rather than divisible capacity;
- **Institutional constraints**, including procurement rules and replacement eligibility criteria.

This structure creates a multi-stage commitment–recourse problem: replacement and procurement decisions are made prior to observing future maintenance trajectories and demand shocks, while operational adjustments occur after uncertainty is realized.

Two features make the problem analytically challenging:

1. **Nonlinear and heterogeneous cost growth**: maintenance and downtime costs accelerate at different rates across vehicle classes.
2. **Censoring and observational bias**: leased and owned vehicles are observed over different service horizons.

Existing rule-based approaches effectively assume that $C(x, \xi)$ follows a homogeneous and deterministic structure, replacing the stochastic expectation with a fixed age or mileage threshold. The dissertation instead models the uncertainty explicitly, estimates cost trajectories empirically, and aligns predictive modeling with the economic decision function itself.

The remainder of the dissertation develops analytical tools that address each dimension of this formal structure: stochastic optimization for commitment and recourse decisions, interpretable machine learning for heterogeneous cost estimation, empirical analysis of diminishing returns in lifecycle cost, and a hybrid decision-focused framework that integrates prediction and optimization.

1.1.2 Dissertation Contributions

This dissertation makes four primary contributions:

1. It formulates public-sector fleet replacement as a structured two-stage stochastic planning problem and evaluates classical decomposition methods within a calibrated agency setting.
2. It develops an interpretable machine-learning framework that quantifies heterogeneous lifecycle cost drivers while preventing temporal and organizational information leakage.
3. It empirically identifies class-specific diminishing-returns replacement ages using nonparametric smoothing of long-horizon fleet data, bridging descriptive cost behavior and economic replacement theory.
4. It develops and evaluates a decision-coupled learning framework that aligns predictive modeling with downstream economic policy quality, measured through equivalent annual cost and policy regret.

Collectively, these contributions integrate uncertainty modeling, interpretability, and decision-focused evaluation into a unified analytical framework for public-sector capital planning.

1.2 Data and Case Setting (Overview)

The empirical analyses draw on sixteen years (2009–2025) of state fleet data integrating work orders, fuel transactions, mileage, and purchase–sale records across

major vehicle classes. Predictors include vehicle attributes, duty cycles, and maintenance histories; outcomes capture operating and downtime costs. Table 6.1 in Chapter 4 summarizes the baseline composition of owned and leased vehicles.

The fleet is heterogeneous in both composition and use. Sedans and pickups dominate overall counts, but specialized units (police, heavy equipment) and agency-specific duty cycles create diverse cost trajectories. Leased vehicles cluster in younger age bands, whereas owned vehicles exhibit long right tails in age and mileage. These empirical patterns motivate the dissertation's emphasis on heterogeneity, censoring, and uncertainty.

1.3 Institutional Context and Limitations of Current Replacement Practices

Public-sector fleet agencies operate within a layered institutional environment that materially shapes replacement and procurement decisions. Unlike private firms that optimize primarily on financial grounds, state fleets must balance budget structures, procurement law, operational constraints, and public accountability. These institutional features both constrain analytical flexibility and explain the persistence of rule-based replacement policies.

1.3.1 Budget and Capital Structure

Most U.S. states—including the case agency—separate capital and operating budgets. Vehicle acquisition is financed through capital appropriations, while maintenance, fuel, and downtime are recorded as operating expenditures. This structural separation creates intertemporal tension: deferring replacement reduces short-term capital outlay but often increases operating cost through maintenance escalation, fuel inefficiency, and higher downtime risk.

Annual budget cycles further incentivize short-term expenditure smoothing rather than lifecycle efficiency. As a result, replacement timing is often driven by available capital rather than by economically optimal lifecycle considerations.

1.3.2 Procurement Rules and Administrative Simplicity

Vehicle acquisition is governed by statutory requirements, competitive bidding, vendor contracts, and eligibility criteria. To simplify administration, many agencies rely on fixed age–mileage matrices (e.g., retire sedans at 120,000 miles or ten years). While administratively transparent, such rules implicitly assume homogeneous cost trajectories and deterministic future conditions.

These rules neglect heterogeneity in duty cycles, maintenance environments, and utilization intensity. They also ignore uncertainty in maintenance escalation, salvage value, demand variability, and supply-chain lead times, which can range from six to eighteen months. Consequently, rule-based policies often misalign replacement timing with true economic life.

1.3.3 Operational Heterogeneity and Censoring Effects

Fleet systems are inherently heterogeneous. Some vehicles accumulate more than 25,000 miles annually, while others operate fewer than 5,000. Mission-critical units exhibit high downtime sensitivity, whereas administrative vehicles face different risk profiles. Leasing arrangements further distort observed cost patterns: leased vehicles are rarely observed beyond five to six years, while owned vehicles frequently exceed twelve to fourteen years of service.

This censoring creates biased lease–own comparisons and obscures true lifecycle cost trajectories unless utilization and service horizon are explicitly controlled. Without empirical adjustment, administrative rules may reinforce misleading comparisons rather than correct them.

1.3.4 Risk, Uncertainty, and Policy Communication

Most rule-based systems treat maintenance cost growth, catastrophic failure risk, salvage volatility, and supply-chain delays as deterministic averages. Such simplifications overlook asymmetric recourse costs and uncertainty-driven risk that can materially alter optimal replacement timing.

At the same time, public-sector accountability requires analytical tools to be interpretable and defensible. Legislators and auditors frequently ask: "Why this replacement age? Why this budget level?" Fixed thresholds provide simplicity but lack economic justification. Transparent stochastic models and interpretable machine-learning diagnostics provide economically meaningful explanations—such as marginal capacity value, cost acceleration, and risk exposure—that support defensible policy decisions.

Taken together, these institutional constraints and operational realities explain both the persistence and the limitations of current replacement practices. They also motivate the need for analytical frameworks that are uncertainty-aware, empirically grounded, and interpretable within a public governance environment.

1.4 High-Level Research Gaps

Despite progress in lifecycle modeling, stochastic programming, and machine learning, their integration for public-sector fleet management remains limited. Three persistent gaps motivate this dissertation.

Gap 1: Deterministic or Oversimplified Cost Structures

Many fleet planning frameworks rely on deterministic lifecycle cost curves or simplified parametric assumptions. Such models implicitly replace $\mathbb{E}[C(x, \xi)]$ with $C(x, \bar{\xi})$ for an average realization $\bar{\xi}$, thereby neglecting asymmetric recourse costs, nonlinear maintenance escalation, and heterogeneous duty cycles (Pantuso et al., 2014).

In capital-intensive public systems, underestimating adverse demand or failure realizations can be substantially more costly than modest over-preparedness. Deterministic approximations therefore risk systematically misrepresenting the economic value of preparedness and misestimating optimal replacement timing.

Gap 2: Misalignment Between Predictive Loss and Economic Objectives

Machine learning models are typically trained to minimize statistical loss functions such as mean squared error (MSE). However, minimizing prediction error does not guarantee minimization of economic regret. A model that is statistically accurate may still induce suboptimal replacement timing, misallocate capital, or underrepresent tail risks that materially affect budget outcomes.

This disconnect arises because conventional predictive objectives are agnostic to the downstream decision function: they penalize average forecast deviation rather than economically consequential errors near replacement thresholds or capacity constraints. As emphasized in decision-focused learning research (Elmachtoub and Grigas, 2022), statistical optimality and decision optimality are not equivalent.

Gap 3: Limited Integration of Interpretability and Policy Diagnostics

Public-sector decision environments require interpretable methods, documented logic, and defensible outcomes. Yet many optimization and machine-learning frameworks remain opaque, providing limited insight into cost drivers, marginal values, or policy sensitivity.

Few studies combine interpretable predictive modeling, scenario-based optimization, and economically meaningful diagnostics—such as shadow prices, regret measures, and replacement thresholds—within a unified analytical pipeline. Without such

integration, advanced analytics risk being technically sophisticated but institutionally unusable.

Gap 4: Limited Empirical Identification of Economic Life and Diminishing Returns

Classical replacement theory characterizes economic life as the age at which marginal operating cost growth exceeds the annualized benefit of continued ownership (Jardine and Tsang, 2021; Hartman, 2004). In practice, however, public-sector fleets rarely estimate economic life empirically from longitudinal data. Replacement thresholds are typically set through fixed age–mileage matrices or budget-driven heuristics rather than data-driven identification of cost inflection points.

While stochastic programming models incorporate uncertainty and machine learning models capture nonlinear cost behavior, relatively few studies systematically estimate class-specific diminishing-returns ages using long-horizon operational data. As a result, agencies lack transparent, empirically grounded benchmarks that connect observed lifecycle cost behavior to replacement timing decisions.

This gap motivates a dedicated empirical analysis of cost–age trajectories to identify economically meaningful tipping ages that bridge descriptive lifecycle patterns and prescriptive replacement policy.

Bridging the Gaps: Chapter-Level Contributions

The four analytical components of this dissertation are structured to address these gaps systematically rather than independently. Each chapter responds to a specific limitation in the existing literature while reinforcing the broader integrated framework.

Chapter 3 addresses Gap 1 by replacing deterministic replacement rules with a two-stage stochastic optimization model that explicitly captures commitment–recourse asymmetry, demand variability, downtime risk, and salvage uncertainty. By evaluating classical decomposition methods within a calibrated public-sector setting, the chapter demonstrates how uncertainty-aware optimization alters replacement and capacity decisions relative to deterministic benchmarks.

Chapter 4 responds to Gap 3 by developing an interpretable machine-learning framework for lifecycle cost prediction. Tree-based ensemble models combined with SHAP-based explanations quantify nonlinear age effects, ownership heterogeneity, utilization differences, and agency-level variation, while validation procedures prevent temporal and cross-organizational information leakage. This ensures predictive insight remains transparent and institutionally defensible.

Chapter 5 addresses Gap 4 by empirically estimating class-specific diminishing-returns ages using nonparametric smoothing of long-horizon fleet data. Rather than relying on fixed age–mileage thresholds, this chapter identifies data-driven economic tipping points in operating cost trajectories, providing transparent benchmarks that connect descriptive cost behavior to prescriptive replacement timing.

Chapter 6 responds to Gap 2 by integrating prediction and optimization through a Decision–Coupled Learning framework. Predictive cost signals are evaluated using downstream economic loss—measured through equivalent annual cost and policy regret—rather than statistical error alone. This alignment ensures that model improvement is judged by its impact on decision quality, thereby bridging the traditional divide between predictive analytics and prescriptive optimization.

Collectively, these chapters transform fleet replacement from a rule-based or sequential modeling exercise into an integrated, uncertainty-aware, and economically grounded decision system. The framework explicitly links empirical cost behavior, interpretable prediction, and stochastic optimization into a coherent analytical pipeline suitable for public-sector governance environments.

1.5 Key Findings and Impact

The integrated framework developed here yields:

- **Interior-optimal replacement ages** rather than fixed rules;
- **Transparent SHAP-based diagnostics** for cost drivers;
- **Risk-aware stochastic policies** accounting for downtime and salvage uncertainty;
- **Hybrid predictive–prescriptive decisions** that reduce EAC while maintaining interpretability.

For practice, the project provides a reproducible pipeline linking raw data, interpretable prediction, optimization under uncertainty, and decision-focused integration. For research, it demonstrates how machine learning and stochastic programming are aligned to produce economically meaningful policies.

1.6 Dissertation Organization

The dissertation advances a unified framework for improving fleet planning under uncertainty:

- **Chapter 3** develops a two-stage stochastic optimization model for fleet planning under uncertainty and evaluates classical decomposition methods, including Benders Decomposition, Lagrangian Relaxation, and Progressive Hedging.
- **Chapter 4** constructs interpretable machine learning models for lifecycle cost prediction using tree-based ensembles and SHAP-based diagnostics, identifying nonlinear cost drivers related to age, ownership, utilization, and organizational context.
- **Chapter 5** empirically examines diminishing returns in fleet lifecycle costs, estimating class-specific cost–age trajectories and identifying data-driven tipping ages that signal economically inefficient continued operation.

- **Chapter 6** integrates prediction and optimization into a hybrid Decision–Coupled Learning framework that aligns predictive modeling with economic decision quality under uncertainty.
- **Chapter 7** synthesizes implications for multi-year fleet planning, electrification pathways, and public-sector policy, and outlines directions for future research.

This structure reflects the dissertation's overarching aim: combining uncertainty modeling, interpretable ML, and decision-focused integration into a transparent, defensible framework for public-sector fleet management.

1.6.1 Scope and Boundaries

The dissertation focuses on annualized replacement and capacity decisions within a one-period stochastic planning structure, calibrated to multi-year empirical data. Fully dynamic multi-period depreciation, real options modeling, and fully differentiable learning-through-optimization are outside the scope of this study.

The emphasis is instead on transparency, tractability, and institutional implementability within real public-sector constraints.

Chapter 2

Literature Review and Theoretical Positioning

This chapter situates the dissertation within the intersecting literatures of public-sector fleet management, lifecycle cost analysis, stochastic programming, interpretable machine learning, and decision-focused optimization. Each of these research streams has matured substantially in isolation, offering powerful tools for modeling cost dynamics, uncertainty, and predictive relationships. However, their integration into transparent, uncertainty-aware decision frameworks for large public fleets remains limited. In practice, fleet replacement policies are often derived from simplified heuristics or predictive models that are weakly connected to downstream economic objectives and institutional constraints. The purpose of this chapter is to review the theoretical foundations underlying each component of the framework, identify persistent gaps in existing approaches, and motivate the integrated analytical pipeline developed in the subsequent chapters.

2.1 Overview

Public-sector fleet agencies operate under persistent uncertainty in operational demand, maintenance needs, downtime risk, and capital appropriations. Replacement and procurement decisions must be made well in advance of realized conditions and within statutory, fiscal, and political constraints. The literature reviewed here reflects four complementary perspectives:

1. lifecycle cost modeling and asset-level cost dynamics;
2. stochastic programming for fleet and capacity planning under uncertainty;
3. machine learning for cost prediction and interpretability;
4. hybrid predictive–prescriptive approaches linking learning with optimization.

Together, these perspectives motivate the dissertation's objective: to develop interpretable, economically grounded, and uncertainty-aware decision tools for public-sector fleet management. The stochastic optimization components developed in later chapters build on established formulations of two-stage stochastic programming and sample-based scenario modeling (Shapiro et al., 2009; King and Wallace, 2012; Kleywegt et al., 2002).

Classical fleet planning models emphasized optimal capacity sizing and maintenance policies under uncertainty, establishing the commitment–recourse trade-offs that motivate modern stochastic fleet models (Berman and Larson, 1985).

At the same time, public-sector decision environments impose additional requirements of transparency and accountability, motivating the use of interpretable models for high-stakes asset decisions (Rudin, 2019).

Recent methodological advances address these challenges by embedding predictive models directly within optimization problems, aligning learning with downstream decision quality rather than statistical accuracy alone (Donti et al., 2017; Elmachtoub and Grigas, 2022).

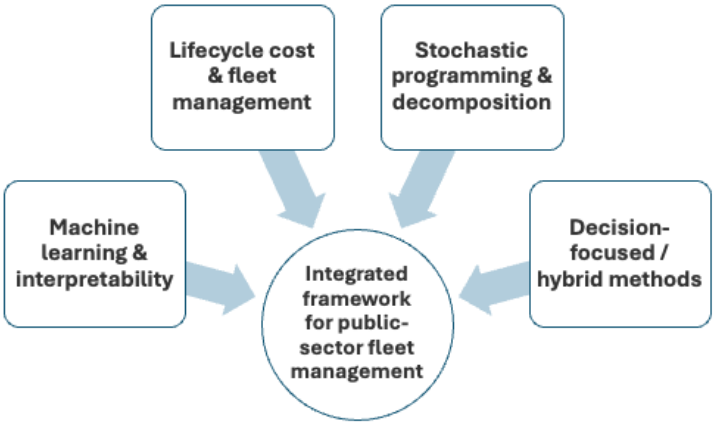


Figure 2.1: Conceptual positioning of the dissertation within four major research streams: lifecycle cost and fleet management, stochastic programming and decomposition, machine learning and interpretability, and decision-focused hybrid methods.

This hybridization of machine learning and optimization reflects a broader trend documented across operations research and artificial intelligence, where learning-augmented optimization is increasingly used to support complex, data-driven decision systems (Bengio et al., 2021).

2.2 Fleet Management and Lifecycle Cost Modeling

Lifecycle cost analysis (LCC) provides the economic foundation for rational fleet replacement decisions by jointly accounting for acquisition cost, maintenance escalation, operating expense, downtime risk, and residual value. Classical replacement theory and subsequent fleet applications show that optimal replacement timing emerges from the interaction of these cost components rather than from fixed age or mileage rules (Jardine and Tsang, 2021; Hartman, 2004).

Most LCC frameworks treat maintenance cost as an age- and usage-dependent process, often exhibiting nonlinear acceleration at later asset ages. Recent studies reinforce the importance of stochastic and nonlinear effects in fleet cost behavior. Electrification-focused analyses show that maintenance and energy cost patterns vary substantially across powertrains, duty cycles, and operating environments, producing cost trajectories that differ markedly from those of conventional vehicles (Ritari et al., 2020; Frieß and Pferschy, 2024). Related work on alternative-fuel and emission-focused fleet planning highlights long-term cost uncertainty and environmental externalities (Ercan et al., 2015; Quarles et al., 2020).

Government fleets differ from commercial fleets in two important respects. First, statutory replacement rules and procurement cycles constrain flexibility. Second, public oversight requires defensible and well-documented capital decision processes. The NCHRP synthesis on fleet replacement practices (National Cooperative Highway Research Program, 2019) documents wide variation in replacement rules, limited use of data-driven methods, and a need for standardized, economically grounded approaches—motivating the analytical framework developed in this dissertation.

Table 2.1 summarizes representative studies in lifecycle cost modeling and fleet replacement. Collectively, these works demonstrate that economically efficient replacement timing depends on the joint evolution of acquisition, operating, maintenance, and residual costs, and that uniform age-based rules often fail to reflect heterogeneous duty cycles and asset use in public-sector fleets.

2.3 Stochastic Programming and Decomposition Methods

Two-stage stochastic programming provides a canonical framework for modeling commitment–recourse decisions under uncertainty, in which first-stage decisions are

Table 2.1: Selected studies on lifecycle cost modeling and fleet replacement.

Study	Context	Method	Key Insight
Jardine and Tsang (2021)	General asset management	Maintenance and replacement theory	Formalizes economic life as a balance between operating cost growth and replacement cost under uncertainty.
Hartman (2004)	Industrial assets	Replacement optimization	Shows optimal replacement timing depends on cost growth and utilization rather than fixed age rules.
Ercan et al. (2015)	Transit bus fleets	Lifecycle cost and emissions analysis	Demonstrates how powertrain and duty cycle alter lifecycle cost and optimal replacement timing.
Quarles et al. (2020)	Public-sector fleets	Lifecycle cost evaluation	Shows large variation in cost trajectories across fleet technologies and operating contexts.
Frieß and Pferschy (2024)	Public transit systems	Mixed-fleet optimization	Highlights uncertainty and heterogeneity in zero-emission fleet lifecycle cost.
National Cooperative Highway Research Program (2019)	State DOT fleets	Survey and practice review	Documents wide variation in replacement rules and limited use of data-driven lifecycle analysis.

made before uncertainty is realized and second-stage actions adapt to scenario outcomes (Shapiro et al., 2009). In transportation and logistics, stochastic programming has been applied to fleet sizing, dynamic allocation, network design, and emergency response planning (Powell, 1986; Santoso et al., 2005; Wang et al., 2022). In these settings, decisions trade inexpensive but inflexible commitments against expensive but flexible recourse.

Decomposition methods are essential for solving such models at scale. Key families include:

- **Benders Decomposition (BD)**, which exploits separability between first-stage decisions and scenario subproblems (Benders, 1962; Rahmaniani et al., 2017);
- **Lagrangian Relaxation (LR)**, which dualizes coupling constraints to obtain separable scenario problems and strong dual bounds (Geoffrion, 1974; Guignard, 2003);
- **Progressive Hedging (PH)**, which enforces nonanticipativity through quadratic penalties and supports scenario parallelism (Rockafellar and Wets, 1991; Watson et al., 2012).

Sampling-based approaches and integer extensions, including the integer L-shaped method and sample average approximation (SAA), further enable large stochastic fleet models (Laporte and Louveaux, 1993; Kleywegt et al., 2002). Recent advances emphasize cut strengthening, stabilization, and adaptive selection strategies (Rahmaniani et al., 2017; Ramírez-Pico et al., 2023), underscoring the continued relevance of decomposition methods for computationally intensive fleet planning. Table 2.2 situates stochastic programming and decomposition methods alongside interpretable machine learning and decision-focused optimization, highlighting how these literature streams collectively inform the dissertation's integrated framework. These works motivate and support the use of scenario-based modeling and decomposition methods for large-scale public-sector fleet problems, where deterministic formulations quickly become computationally intractable.

Table 2.2: Summary of literature streams, research gaps, and dissertation contributions.

Domain	Representative Literature	Identified Gaps	Dissertation Contribution
Stochastic Optimization and Decomposition	(Benders, 1962; Geoffrion, 1974; Birge and Louveaux, 2011; Shapiro et al., 2009; Watson et al., 2012)	Predominantly theoretical or industrial applications; limited empirical validation, institutional applicability, and public-sector fleet focus.	A calibrated two-stage stochastic fleet model evaluated using Benders decomposition, Lagrangian relaxation, and Progressive Hedging, emphasizing interpretability, policy relevance, and preparedness under uncertainty.
Interpretable Machine Learning for Cost Modeling	(Breiman, 2001; Friedman, 2001; Ke et al., 2017; Lundberg and Lee, 2017; Lundberg et al., 2020; Ribeiro et al., 2016; Murdoch et al., 2019)	Emphasis on predictive accuracy over interpretability and policy alignment; limited application to lifecycle cost modeling in public-sector fleets.	Transparent Random Forest and LightGBM models with SHAP- and LIME-based explanations that quantify nonlinear lifecycle cost drivers, ownership effects, and organizational heterogeneity.
Decision-Focused / Hybrid Predict–Then–Optimize	(Elmachtoub and Grigas, 2022; Ban and Rudin, 2019; Bertsimas and Kallus, 2020)	Few empirical integrations of prediction with stochastic optimization; limited evidence in public-sector decision contexts.	A Decision–Coupled Learning framework embedding equivalent annual cost objectives into predictive training, aligning model accuracy with economic performance and risk-aware fleet policy outcomes.

2.4 Machine Learning for Cost Modeling and Interpretability

Machine learning (ML) provides flexible, nonparametric models capable of capturing nonlinear cost trajectories, heterogeneous usage patterns, and organizational variation. Tree-based ensemble methods—including gradient boosting and related approaches—perform well in high-variance, high-dimensional fleet cost data (Friedman, 2001; Chen and Guestrin, 2016; Natekin and Knoll, 2013; Ke et al., 2017).

However, predictive accuracy alone is insufficient in public-sector settings. This concern is especially acute in high-stakes public-sector applications, where post-hoc explanations of opaque models may be insufficient for accountability, motivating the use of inherently interpretable or transparently explainable approaches (Rudin, 2019).

Agencies must justify replacement decisions to auditors, legislators, and other stakeholders. SHAP values provide additive, locally consistent explanations of complex models, enabling transparent decomposition of cost predictions into age, utilization, ownership, and organizational effects (Lundberg and Lee, 2017; Lundberg et al., 2020; Murdoch et al., 2019).

Validation for dependent and hierarchical fleet data requires special care. VIN-safe and agency-blocked validation schemes prevent information leakage across multi-year vehicle histories and organizational units, producing realistic estimates of predictive performance (Varma and Simon, 2006; Roberts et al., 2017).

Table 2.2 summarizes key contributions in machine learning, explainability, and decision-focused optimization that inform the hybrid framework developed later in the dissertation. These studies reflect a shift away from purely predictive accuracy toward models that support transparent, decision-relevant analytics in high-stakes settings.

2.5 Decision-Focused and Hybrid Learning for Prescriptive Optimization

Traditional predict–then–optimize workflows assume that improved predictive accuracy yields better decisions. In practice, statistical loss functions such as mean squared error are weak proxies for economic decision quality, and small prediction errors near decision thresholds can induce large policy changes (Elmachtoub and Grigas, 2022). This concern is particularly salient in public-sector decision contexts, where opaque models may be unacceptable regardless of predictive performance, reinforcing the need for interpretable or optimization-aligned learning approaches (Rudin, 2019).

Decision-focused learning embeds optimization objectives directly into model training, aligning learning with downstream economic outcomes. Approaches include optimization-aware loss functions, predictive prescriptions, learning-augmented stochastic programming, and surrogate-based methods for mixed-integer optimization (Bertsimas and Kallus, 2020; Elmachtoub and Grigas, 2022; Sarayloo et al., 2023; Triantafyllou and Papathanasiou, 2024). These ideas motivate the hybrid Decision–Coupled Learning framework developed in Chapter 6.

2.6 Empirical Context and Data Challenges

The dissertation draws on a 16-year dataset (2009–2025) covering more than 6,000 state fleet assets across multiple vehicle classes. This empirical setting introduces several methodological challenges:

- **Censoring and ownership imbalance**, with leased vehicles rarely exceeding five years of age and owned vehicles extending beyond 15 years;
- **High variance across agencies** due to heterogeneous duty cycles and maintenance practices;

- **Temporal dependence**, requiring VIN-safe validation;
- **Non-stationarity**, including supply-chain shocks and inflation.

These complexities justify the combined use of interpretable ML, stochastic programming, and hybrid decision-focused learning.

2.7 Comparative Limitations of Existing Approaches

Although lifecycle cost modeling, stochastic programming, machine learning, and decision-focused learning each provide powerful analytical tools, they remain largely siloed in the literature. Understanding their comparative strengths and limitations clarifies the need for an integrated framework.

2.7.1 Deterministic Lifecycle Cost Models

Classical lifecycle cost (LCC) models provide transparent and economically grounded replacement rules based on expected cost trajectories. These models offer interpretability and institutional defensibility; however, they typically assume deterministic cost evolution and homogeneous operating conditions. Such assumptions limit their ability to accommodate volatility in maintenance, downtime shocks, salvage uncertainty, or budget constraints. As a result, deterministic LCC approaches may produce economically appealing replacement ages that are not robust under uncertainty.

2.7.2 Pure Stochastic Optimization Models

Two-stage and multistage stochastic programming models address uncertainty explicitly by modeling commitment and recourse decisions across scenario sets. These formulations provide risk-aware capacity planning and formally capture nonanticipativity constraints. However, many applications rely on stylized or parametric cost inputs rather than empirically calibrated lifecycle behavior. Without data-driven

cost estimation, stochastic models may misrepresent the underlying maintenance and utilization dynamics of heterogeneous fleets.

2.7.3 Pure Predictive Machine Learning Models

Machine learning approaches excel at capturing nonlinear and high-dimensional relationships in operational data. In fleet applications, tree-based ensemble methods can uncover heterogeneous aging effects, organizational variation, and ownership structure impacts. Yet predictive accuracy alone does not ensure economic optimality. Models trained to minimize statistical loss functions (e.g., mean squared error) may still induce suboptimal replacement policies if errors are concentrated near decision thresholds or if cost variance is mischaracterized. In public-sector contexts, reliance on predictive accuracy without explicit policy linkage risks producing analytically sophisticated but operationally misaligned recommendations.

2.7.4 Decision-Focused and Prescriptive Learning

Recent advances in decision-focused learning seek to bridge prediction and optimization by embedding downstream objectives into model training. These approaches align learning with economic regret rather than predictive error. However, most studies remain methodological, focusing on stylized optimization problems or fully differentiable settings. Empirical implementations in large, public-sector, mixed-integer decision environments remain limited. Moreover, few studies integrate interpretable predictive diagnostics with scenario-based stochastic programming in a unified decision system.

2.7.5 Implications for Public-Sector Fleet Management

Public fleet systems operate at the intersection of these limitations. Deterministic lifecycle rules lack uncertainty modeling; stochastic programs often lack empirical calibration; predictive models lack economic alignment; and hybrid methods rarely

address interpretability and institutional constraints simultaneously. This fragmentation motivates the integrated framework developed in this dissertation, which explicitly links empirical lifecycle behavior, interpretable machine learning, stochastic optimization, and decision-coupled evaluation within a transparent and auditable structure.

2.8 Synthesis and Research Positioning

While lifecycle cost modeling, stochastic programming, and machine learning have each matured substantially, their integration for public-sector fleet management remains limited. Existing studies typically address these components in isolation—either focusing on deterministic lifecycle analysis, optimization under stylized uncertainty, or predictive modeling divorced from downstream decision consequences. As a result, agencies often lack analytically coherent tools that connect observed cost behavior to uncertainty-aware and policy-relevant replacement decisions.

This dissertation addresses these gaps through four complementary contributions:

- developing and systematically comparing decomposition-based stochastic optimization models for fleet planning under demand, cost, and recourse uncertainty (Chapter 3);
- constructing interpretable machine learning models that correct for censoring, information leakage, and nonlinear lifecycle cost dynamics, while remaining transparent and auditable for public-sector use (Chapter 4);
- empirically identifying class-specific diminishing-returns ages in fleet lifecycle costs using nonparametric smoothing of long-horizon operational data, providing transparent and data-driven economic-life benchmarks that bridge descriptive cost analysis and replacement policy design (Chapter 5);

- introducing a hybrid Decision–Coupled Learning framework that aligns predictive modeling with economic decision quality by embedding optimization objectives directly into the learning process (Chapter 6).

Together, these contributions establish a transparent, uncertainty-aware, and empirically grounded analytical framework for public-sector fleet replacement and procurement planning. The framework explicitly links observed lifecycle cost behavior, predictive modeling, and prescriptive optimization, providing a coherent foundation for evidence-based and institutionally defensible fleet policy in large government fleets.

Chapter 3

Stochastic Optimization for Fleet Planning under Uncertainty

Public-sector fleets operate in environments where procurement decisions must be made months or years in advance, long before workload requirements, vehicle utilization, or maintenance needs are fully known. Short-term remedies—such as emergency leases or spot purchases—are often substantially more expensive, creating a core commitment–recourse trade-off. Two-stage stochastic programming provides a natural structure for capturing this imbalance: first-stage ownership decisions must be selected before uncertainty is realized, while second-stage recourse actions absorb unexpected shortfalls (Shapiro et al., 2009; King and Wallace, 2012; Powell and Topaloglu, 2005). Early OR work on fleet planning emphasized capacity sizing and renewal decisions under demand uncertainty (Berman and Larson, 1985), establishing a foundation that was later extended through stochastic programming and multi-period replacement models. Stochastic fleet planning models have appeared across logistics, transportation, and asset renewal settings (Santoso et al., 2005; Escudero et al., 2007; Pantuso, 2016; Wang et al., 2018; Khalili-Damghani et al., 2022). Demand and fuel-price variability can meaningfully shift optimal replacement and procurement timing (Zheng and Chen, 2016); maintenance uncertainty influences reserve fleet

sizing and idle-capacity trade-offs (Raposo et al., 2018); and renewal planning that jointly considers economic cost and environmental performance emphasizes the multi-objective nature of long-horizon fleet decisions (Ercan et al., 2015; Frieß and Pferschy, 2024). More broadly, stochastic network and supply chain design studies (Govindan et al., 2017; Eskandarpour et al., 2015; Wang et al., 2022) reinforce the value of explicitly modeling uncertainty in capital-intensive infrastructure and asset decisions.

This chapter presents three classical decomposition approaches—Benders Decomposition (BD), Lagrangian Relaxation (LR), and Progressive Hedging (PH)— and reports their computational performance on a stylized but realistic state-agency fleet instance. The focus is both algorithmic and managerial: how quickly and accurately each method converges, and what insights the resulting policies offer regarding preparedness, flexibility, and cost-efficiency.

3.1 Structural Derivation of the Stochastic Fleet Model

This chapter formulates public-sector fleet replacement as a two-stage stochastic programming problem. Before presenting the mathematical formulation, it is useful to derive the model structure intuitively from the institutional and economic characteristics of fleet management.

3.1.1 Commitment and Recourse Structure

Fleet replacement decisions exhibit natural temporal asymmetry. Capital decisions—such as purchasing or retaining vehicles—must be made before future operating conditions are known. Once a fiscal year begins, maintenance realizations, downtime events, and demand levels unfold stochastically.

This timing structure creates a classic commitment–recourse setting:

- **First-stage decisions** determine owned fleet capacity and replacement actions prior to observing uncertainty.
- **Second-stage decisions** adjust operations after uncertainty is realized, potentially through short-term leasing, emergency procurement, or reallocating idle units.

The economic trade-off is fundamental: committing too little capacity increases recourse cost under high-demand or high-failure scenarios, while committing too much capacity increases fixed ownership cost and idle risk.

3.1.2 Sources of Uncertainty

Uncertainty in fleet systems arises from multiple interacting sources:

1. **Maintenance cost escalation**: aging vehicles exhibit nonlinear and heterogeneous repair behavior.
2. **Downtime events**: failure realizations affect service availability.
3. **Demand variability**: agency mission intensity varies across time and regions.
4. **Residual value and capital cost fluctuation**.

These uncertainties are not merely noise around a deterministic average. They directly alter the marginal economic value of replacement and capacity decisions.

3.1.3 Why Deterministic Models Are Insufficient

Deterministic lifecycle cost models implicitly replace $\mathbb{E}[C(x, \xi)]$ with $C(x, \bar{\xi})$ for some average realization $\bar{\xi}$. This approximation neglects asymmetry in recourse cost: underestimating demand or failure risk can be significantly more expensive than overestimating it.

For public agencies operating under procurement delays and capital rigidity, the cost of being underprepared often dominates the cost of modest over-preparedness. This asymmetry motivates explicit modeling of scenario-based uncertainty.

3.1.4 Resulting Two-Stage Structure

These structural features lead naturally to a two-stage stochastic program of the form:

$$\min_{x} \; c^{\text{own}} x + \mathbb{E}_{s \in S}\big[Q(x, \xi_s)\big],$$

where first-stage decisions x determine owned capacity and $Q(x, \xi_s)$ represents scenario-dependent recourse cost under realization ξ_s.

The recourse function captures emergency procurement, downtime penalties, and demand shortfalls, linking long-term ownership decisions to short-term operational outcomes.

3.1.5 Implications for Decomposition

The scenario separability of the recourse function implies that, conditional on first-stage decisions, each scenario can be evaluated independently. This structure motivates decomposition approaches such as Benders Decomposition, Lagrangian Relaxation, and Progressive Hedging.

Thus, decomposition methods are not introduced as algorithmic afterthoughts; they emerge directly from the economic separability inherent in the fleet planning problem.

The formal mathematical model presented in the following section makes these relationships explicit.

3.2 Two-Stage Stochastic Programming Preliminaries

Before presenting the calibrated fleet model, it is useful to recall the canonical form of a two-stage stochastic program and the structures exploited by decomposition methods.

Let $x \in X \subseteq \mathbb{Z}^{n_1} \times \mathbb{R}^{n_2}$ denote first-stage (here-and-now) decisions, and let $\tilde{\xi}$ denote a random vector capturing uncertain parameters such as demand, costs, or failure rates. A generic two-stage stochastic mixed-integer program can be written as (Shapiro et al., 2009; King and Wallace, 2012):

$$\min_{x \in X} \left\{ c^\top x + \mathbb{E}_{\tilde{\xi}} \left[Q(x, \tilde{\xi}) \right] \right\}. \tag{3.1}$$

For the test instance used in the method comparison, the first-stage and recourse models are linear. Accordingly, the deterministic equivalent reduces to a (potentially large) but highly structured linear program.

For a linear recourse structure, the scenario-specific problem has the form

$$\begin{aligned} Q(x, \xi) \;=\; \min_{y} \quad & q(\xi)^\top y && (3.2) \\ \text{s.t.} \quad & T(\xi)x + W(\xi)y \;\geq\; h(\xi), && (3.3) \\ & y \in Y(\xi), && (3.4) \end{aligned}$$

where y denotes second-stage (recourse) decisions, $q(\xi)$ represents scenario-dependent costs, and the matrices $T(\xi)$ and $W(\xi)$ couple first- and second-stage decisions through the constraints. The feasible set $Y(\xi)$ may include integrality, capacity, or sign restrictions.

In practice, the expectation in (3.1) is approximated via a finite set of scenarios $s \in S$ with probabilities p_s, leading to the sample average approximation (SAA):

$$\min_{x \in X} \left\{ c^\top x + \sum_{s \in S} p_s \, Q(x, \xi_s) \right\}. \tag{3.5}$$

Theoretical and computational properties of SAA for discrete stochastic optimization are well documented in Kleywegt et al. (2002) and Shapiro et al. (2009).

When the second-stage problem is a linear program and x is fixed, each recourse problem admits a dual formulation. For a given scenario s with data (q_s, T_s, W_s, h_s),

the primal recourse problem

$$Q(x, \xi_s) = \min_{y} \quad q_s^\top y \tag{3.6}$$

$$\text{s.t.} \quad T_s x + W_s y \geq h_s, \tag{3.7}$$

$$y \geq 0 \tag{3.8}$$

has the dual

$$Q(x, \xi_s) = \max_{\pi_s} \quad (h_s - T_s x)^\top \pi_s \tag{3.9}$$

$$\text{s.t.} \quad W_s^\top \pi_s \leq q_s, \tag{3.10}$$

$$\pi_s \geq 0, \tag{3.11}$$

where π_s are the dual variables associated with the recourse constraints. Expressions of the form (3.9) underpin Benders Decomposition: for any fixed x, dual optimal solutions generate supporting hyperplanes (Benders cuts) that approximate the expected recourse cost in the master problem (Benders, 1962; Rahmaniani et al., 2017). The fleet-specific model introduced in Section 3.3 is a direct instantiation of this two-stage structure, with ownership expansions as first-stage variables and emergency procurement as linear recourse.

3.3 Model

This section formalizes a one-period, two-stage stochastic test instance for fleet planning. The agency selects first-stage ownership expansions before demand is known; after demand is realized, any remaining shortfall is covered by expensive emergency procurement. The objective is to minimize expected total cost while maintaining feasibility under all demand scenarios.

3.3.1 Sets and Indices

- V: set of vehicle types

$$V = \{\text{sedans},\ \text{pickups},\ \text{vans},\ \text{SUVs}\}.$$

- S: set of demand scenarios

$$S = \{\text{above_avg},\ \text{average},\ \text{below_avg}\}$$

All scenarios are treated as equiprobable, with $p_s = 1/3$ for each $s \in S$. This mirrors a simple SAA structure with three representative demand regimes.

3.3.2 Parameters

- x_v^0 — initial owned fleet of type v.
- c_v^{own} — annual ownership and operating cost per unit of type v.
- c_v^{buy} — emergency purchase or short-term lease cost per unit.
- c_v^{idle} — idle cost per unit of unused owned capacity of type v.
- $d_{v,s}$ — demand for type v in scenario s.
- p_s — probability of scenario s.

3.3.3 Calibration Values Used in Experiments

Initial fleet.

$$x^0 = \{\text{sedans: } 375,\ \text{pickups: } 650,\ \text{vans: } 350,\ \text{SUVs: } 550\}.$$

Annual ownership cost.

$$c^{\text{own}} = (1850,\ 2200,\ 2000,\ 2000).$$

Emergency purchase cost.

$$c^{\text{buy}} = (30000,\ 50000,\ 45000,\ 55000).$$

Idle carrying cost. Idle cost is modeled as a proportional carrying cost equal to 30% of annual ownership cost:

$$c_v^{\text{idle}} = 0.3\, c_v^{\text{own}}, \qquad \forall v \in V.$$

Demand scenarios.

$$d_{\cdot,\text{above_avg}} = (530,\ 900,\ 560,\ 810),$$
$$d_{\cdot,\text{average}} = (500,\ 875,\ 530,\ 780),$$
$$d_{\cdot,\text{below_avg}} = (470,\ 845,\ 500,\ 750).$$

3.3.4 Decision Variables

- $x_v \in \mathbb{Z}_+$ — additional owned vehicles of type v (first stage).
- $y_{v,s} \geq 0$ — emergency procurement for type v in scenario s (recourse).
- $w_{v,s} \geq 0$ — idle owned vehicles of type v in scenario s.

In practice, vehicle acquisitions are discrete; accordingly, first-stage decisions x_v are modeled as integer. Recourse variables $y_{v,s}$ can likewise be restricted to integers; however, they are treated as continuous here to preserve a transparent linear recourse structure and to isolate decomposition behavior in a controlled testbed.

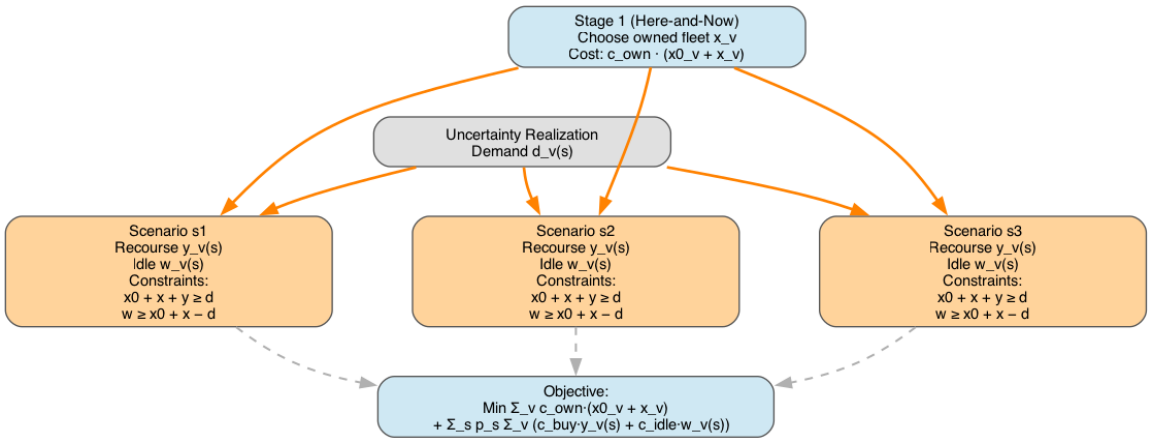


Figure 3.1: Structure of the two-stage fleet model. First-stage ownership decisions are made before demand is realized; scenario-specific recourse actions ensure feasibility under each demand outcome.

3.3.5 Two-Stage Stochastic Program

$$\min_{x,y,w} \quad \sum_{v\in V} c_v^{\text{own}}(x_v^0 + x_v) + \sum_{s\in S} p_s \sum_{v\in V} \left(c_v^{\text{buy}} y_{v,s} + c_v^{\text{idle}} w_{v,s} \right) \tag{3.12}$$

$$\text{s.t.} \quad x_v^0 + x_v + y_{v,s} - w_{v,s} = d_{v,s}, \qquad \forall v \in V,\ \forall s \in S, \tag{3.13}$$

$$x_v \in \mathbb{Z}_+, \qquad \forall v \in V, \tag{3.14}$$

$$y_{v,s} \geq 0,\ w_{v,s} \geq 0, \qquad \forall v \in V,\ \forall s \in S. \tag{3.15}$$

Constraint (3.13) enforces a scenario-wise flow balance. If realized demand exceeds owned capacity, recourse $y_{v,s}$ covers the shortfall. If owned capacity exceeds demand, the surplus is captured by $w_{v,s}$ and incurs idle cost. Because both recourse and idle capacity carry positive cost, the model selects the economically optimal hedge between excess ownership and emergency procurement. At optimum, $y_{v,s}$ and $w_{v,s}$ are not simultaneously positive, so the balance constraint induces the intended complementarity between recourse and idle capacity.

3.4 Economic Interpretation of Dual Variables and Recourse Sensitivity

While the two-stage stochastic formulation provides a mathematical representation of fleet replacement decisions under uncertainty, the associated dual variables and recourse sensitivities admit clear economic interpretations that are particularly relevant in public-sector contexts.

3.4.1 Shadow Prices of Capacity and Demand

In each scenario subproblem, constraints linking owned capacity and realized demand generate dual variables that measure the marginal value of relaxing those constraints. Economically, these multipliers represent the *shadow price of insufficient capacity* under a given realization of maintenance, downtime, and demand conditions.

If the dual variable associated with a demand constraint is positive, it indicates that the system is capacity-constrained in that scenario: an additional owned unit would reduce expected recourse cost. Conversely, a zero multiplier suggests slack capacity or over-provisioning.

Thus, dual variables quantify the scenario-specific marginal economic value of fleet expansion decisions.

3.4.2 Recourse Function as a Value-of-Flexibility Measure

The recourse function

$$Q(x, \xi_s)$$

captures the minimum second-stage cost given first-stage decisions x under scenario ξ_s. Its slope with respect to x reflects how sensitive operational cost is to ownership decisions.

Steep recourse gradients indicate that under-investment in capacity leads to expensive emergency procurement or downtime penalties. Flatter gradients suggest diminishing marginal benefit from additional owned units.

From a policy perspective, the curvature of the expected recourse function reveals the economic trade-off between preparedness and idle capacity.

3.4.3 Interpretation of Benders Cuts

In Benders decomposition, optimality cuts are constructed using dual solutions of the scenario subproblems. Each cut can be interpreted as a linear approximation of the expected recourse cost around a given first-stage decision.

Economically, these cuts represent accumulated knowledge about the marginal cost of under-capacity across observed scenarios. As cuts are added, the master problem internalizes increasingly refined estimates of the true expected recourse surface.

This process has a natural managerial interpretation: the model progressively "learns" the cost of insufficient preparation.

3.4.4 Lagrangian Multipliers and Coordination Penalties

In Lagrangian relaxation and Progressive Hedging approaches, multipliers associated with nonanticipativity constraints measure the degree of disagreement across scenario-specific first-stage decisions.

Large multipliers signal that scenarios favor materially different capacity levels, reflecting high uncertainty or asymmetric cost exposure. Small multipliers indicate structural stability in replacement decisions across scenarios.

Thus, these values provide insight into how sensitive replacement policy is to uncertainty realizations.

3.4.5 Policy-Relevant Sensitivity Insights

Beyond algorithmic utility, dual variables and recourse sensitivities offer diagnostic value:

- identifying vehicle classes where capacity shortages are most costly,
- quantifying the marginal benefit of incremental budget increases,
- detecting regimes where replacement timing is highly sensitive to uncertainty.

These interpretations reinforce that stochastic optimization is not merely a computational device. It produces economically meaningful shadow prices and policy signals that support transparent justification of fleet decisions.

3.5 Positioning within Stochastic Fleet and Capacity Planning

Two-stage stochastic programming has long been used to model advance capacity commitment with costly recourse under demand uncertainty (Santoso et al., 2005; Escudero et al., 2007; Wallace and Fleten, 2003). In these settings, acquisition decisions are capital-intensive and discrete, while recourse actions—such as emergency procurement or reallocation—provide flexibility at higher marginal cost. Deterministic equivalents grow rapidly with scenario count, motivating decomposition strategies that exploit scenario separability (Powell and Topaloglu, 2005; Mulvey and Ruszczy'nski, 1995; Watson et al., 2012).

In fleet and renewal contexts, uncertainty in demand, fuel price, and deterioration materially alters optimal replacement timing and acquisition rates (Zheng and Chen, 2016; Raposo et al., 2018). Environmental and technology-transition studies further emphasize that long-horizon fleet planning is inherently multi-objective, balancing economic cost, reliability, and emissions considerations (Ercan et al., 2015; Frieß and Pferschy, 2024).

Risk-averse stochastic models incorporate measures such as CVaR to protect against unfavorable realizations (Noyan, 2012, 2016). Related fleet size and mix studies further highlight how managerial priorities influence optimal composition decisions (Sarangi et al., 2023).

However, relatively few studies provide transparent, side-by-side evaluation of classical decomposition approaches within calibrated public-sector fleet instances.

This chapter contributes to this literature by isolating the structural commitment–recourse trade-off in a calibrated state-agency fleet setting and systematically comparing Benders Decomposition, Lagrangian Relaxation, and Progressive Hedging. The emphasis is not on new algorithmic theory, but on interpretability, convergence behavior, and economically meaningful policy signals in a public-sector decision environment.

3.6 Methods

Three decomposition strategies are evaluated on the model in Section 3.3: Benders Decomposition (BD), Lagrangian Relaxation (LR), and Progressive Hedging (PH). Each exploits scenario separability in a distinct way:

- **Benders Decomposition (BD)** partitions the problem into a master problem over first-stage decisions and scenario-specific recourse subproblems, generating cuts that approximate the expected recourse cost (Benders, 1962; Rahmaniani et al., 2017; Laporte and Louveaux, 1993).
- **Lagrangian Relaxation (LR)** dualizes the coupling constraints to obtain fully separable scenario-level subproblems, updated iteratively through multipliers (Geoffrion, 1974; Fisher, 1981; Guignard, 2003).
- **Progressive Hedging (PH)** enforces nonanticipativity through quadratic penalties, allowing scenarios to be solved independently while driving first-stage decisions toward consensus (Rockafellar and Wets, 1991; Watson et al., 2012).

Algorithm 1: Benders Decomposition for the Fleet Model

Initialize: $LB \leftarrow -\infty$, $UB \leftarrow +\infty$, $k \leftarrow 0$. Initialize the master problem with first-stage variables x, recourse proxy θ, and no Benders cuts.

Repeat until $UB - LB \leq \varepsilon$**:**

(1) Solve the master problem to obtain $x^{(k)}$ and master objective $z_M^{(k)}$.

(2) For each scenario $s \in S$: Solve the recourse LP at $x^{(k)}$ to obtain $Q_s(x^{(k)})$ and an optimal dual solution $\pi_s^{(k)}$. Add the Benders optimality cut:

$$\theta \;\geq\; \sum_{s \in S} p_s \left(h_s^\top \pi_s^{(k)} - (T_s x)^\top \pi_s^{(k)} \right).$$

(3) Update bounds:

$$LB \leftarrow z_M^{(k)}, \qquad UB \leftarrow \sum_{v \in V} c_v^{\text{own}} (x_v^0 + x_v^{(k)}) + \sum_{s \in S} p_s Q_s(x^{(k)}).$$

(4) Set $k \leftarrow k + 1$.

Return: $x^\star$ and expected cost UB.

Logic-based Benders variants (Hooker and Ottosson, 2003; Hooker, 2007) and learning-augmented decomposition methods (Sarayloo et al., 2023) provide additional perspectives on how structural information and predictive signals can improve large-scale stochastic optimization; these ideas motivate the hybrid framework in Chapter 6. When first-stage decisions are integer-valued, classical Benders decomposition may suffer from weak cuts and slow convergence. Combinatorial and strengthened cut strategies have therefore been proposed to improve performance in mixed-integer settings (Laporte and Louveaux, 1993; Codato and Fischetti, 2006; Rahmaniani et al., 2017).

Algorithms 1–3 summarize the three decomposition methods specialized to the fleet planning problem. In the implementation used here, all master and scenario subproblems are solved with a commercial optimization solver (LP/MILP), and convergence is monitored via global bounds (BD and LR) or nonanticipativity residuals (PH). The algebraic formulations and algorithmic details follow the

Algorithm 2: Lagrangian Relaxation for the Fleet Model

Initialize: multipliers $\lambda_s^{(0)}$ for all $s \in S$, step sizes $\{\alpha_k\}$, and $k \leftarrow 0$.

Repeat until convergence:

(1) For each scenario $s \in S$: solve the Lagrangian-relaxed subproblem with multipliers $\lambda_s^{(k)}$ to obtain $x_s^{(k)}$ (and recourse variables if needed).

(2) Compute consensus:

$$\bar{x}^{(k)} \;=\; \sum_{s \in S} p_s x_s^{(k)}.$$

(3) Evaluate the Lagrangian dual bound $LB^{(k)}$.

(4) Update multipliers for each scenario:

$$\lambda_s^{(k+1)} \;=\; \lambda_s^{(k)} + \alpha_k \left(x_s^{(k)} - \bar{x}^{(k)} \right).$$

(5) Set $k \leftarrow k + 1$.

Return: consensus solution $\bar{x}$ and best dual bound $\max_k LB^{(k)}$.

Algorithm 3: Progressive Hedging for the Fleet Model

Initialize: Solve each scenario independently to obtain $x_s^{(0)}$ for all $s \in S$. Set $\bar{x}^{(0)} = \sum_{s \in S} p_s x_s^{(0)}$, $\lambda_s^{(0)} = 0$, and $k \leftarrow 0$.

Repeat until $\sum_{s \in S} p_s \| x_s^{(k)} - \bar{x}^{(k)} \|^2 \leq \varepsilon$**:**

(1) For each scenario $s \in S$, solve the penalized subproblem:

$$\min_{x_s, y_s} \; f_s(x_s, y_s) + \lambda_s^{(k)\top} x_s + \frac{\rho}{2} \| x_s - \bar{x}^{(k)} \|^2,$$

to obtain $x_s^{(k+1)}$.

(2) Update consensus:

$$\bar{x}^{(k+1)} = \sum_{s \in S} p_s x_s^{(k+1)}.$$

(3) Update multipliers:

$$\lambda_s^{(k+1)} = \lambda_s^{(k)} + \rho \left(x_s^{(k+1)} - \bar{x}^{(k+1)} \right).$$

(4) Set $k \leftarrow k + 1$.

Return: consensus first-stage decision $\bar{x}$.

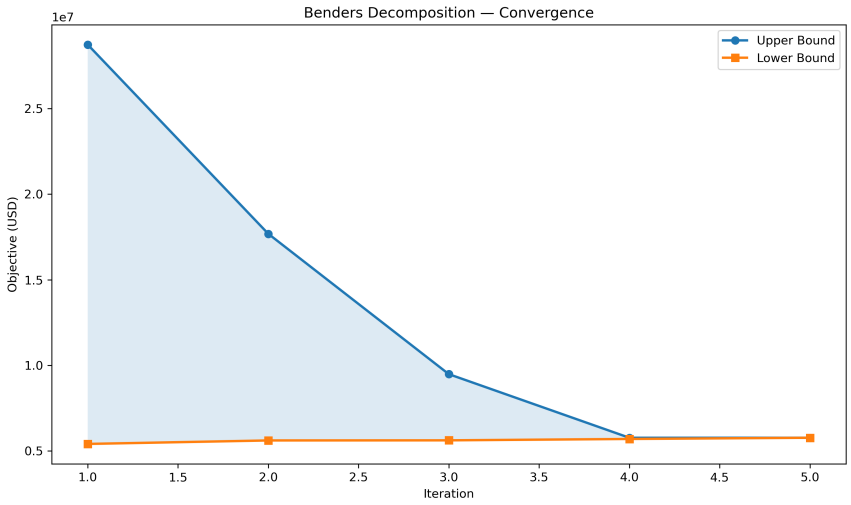


Figure 3.2: Benders convergence: lower and upper bounds tighten monotonically and meet at \$5,770,750 after five iterations.

Table 3.1: Excess owned capacity (owned fleet minus scenario demand) under the optimal expansion.

Scenario	Sedans	Pickups	Vans	SUVs
Above Avg	0	0	0	0
Average	30	25	30	30
Below Avg	60	55	60	60

canonical treatments in Benders (1962); Rahmaniani et al. (2017); Geoffrion (1974); Fisher (1981); Rockafellar and Wets (1991); Watson et al. (2012).

3.7 Computational Results

This section presents the numerical performance of the decomposition methods on the four-vehicle-type, three-scenario test instance. All three methods are applied to the same calibrated data; only the solution strategy differs.

All three methods recover the same optimal first-stage expansion,

$$\Delta x^* = (155,\ 250,\ 210,\ 260),$$

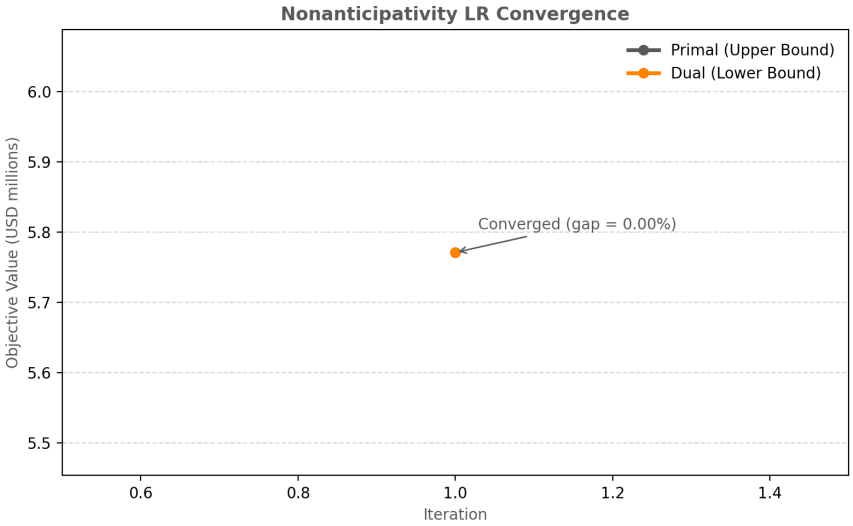


Figure 3.3: Lagrangian Relaxation convergence: the dual bound matches the optimal objective on this instance after one multiplier update.

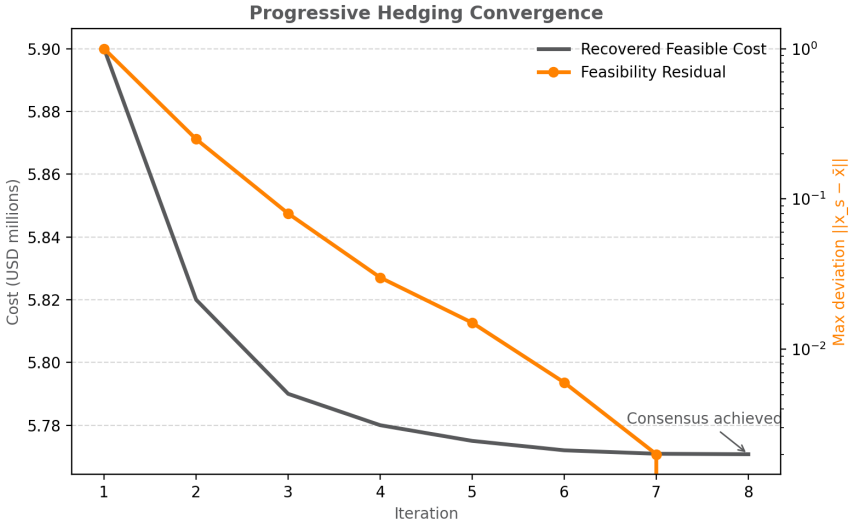


Figure 3.4: Progressive Hedging convergence: scenario decisions align within eight iterations as nonanticipativity residuals fall below the stopping tolerance.

Table 3.2: Comparison of decomposition methods on the test instance.

Metric	BD	LR	PH
Iterations	5	1	8
Optimal cost	\$5,770,750	same	same
Optimal expansion	(155, 250, 210, 260)	same	same
Recourse purchases	0	0	0

Notes: The expansion vector corresponds to (sedans, pickups, vans, SUVs).

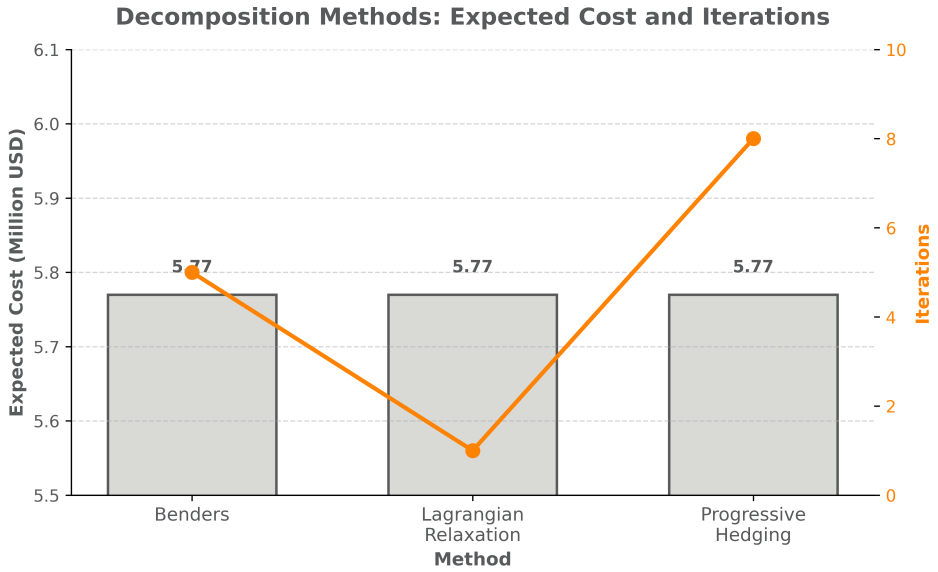


Figure 3.5: Comparison of the three decomposition methods on the calibrated fleet instance. All methods identify the same optimal expansion plan and expected cost.

which brings the total owned fleet exactly to the above-average demand level in each vehicle class. Thus, emergency procurement is unnecessary in all scenarios, and expected cost is limited to ownership costs. In the average and below-average scenarios, the excess capacity appears as idle units, summarized in Table 3.1. Although this instance is intentionally compact, it provides a controlled benchmark for interpreting convergence behavior before introducing vehicle-level heterogeneity and data-driven cost uncertainty in later chapters.

3.8 Managerial Insights

The decomposition results reveal a consistent strategic pattern in the agency's optimal fleet decision. All three methods—BD, LR, and PH—select the same expansion plan, ensuring that even the highest-demand scenario requires no emergency procurement. This alignment across methods underscores the robustness of the recommended policy rather than any dependence on a particular algorithm.

Several managerial insights emerge:

- **Preparedness versus flexibility.** The optimal policy deliberately *over-provisions* capacity relative to average and below-average demand. Modest idle capacity is economically justified to avoid the much higher cost of emergency recourse. The agency effectively treats the above-average scenario as a design point.
- **Economic value of avoiding recourse.** Because emergency procurement is substantially more expensive than standard ownership, the model consistently selects first-stage expansions that eliminate recourse in all scenarios. This behavior reflects a "risk premium" on emergency purchasing, a pattern often observed in public-sector fleet policy and in other capital-intensive systems where unmet demand is heavily penalized.
- **Differential value across vehicle types.** Pickups, SUVs, and vans receive proportionally larger expansions than sedans, mirroring their higher peak-demand levels. The decomposition methods help quantify the marginal value of capacity by vehicle type, which is useful for category-specific budgeting and replacement planning.
- **Method selection for policy analysis.** The three methods provide complementary diagnostic features:
  - **BD** yields interpretable duals and explicit lower/upper bounds, supporting auditability and transparent documentation.

  - **LR** produces strong dual bounds quickly and is well-suited for rapid what-if and sensitivity analyses on larger variants.
  - **PH** offers straightforward scenario parallelism and residual-based convergence, which can be attractive when strict optimality certificates are less critical than feasibility and robustness.
- **Scalability and practical readiness.** Although the test instance is intentionally compact, the convergence behavior and diagnostics illustrate which approaches are most promising when additional vehicle classes, cost components, and temporal structure are introduced.

Collectively, these insights emphasize that decomposition methods do more than solve the optimization problem: they also reveal economic and operational logic that supports budget justification and long-term planning in public agencies.

3.9 Discussion and Conclusion

The decomposition experiments on the four-type, three-scenario fleet instance highlight the shared foundations and differing algorithmic behaviors of three classical stochastic programming methods. Despite relying on distinct mechanisms—cuts, multipliers, or quadratic penalties—all three methods converge to the same ownership expansion vector and expected cost. This reinforces a reassuring conclusion: *the fleet recommendation is structural, not an artifact of the solution technique.*

At the same time, each method provides distinct diagnostic information that is valuable for analysts and policymakers:

- **Benders Decomposition** produces an informative sequence of bounds and cuts that quantify the marginal value of first-stage capacity, making the role of uncertainty in shadow prices visible over iterations.

- **Lagrangian Relaxation** (nonanticipativity-based) achieves extremely fast convergence on this instance and reveals the strength of the coupling constraints across scenarios, suggesting its value for larger sensitivity studies.
- **Progressive Hedging** illustrates scenario-wise behavior over iterations, showing how local decisions reconcile into a unified first-stage plan. Its parallel structure makes it a practical candidate for scaling to larger scenario sets or multi-period formulations.

From a fleet-management standpoint, the optimal solution has a clear interpretation: the agency expands ownership just enough to cover peak demand without relying on emergency procurement, accepting idle capacity in lower demand scenarios as the cost of preparedness. This mirrors cost–service trade-offs observed in public-sector asset and capacity planning, where modest over-capacity can be economically efficient when unmet demand is heavily penalized and service reliability is paramount (Powell, 1986; National Cooperative Highway Research Program, 2019).

Beyond the specifics of this stylized instance, the chapter establishes a transparent benchmark that supports the dissertation's subsequent chapters:

- The interpretable ML models developed in Chapter 4 provide granular, data-driven estimates of lifecycle cost and usage heterogeneity that can enrich the stochastic model.
- The hybrid decision-focused learning framework in Chapter 6 builds directly on both components, using predictive signals and optimization structure jointly to improve policy performance.
- The decomposition methods evaluated here provide a methodological backbone for multi-period, multi-agency, or electrification-sensitive extensions, including risk-averse and distributionally robust variants (Bertsimas et al., 2011b; Wang et al., 2025).

This chapter compares classical decomposition methods within a controlled stochastic fleet planning setting, establishing a methodological baseline for subsequent, data-driven extensions.

Natural directions for research include multi-year fleet renewal with vehicle-level age dynamics, integration across multiple agencies or regions, data-driven scenario generation from predictive models, and hybrid algorithmic strategies that combine elements of BD, LR, and PH. These extensions bridge classical stochastic optimization with the interpretable machine learning and hybrid approaches developed in the later chapters.

Chapter 4

Machine Learning and Explainable Cost Modeling

Building on the stochastic optimization framework developed in Chapter 3, this chapter constructs the predictive and interpretability layer required for heterogeneity-aware fleet planning. Whereas the two-stage model formalizes the commitment–recourse trade-off under uncertainty, it requires accurate and structurally meaningful cost estimates as inputs. Machine learning provides these estimates at the vehicle and agency level, capturing nonlinear age effects, ownership interactions, and organizational heterogeneity that cannot be represented through average cost assumptions.

Lifecycle cost in public-sector fleets is neither linear nor stationary. Maintenance exposure, utilization intensity, ownership structure, and agency-specific practices interact in complex ways. Ignoring this structure can distort replacement timing and budget allocation. Accordingly, the goal of this chapter is not merely predictive accuracy, but predictive structure: a model that captures systematic cost drivers while remaining interpretable and auditable.

In public-sector decision environments, interpretability is essential. Analytic outputs must be transparent, defensible, and aligned with governance requirements (Molnar, 2022; Murdoch et al., 2019). Practice syntheses further emphasize that

agency rules and institutional context shape fleet composition and replacement behavior, underscoring the need for tools that support—rather than replace—managerial judgment (National Cooperative Highway Research Program, 2019).

Following established definitions and diagnostic principles for interpretable machine learning (Murdoch et al., 2019; Molnar, 2022; Chen et al., 2022), this chapter adopts SHAP-based explanations for tree ensemble models as a transparent and theoretically grounded approach to decomposing lifecycle cost predictions (Lundberg et al., 2020). These explanations provide the interpretive bridge between statistical learning and operational fleet policy.

4.1 Data and Feature Preparation

The analysis draws on administrative, operational, and cost records from the state's enterprise fleet system, covering fiscal years 2015–2025. These sources were consolidated into a unified vehicle-level panel suitable for both predictive modeling and integration with the stochastic optimization framework. The empirical setting parallels prior lifecycle cost studies in public-sector and transit systems that emphasize operating cost, replacement timing, and institutional constraints rather than acquisition price alone (National Cooperative Highway Research Program, 2019).

4.1.1 Fleet Dataset Overview

Table 6.1 summarizes fleet composition by ownership type and vehicle category. The fleet comprises over 6,300 vehicles distributed across multiple agencies and duty classes. Light-duty sedans and SUVs are predominantly leased, while heavy-duty, police, and specialized assets are almost entirely owned. This structural heterogeneity implies that replacement policy cannot be uniform: ownership structure, duty class, and utilization must be modeled jointly.

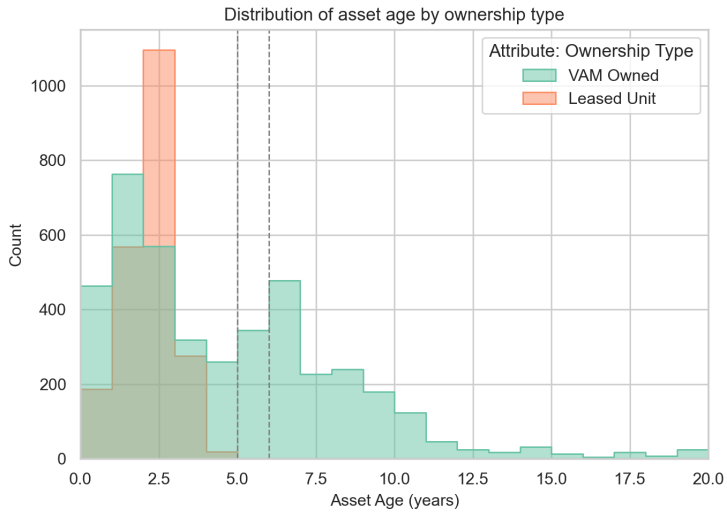


Figure 4.1: Distribution of asset age by ownership type. Leased vehicles are concentrated in early service years due to contractual turnover and programmatic replacement policies, whereas owned vehicles extend to substantially higher ages. Vertical dashed lines indicate the 4–7 year policy window commonly associated with mid-life replacement discussions.

Figure 4.1 illustrates the age distribution by ownership type. Leased vehicles are concentrated in early service years due to contractual turnover and programmatic replacement policies, whereas owned vehicles extend into substantially higher ages. This differential age support creates a form of observational censoring: certain ownership–age combinations are sparse or absent in the historical data. As a result, descriptive averages may conflate policy rules with underlying cost structure. Predictive modeling is therefore required to estimate lifecycle cost trajectories while accounting for uneven empirical support.

4.2 Data Pipeline and Preprocessing

The machine learning workflow integrates preprocessing, feature engineering, model training, validation, and SHAP-based explainability within a unified pipeline (Figure 4.2). All processing was implemented in Python using `pandas` and `scikit-learn`, with ensemble tree-based models serving as the primary predictive architecture due

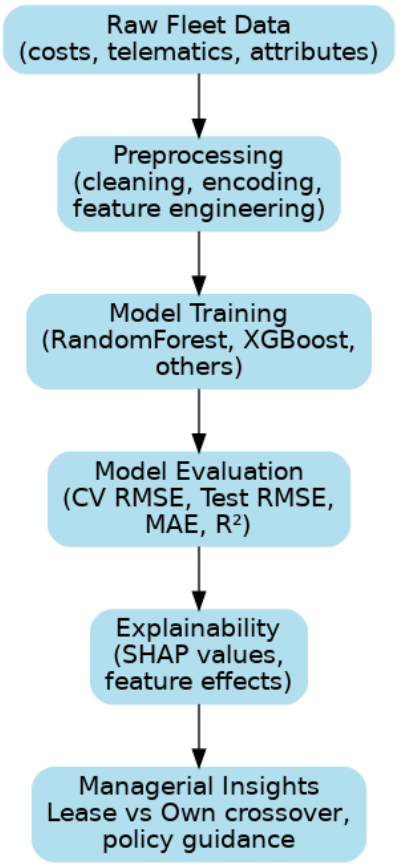


Figure 4.2: Machine learning workflow for lifecycle cost modeling, from raw data to predictive models and explainable outputs. The pipeline enforces leakage-safe validation and supports direct integration with the stochastic optimization model.

to their ability to capture nonlinearities and high-order interactions in heterogeneous lifecycle cost data (Natekin and Knoll, 2013).

4.2.1 Features

Table 6.2 lists the predictors retained in the analytic dataset. These variables span vehicle characteristics, organizational context, and operational exposure, enabling the model to capture nonlinear cost progression and ownership–age interactions.

Additional utilization and cost components (e.g., mileage, fuel cost, repair cost) are incorporated in specific model variants used for sensitivity analysis but are omitted from the core feature table for brevity.

Table 4.1: Key features retained for predictive and optimization modeling.

Feature	Type	Description
Asset Age	Continuous	Years since acquisition
Days in Service	Continuous	Cumulative operational exposure
Dormant Duration (Days)	Continuous	Time unassigned or inactive
Ownership Type	Binary	Owned vs. leased indicator
Vehicle Type	Categorical	Vehicle class (sedan, pickup, SUV, etc.)
Asset Type	Categorical	Functional duty classification
Make / Manufacturer	Categorical	Vehicle identity attributes (high cardinality)
Agency	Categorical	Managing agency
Using Dept	Categorical	Department-level assignment context
Grand Division	Categorical	Regional classification within the state

4.2.2 Target Variable

The primary target variable, *Total Cost w/o Dep*, measures annual operating and maintenance cost excluding depreciation. This focuses on the controllable portion of lifecycle cost most relevant for replacement and budgeting decisions, consistent with total cost of ownership (TCO) frameworks (Ercan et al., 2015; Quarles et al., 2020).

Because maintenance cost exhibits heavy-tailed and heteroskedastic behavior, models were trained both on the original dollar scale and on the log-transformed scale, log(cost). Log-target training improves variance stabilization and generalization while predictions are evaluated and interpreted on the original dollar scale.

4.2.3 Preprocessing and Validation Design

Preprocessing steps include:

- standardization of continuous variables;
- encoding of categorical variables (one-hot or target encoding depending on cardinality);

- integration within a single pipeline object to ensure identical transformations across cross-validation folds and hold-out testing.

These validation safeguards follow best practices for dependent and hierarchical datasets (Varma and Simon, 2006; Roberts et al., 2017). Together, they ensure that reported performance reflects realistic forecasting constraints (predicting future years and unseen agencies) rather than optimistic reuse of information across time or organizational units.

4.2.4 Validation Strategy and Leakage Prevention

Fleet cost data exhibits both temporal dependence and hierarchical structure. Vehicles generate repeated observations across years, and naive random splits can introduce severe information leakage by allowing future service history to inform earlier predictions.

To preserve chronological and structural integrity, validation folds are constructed to ensure:

- all observations from a given vehicle (VIN) remain within a single fold;
- training data temporally precede validation data;
- agency-level blocking is evaluated to assess cross-organizational generalization.

These safeguards align with theoretical results demonstrating that standard cross-validation under dependence yields optimistically biased error estimates (Arlot and Celisse, 2010; Varma and Simon, 2006; Roberts et al., 2017).

4.3 Why Tree-Based Ensembles? Methodological Rationale

Lifecycle cost data exhibit pronounced nonlinearities and interaction effects:

- maintenance expenditures often increase nonlinearly at age or mileage thresholds;

- utilization interacts with age and ownership structure in shaping cost exposure;
- organizational effects (agency, division) shift baseline cost levels;
- observational censoring differs systematically between leased and owned assets.

Linear or low-order parametric models require extensive manual interaction specification to approximate these structures and remain sensitive to functional-form misspecification. Tree-based ensemble models are well suited to this setting because they flexibly capture nonlinear effects, high-order interactions, and heterogeneous response patterns without imposing global smoothness or additivity assumptions (Natekin and Knoll, 2013).

In addition, lifecycle maintenance cost displays heavy-tailed and heteroskedastic behavior. Training on the log-transformed target log(cost) stabilizes variance and improves generalization in the presence of upper-tail risk. Under VIN-safe cross-validation and leakage-free hold-out evaluation, variance-stabilized gradient boosting models achieved superior predictive performance relative to raw-target alternatives.

Among candidate architectures, log-target XGBoost provided the strongest balance of accuracy, calibration, and tail control. In particular, it substantially reduced upper-tail error while preserving interpretability through SHAP-based decomposition. Random Forest and conventional gradient boosting serve as useful benchmarks and diagnostic comparators; however, given the structural properties of the data, the log-target XGBoost model is adopted as the primary predictive engine for subsequent explainability and counterfactual analysis.

4.4 Predictive Modeling Layer

Several ensemble architectures were trained to predict annual lifecycle cost under leakage-safe evaluation. Because maintenance expenditures exhibit heavy-tailed and heteroskedastic behavior, both raw-target and log-transformed (logY) specifications were evaluated. Training on log(cost) stabilizes variance and reduces sensitivity to

rare but extreme repair events, while predictions are interpreted on the original dollar scale via back-transformation.

Mean absolute error (MAE) served as the primary tuning metric due to its robustness to outliers and direct interpretability in monetary terms. Root mean squared error (RMSE) and R^2 were used as complementary diagnostics capturing dispersion and explained variance. Hyperparameters were tuned through grid and Bayesian search procedures consistent with established ensemble modeling practice (Natekin and Knoll, 2013; Molnar, 2022).

Validation splits were constructed to reflect realistic forecasting conditions. Temporal ordering ensured that future service information was never used to predict earlier observations, and all folds were both VIN-safe and agency-respecting (Section 6.3). These safeguards prevent information leakage across vehicles or organizational units and ensure that reported performance reflects operationally meaningful generalization.

4.5 Model Validation and Benchmarking

We benchmarked baseline and ensemble models under strict VIN-safe cross-validation and a temporally ordered, leakage-free hold-out test set.

Because annual maintenance cost exhibits heavy-tailed and heteroskedastic behavior, both raw-target and log-transformed (logY) specifications were evaluated. Training on log(cost) stabilizes variance and improves generalization in skewed cost environments, with predictions evaluated after back-transformation to the original dollar scale.

Table 4.2 summarizes hold-out performance.

The calibration plot in Figure 4.4 indicates strong agreement between predicted and observed cost across most deciles. Mild underprediction is observed in the extreme upper tail, reflecting conservative smoothing of rare, high-cost repair events under

Table 4.2: Leakage-safe hold-out performance (original dollar scale).

Model	RMSE	MAE	R^2
XGBoost (logY)	**19,044.76**	**5,224.41**	**0.764**
GradientBoosting (logY)	21,059.54	5,832.93	0.712
RandomForest	23,560.60	7,350.46	0.639
GradientBoosting	23,247.59	7,129.55	0.649
DecisionTree	27,475.87	8,706.98	0.509
Ridge	37,473.52	14,814.17	0.087

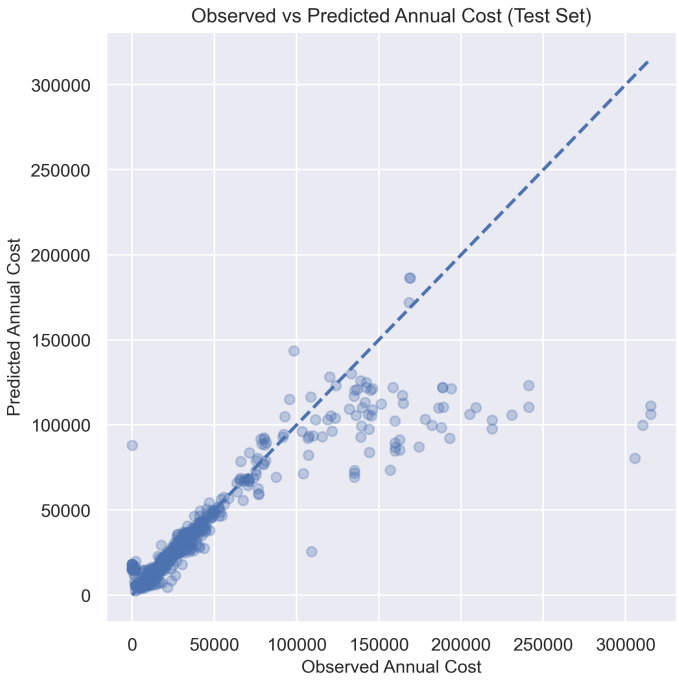


Figure 4.3: Observed versus predicted annual cost on the leakage-free hold-out set using the log-target XGBoost model. Predictions are evaluated on the original dollar scale after back-transformation.

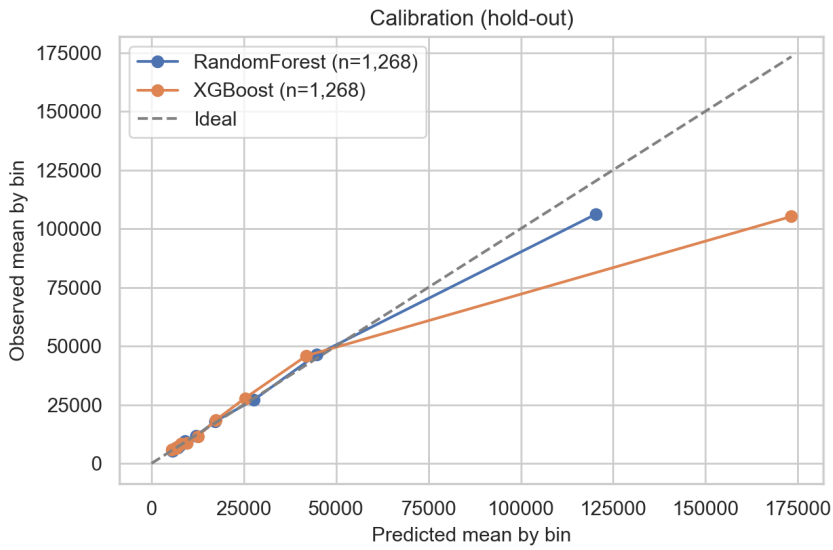


Figure 4.4: Calibration curve by prediction decile for the hold-out set. The log-target XGBoost model remains close to the ideal diagonal across most cost bins, with modest underprediction in the extreme upper tail.

the log-target specification. Importantly, systematic bias is not observed across the central mass of the distribution.

The log-target XGBoost model achieves the strongest predictive performance, with RMSE of \$19,045, MAE of \$5,224, and $R^2 = 0.764$ on the hold-out set. This represents a substantial improvement over raw-target models and indicates that variance-stabilizing transformations materially improve lifecycle cost prediction.

Figure 4.5 examines residual structure across asset age. Error dispersion is largest during the earliest service years, reflecting heteroskedasticity and the influence of rare early-life repair events. Beyond early service, residual variance stabilizes and the LOWESS trend remains centered near zero across mid- and late-life stages, indicating no systematic age-dependent bias in prediction.

The absence of structured residual drift across age is critical for subsequent interpretability analysis. It ensures that SHAP-based lifecycle inference reflects genuine structural effects rather than artifacts of systematic under- or overprediction.

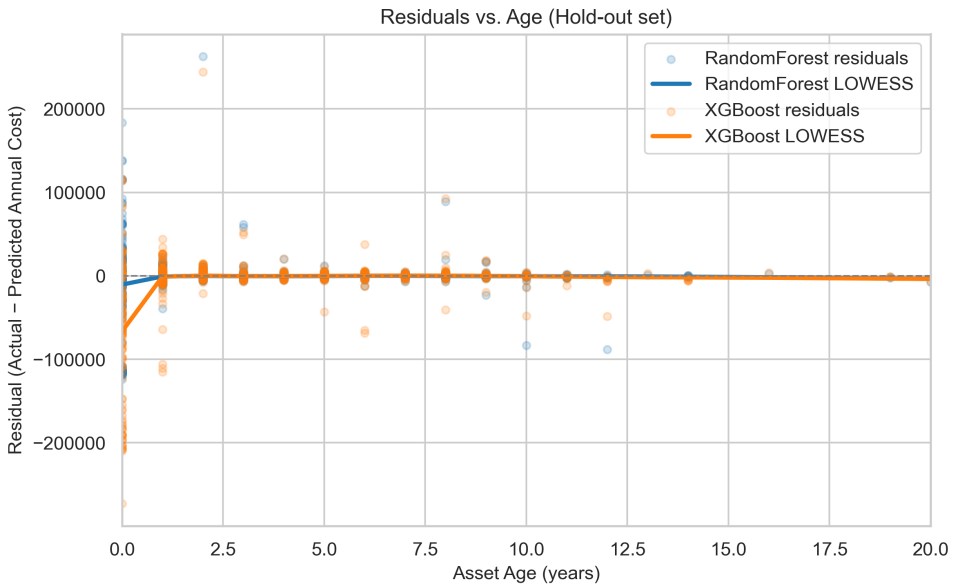


Figure 4.5: Residuals versus asset age on the hold-out set for the log-target XGBoost model. LOWESS smoothing (solid line) highlights lifecycle-dependent dispersion patterns.

Given its superior predictive accuracy, stable calibration, and compatibility with SHAP-based interpretability, the log-target XGBoost model is adopted as the primary predictive engine for subsequent explanation and counterfactual analysis.

4.6 Explainability via SHAP

SHAP values (Lundberg and Lee, 2017; Lundberg et al., 2020) provide an additive, locally consistent decomposition of model predictions derived from cooperative game theory. Each prediction is expressed as a baseline value plus feature-specific contributions, enabling transparent attribution of predicted cost to observable drivers. This framework aligns with public-sector transparency requirements by producing numerically interpretable and auditable explanations, in contrast to heuristic or post-hoc explanation methods (Murdoch et al., 2019; Chen et al., 2022).

Because the predictive model is estimated on log(cost), SHAP values are additive on the log scale. A SHAP shift of δ therefore implies an approximate multiplicative

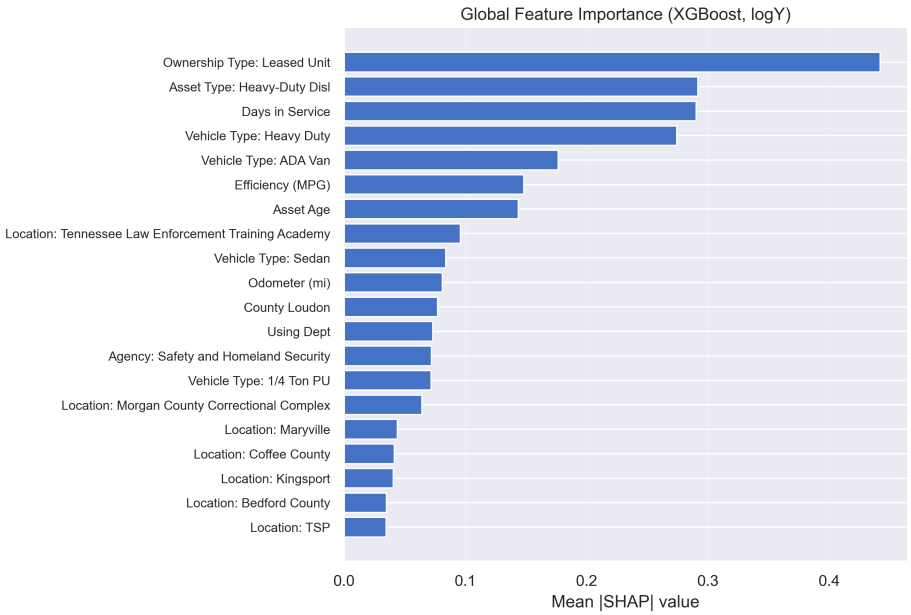


Figure 4.6: Global SHAP feature importance for the log-target XGBoost model. Bars represent mean absolute SHAP value on the log-cost scale, indicating average contribution magnitude to predicted annual cost.

change of $\exp(\delta)$ in predicted dollar cost. Global importance rankings and directional patterns are interpreted on the log scale, while policy-relevant dollar differences are computed using back-transformed model predictions.

4.6.1 Global Insights

Global SHAP feature importance (Figure 4.6) indicates that ownership structure, duty class, and operational exposure are among the strongest determinants of lifecycle cost.

- *Ownership Type (Leased Unit)* exhibits the largest average contribution magnitude, confirming systematic cost differentials between leased and owned vehicles conditional on age, duty class, and utilization.
- *Asset Type and Vehicle Class* (e.g., heavy-duty diesel, ADA van) strongly influence predicted cost levels, reflecting structural maintenance intensity differences across duty categories.

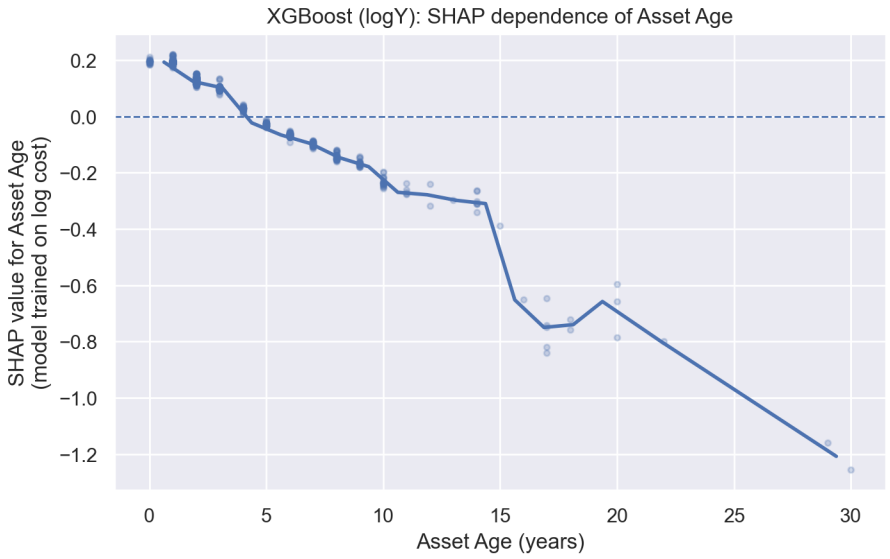


Figure 4.7: SHAP dependence plot for Asset Age under the log-target XGBoost model. Points represent individual observations; the smoothed curve highlights the nonlinear marginal contribution of age to predicted log annual cost.

- *Asset Age* remains a significant lifecycle driver, consistent with established cost-progression and failure-time theory (Ercan et al., 2015), though its effect operates jointly with ownership and utilization rather than dominating independently.
- *Days in Service* captures operational exposure and interacts with age to shape cost trajectories.

Figure 4.7 reveals a nonlinear marginal age effect. Early service years are associated with positive SHAP contributions relative to the model baseline, reflecting elevated cost variability during deployment and warranty transitions. As vehicles transition into mid-life years, the marginal contribution of age to predicted cost declines, suggesting relative stabilization conditional on ownership and utilization.

Beyond approximately 15 years, SHAP values become increasingly negative, suggesting that extreme ages correspond to systematically lower predicted cost relative to the global baseline. This pattern likely reflects selective survival and usage filtering: only particular asset types and agencies retain vehicles into very late service years.

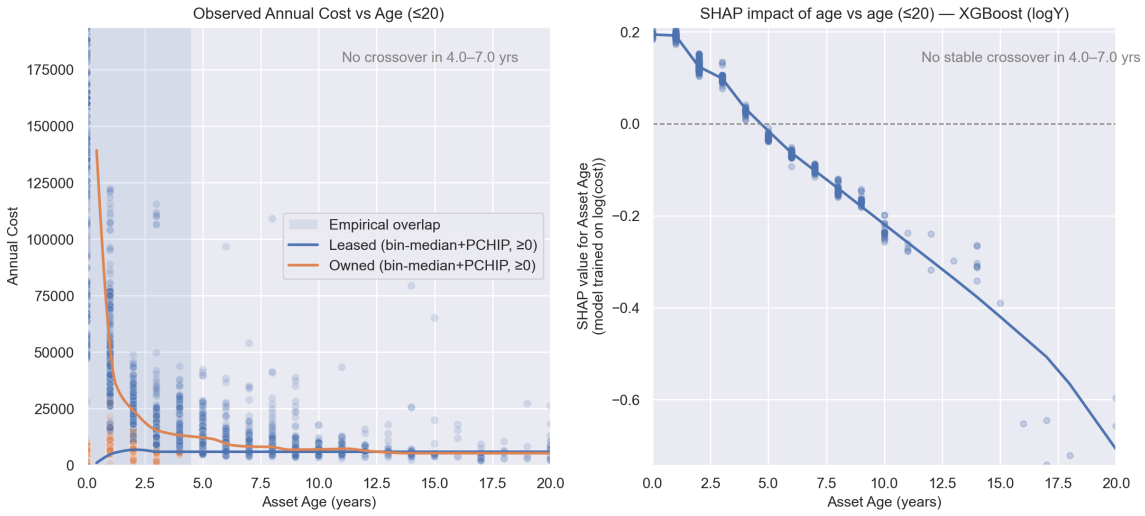


Figure 4.8: Observed annual cost patterns (left) and SHAP-implied model contributions (right) by asset age. The right panel reflects SHAP values from the log-target XGBoost model; additive contributions on the log scale correspond to multiplicative effects in dollar terms. Both panels show no lease–own crossover in the 4–7 year range once empirical support constraints are accounted for, countering naive interpretations based on raw averages.

Figure 4.8 integrates descriptive and model-based evidence. While raw averages may suggest apparent lease–own intersections, these pivots arise from unequal empirical age support and ownership censoring. Restricting attention to the supported age range and examining model-implied marginal contributions reveals no stable crossover in the 4–7 year interval.

This result challenges heuristic replacement rules based solely on age thresholds and reinforces the necessity of model-based lifecycle evaluation that accounts for nonlinear effects and structural ownership differences.

4.6.2 Counterfactual and Age–Ownership Effects

Counterfactual simulations quantify the cost differential between ownership and leasing for identical vehicle characteristics. For each vehicle record, the trained log-target XGBoost model is evaluated twice—once under "owned" and once under "leased"—while holding all other attributes fixed. Predicted log-cost outputs are

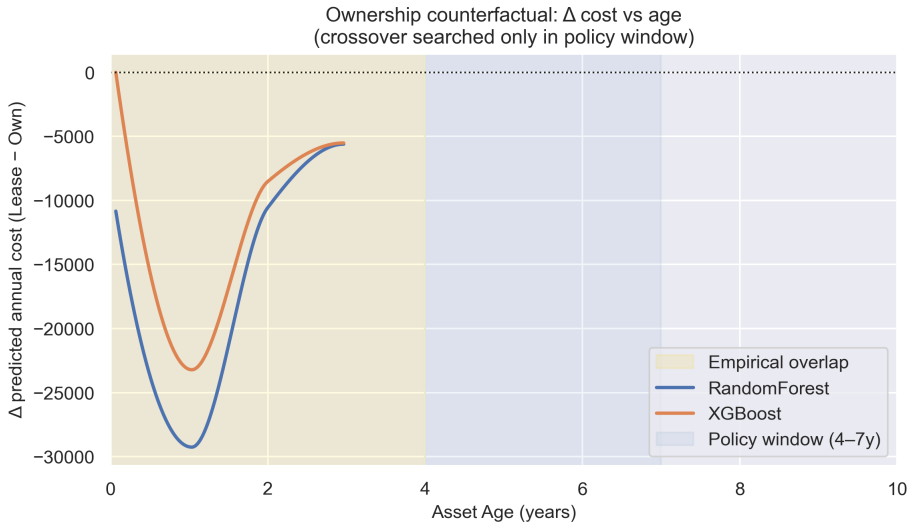


Figure 4.9: Counterfactual ownership–lease cost differential across asset age. Negative values indicate a cost advantage for ownership. The shaded region reflects empirically supported age overlap in the data, and crossover analysis is restricted to the 4–7 year policy window corresponding to practical replacement decision horizons.

then back-transformed to the original dollar scale before computing ownership–lease differences.

Figure 4.9 summarizes the resulting cost differential by asset age. Several patterns emerge:

- Within the empirically supported age range, ownership exhibits a consistent cost advantage relative to leasing once vehicle characteristics and utilization are held fixed.
- No stable crossover is observed within the 4–7 year policy window, indicating that the frequently cited mid-life lease–own pivot is not supported after controlling for censoring and compositional effects.
- The magnitude of the ownership advantage narrows with age but remains economically meaningful within the supported region.

Uncertainty bands in Figure 4.10 visualize 95% confidence intervals for the cost differential, highlighting increasing dispersion at higher ages. This age-dependent

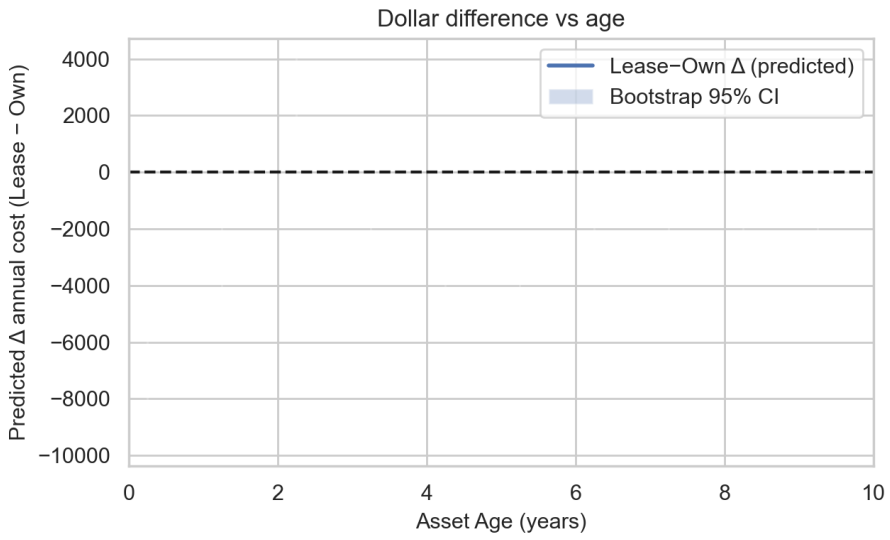


Figure 4.10: Mean and 95% confidence interval of cost difference by age. Predictive uncertainty increases with vehicle age, reinforcing the importance of uncertainty-aware replacement planning.

risk amplification is directly relevant for stochastic programming formulations that incorporate recourse and downside exposure (Noyan, 2012, 2016).

Dollar differences are computed after back-transforming predictions from the log scale, ensuring that counterfactual comparisons reflect expected annual cost on the original monetary scale.

4.6.3 Partial Dependence: Joint Effects

The joint surface in Figure 4.11 highlights the nonlinear and interaction-driven structure of lifecycle cost. Across the empirically supported age range, owned vehicles exhibit systematically higher predicted annual cost than leased vehicles, with the differential remaining negative throughout the policy window (i.e., leased vehicles are consistently less costly than owned vehicles).

While the magnitude of the differential narrows after early service years, no crossover is observed in the 4–7 year decision horizon. This reinforces that ownership

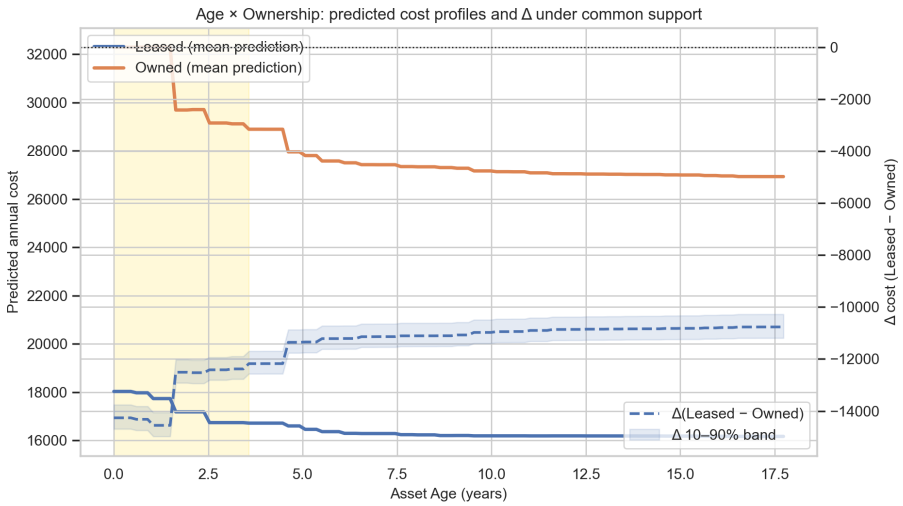


Figure 4.11: Partial dependence of annual cost on asset age and ownership type. Costs rise nonlinearly with age and exhibit distinct ownership effects across the lifecycle.

effects operate jointly with age and utilization rather than through a simple threshold rule.

These nonlinear interaction patterns underscore the importance of embedding predictive structure within downstream optimization models rather than relying on age-only replacement heuristics.

4.7 Summary and Integration with Optimization

This chapter establishes the predictive and interpretive foundation required to embed data-driven cost structure within the stochastic fleet planning framework introduced in Chapter 3. Rather than treating lifecycle cost as an exogenous input, we construct and rigorously validate a heterogeneity-aware predictive model that captures nonlinear age dynamics, ownership interactions, and agency-specific operational effects.

The primary contributions are:

- robust prediction of annual lifecycle cost using ensemble models suited to heavy-tailed and heteroskedastic fleet data;
- transparent decomposition of predicted cost through SHAP-based explanations, enabling auditability and policy interpretability;
- empirical identification of age–ownership interactions and utilization-driven risk patterns that challenge age-only replacement rules;
- validation protocols (VIN-safe and agency-blocked) that reflect realistic forecasting constraints and prevent leakage-driven overstatement of predictive accuracy.

Formally, let $f_{\boldsymbol{\theta}} : \mathcal{X} \to \mathbb{R}$ denote the trained predictive model parameterized by $\boldsymbol{\theta}$. In this setting, the objective is not merely to minimize statistical loss, but to ensure that $f_{\boldsymbol{\theta}}$ produces cost estimates that improve downstream fleet policy decisions.

Accordingly, predictive performance is evaluated in subsequent chapters through its *prescriptive impact* within the stochastic optimization model—measured in terms of expected annual cost, risk exposure, and policy regret—rather than through RMSE or MAE alone.

These results directly enable the hybrid integration developed in Chapter 6, where ML-derived cost structure, SHAP-informed interpretability, and scenario-based recourse modeling are combined within a Decision–Coupled Learning (DCL) framework. This integration links decision-focused learning and predictive–prescriptive analytics to the public-sector fleet context (Elmachtoub and Grigas, 2022; Sarayloo et al., 2023; Triantafyllou and Papathanasiou, 2024).

Chapter 5

Empirical Diminishing Returns in Fleet Lifecycle Costs

Building on the predictive and interpretability framework developed in Chapter 4, this chapter examines a central question in public-sector fleet management: *At what age does continued vehicle operation become economically inefficient?* While earlier chapters focus on prediction and optimization under uncertainty, the present analysis isolates the empirical cost structure itself, identifying the age at which lifetime-average operating cost per year begins to rise.

Public agencies frequently rely on fixed age or mileage thresholds to guide replacement decisions. Although these heuristics offer administrative simplicity, they implicitly assume homogeneous cost behavior across vehicle classes and duty cycles. In contrast, this chapter uses long-horizon operational data to estimate *class-specific diminishing-returns ages*, providing transparent, data-driven reference points for lifecycle policy design, capital budgeting, and downstream optimization models.

5.1 Background and Motivation

Public-sector fleet management has historically relied on heuristic replacement rules, most commonly defined by fixed age or mileage thresholds (Jardine and Tsang, 2021; Hartman, 2004). While such rules are easy to administer and institutionally defensible, they often fail to reflect heterogeneous operating conditions, utilization patterns, and maintenance regimes across vehicle classes.

Classical lifecycle cost theory defines the *economic life* of an asset as the age that minimizes equivalent annual cost (EAC), balancing depreciation, maintenance growth, and salvage value (Cook, 1985; Jardine and Tsang, 2021). Although this framework is well established, public-sector implementations frequently rely on generalized or parametric cost curves that may not reflect realized operating conditions within a specific fleet.

Empirical studies in transportation and equipment management consistently show that operating and maintenance costs accelerate nonlinearly with age, but that the timing of this acceleration varies substantially by vehicle class, duty cycle, and operating environment. This heterogeneity suggests that uniform replacement cycles may induce systematic inefficiencies—overextending high-cost assets while prematurely retiring lower-cost ones.

Recent advances in data availability and interpretable analytics enable a different approach: estimating cost–age relationships directly from observed fleet data using transparent, nonparametric methods. This chapter adopts that perspective, focusing on observable cost inflection points rather than opaque predictive scores or imposed functional forms.

5.2 Diminishing-Returns Age Framework

A central concept in this chapter is the *diminishing-returns age*, defined as the earliest age beyond which average annual operating cost begins to increase. Under this

definition, a diminishing-returns age corresponds to the point at which the lifetime-average operating cost per year reaches its minimum and begins to increase with additional service time. This concept aligns with classical notions of economic life but is estimated empirically rather than derived from assumed cost functions. Unlike classical economic-life calculations that rely on assumed depreciation and salvage profiles, the diminishing-returns age is identified directly from observed operating cost behavior, independent of accounting conventions.

The focus on age-based thresholds reflects the institutional reality of public-sector fleet management. Capital planning, procurement rules, and replacement policies are typically structured around vehicle age rather than continuous usage measures. Identifying economically meaningful age thresholds therefore provides actionable guidance that aligns naturally with existing planning processes. To identify class-specific diminishing-returns ages without imposing a parametric cost form, we apply locally weighted smoothing methods grounded in local polynomial regression theory (Cleveland, 1979; Cleveland and Devlin, 1988; Fan, 1996).

5.3 Data and Descriptive Analysis

5.3.1 Data Sources and Scope

The analysis uses a comprehensive statewide fleet dataset spanning 2009–2025. Data sources include maintenance work orders, fuel and operating expenses, vehicle acquisition records, and surplus disposition outcomes. Vehicles are tracked longitudinally from service entry through disposal when available.

To support class-specific analysis, the study focuses on three dominant light-duty categories: sedans, sport utility vehicles (SUVs), and pickup trucks. These classes collectively represent the majority of the statewide fleet and provide sufficient sample sizes for nonparametric estimation.

5.3.2 Descriptive Cost Patterns

Operating cost exhibits substantial dispersion across vehicles of the same age, reflecting heterogeneity in utilization intensity, maintenance practices, and assignment severity. Summary statistics indicate that average annual operating cost typically declines during early service years, stabilizes during midlife, and increases at later ages—consistent with classical replacement theory but with class-specific timing.

This dispersion motivates the use of smoothing techniques to recover underlying cost–age trends without imposing restrictive parametric assumptions.

5.4 Methods

5.4.1 Annualized Operating Cost

For each vehicle i, cumulative operating and maintenance cost over its observed service life is denoted by C_i, excluding depreciation. Let A_i denote vehicle age in years. The lifetime-average annual operating cost is defined as

$$\bar{C}_i = \frac{C_i}{A_i}.$$

This quantity represents the average cost per year incurred up to age A_i, rather than the marginal cost in a specific year. It aligns with classical economic-life analysis, in which replacement is triggered when cumulative average cost begins to increase.

5.4.2 Vehicle Age Construction

Vehicle age is constructed hierarchically using service-entry dates, sale dates, and model-year fallbacks to ensure robustness to missing or inconsistent records. This approach prioritizes observed service history while maintaining coverage for assets with incomplete administrative data.

5.4.3 Nonparametric Smoothing

The relationship between age and annualized operating cost is estimated using locally weighted regression (LOWESS) (Cleveland, 1979; Cleveland and Devlin, 1988). LOWESS flexibly captures nonlinearities, plateaus, and escalation behavior common in fleet maintenance data without imposing global functional forms.

5.4.4 Tipping-Age Identification

Let $\hat{C}(a)$ denote the LOWESS-smoothed lifetime-average cost curve at age a. The diminishing-returns age τ is defined as

$$\tau = \min \left\{ a \geq 2 : \frac{d\hat{C}(a)}{da} > 0 \right\}.$$

If no positive slope is detected, τ is defined as the global minimum of $\hat{C}(a)$ for $a \geq 2$. In implementation, increases are required to persist over a short age interval to avoid spurious crossings due to local noise.

Sensitivity checks confirm that tipping-age estimates are stable under moderate variation in smoothing bandwidths and minimum-age thresholds.

5.5 Results

5.5.1 Cross-Class Comparison

Figure 5.1 compares smoothed cost–age curves across sedans, SUVs, and pickup trucks. Although all classes exhibit strongly declining average cost per year at early ages, the onset of diminishing returns occurs at different points across vehicle types. As illustrated in Figure 5.1, the diminishing-returns tipping age corresponds to the first sustained increase in lifetime-average operating cost per year for each vehicle class.

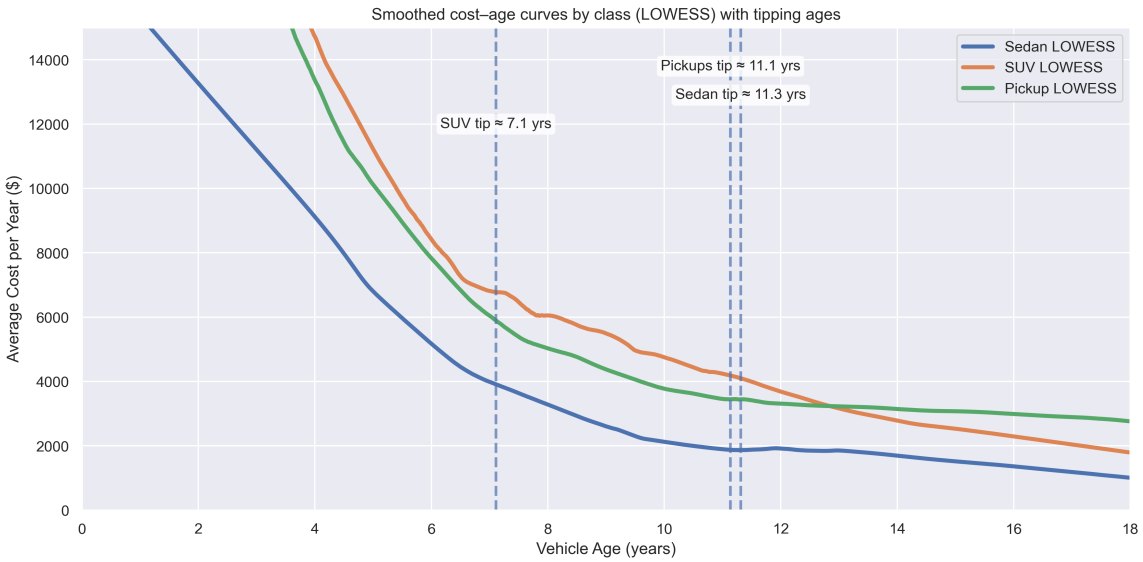


Figure 5.1: Smoothed lifetime-average operating cost per year by vehicle age for sedans, SUVs, and pickup trucks using LOWESS. Dashed vertical lines indicate class-specific diminishing-returns tipping ages. The locations of these tipping points vary substantially across vehicle classes, indicating that uniform replacement cycles are economically inefficient.

Table 5.1: Empirical diminishing-returns results and data coverage by vehicle class.

Class	Tipping Age (yrs)	Avg. Cost at Tip	Usable Vehicles	Usable (%)
Sedans	11.31	$1,867	5,783	97.0%
Pickups	11.14	$3,446	5,072	96.5%
SUVs	7.11	$6,778	4,441	98.6%

SUVs reach a diminishing-returns tipping age earlier (approximately seven years), reflecting earlier maintenance escalation and higher operating burden, whereas sedans and pickup trucks exhibit later tipping points closer to eleven years. These differences imply that a single, uniform replacement cycle would either retire some assets prematurely or retain others beyond their economically efficient service life.

The results reinforce the need for class-specific replacement policies and motivate the hybrid optimization framework developed in subsequent chapters, which explicitly accommodates heterogeneous lifecycle cost trajectories.

Table 5.1 summarizes the estimated diminishing-returns ages alongside data coverage for each vehicle class. Across all three categories, more than 96% of vehicles satisfy validity criteria for age and operating cost, indicating that the estimated tipping points are supported by broad and representative samples rather than narrow subsets of the fleet. Tipping ages are reported on the continuous age grid induced by the smoother; operational replacement policies may round these values to the nearest year.

Observations with zero or negative operating cost—typically arising from accounting adjustments, missing work-order attribution, or atypical service periods—were excluded to prevent distortion of annualized cost measures. No vehicles were excluded due to missing or invalid age values, reflecting the robustness of the hierarchical age construction process.

The high proportion of usable observations across all classes indicates that the estimated tipping ages are not driven by small samples or selective filtering. Instead, they reflect systematic differences in lifecycle cost behavior that persist across thousands of vehicles within each class.

5.5.2 Sedans

Sedans exhibit a long cost plateau, with a clear inflection at approximately 11.3 years, where lifetime-average operating cost per year begins to rise. The corresponding annualized cost at the tipping point is approximately $1,900 per year, indicating that extended service life is economically defensible for this class.

5.5.3 SUVs

SUVs show substantially earlier cost acceleration, with diminishing returns emerging between 7 and 9 years. Annualized operating cost at the tipping point exceeds $6,500 per year, reflecting more severe duty cycles and maintenance intensity.

5.5.4 Pickup Trucks

Pickup trucks display behavior similar to sedans, with cost escalation beginning near 11.1 years at an annualized cost of approximately $3,400. Post-tipping cost growth is steeper than for sedans but remains well below SUV levels.

5.5.5 Cross-Class Interpretation

Across all vehicle classes, initial learning and stabilization effects dominate early service years, while cumulative wear and maintenance complexity drive cost escalation later in life. However, the timing and magnitude of these transitions differ markedly by class, underscoring the inefficiency of uniform replacement rules.

5.6 Policy Implications and Managerial Use

The estimated diminishing-returns ages provide transparent reference points for public-sector fleet policy. Rather than prescribing deterministic replacement rules, these thresholds support prioritization under constrained budgets, ranking vehicles by economic urgency, and justifying capital requests using historical evidence.

Because the thresholds are derived directly from observed operating costs, they are auditable and defensible to oversight bodies. They can be incorporated into rolling capital plans, phased replacement strategies, and scenario-based optimization models.

5.7 Connection to Optimization and Hybrid Models

The diminishing-returns thresholds estimated here are not intended to operate in isolation. Instead, they serve as empirical inputs to the stochastic fleet planning and hybrid decision-focused frameworks developed in subsequent chapters.

In Chapter 6, these empirically estimated cost inflection points are embedded into optimization models that account for uncertainty, budget constraints, and service requirements, enabling prescriptive fleet policies that align predictive insights with economic decision quality.

5.8 Limitations and Future Work

Several limitations suggest directions for future research. First, although hierarchical age construction mitigates missing data, gaps in service-entry dates introduce uncertainty for a subset of assets. Second, annualized cost does not fully normalize for utilization intensity; incorporating mileage, engine hours, or telematics-derived duty-cycle metrics would improve attribution of cost to usage. Third, broad vehicle classes such as SUVs and pickups encompass diverse mission profiles, motivating finer segmentation by specialty use.

Finally, future work can integrate empirical diminishing-returns curves with machine-learning-based forecasting and stochastic optimization to support forward-looking replacement decisions under uncertainty.

5.9 Conclusion

This chapter develops an interpretable, data-driven framework for estimating the economic replacement age of public-sector fleet vehicles. Using nonparametric smoothing of sixteen years of statewide cost data, the analysis identifies class-specific diminishing-returns ages that differ markedly across sedans, SUVs, and pickup trucks.

By grounding lifecycle policy in observable cost behavior rather than fixed heuristics, the proposed framework supports transparent, economically sound fleet management and provides a critical empirical bridge between predictive analytics and optimization-based decision support. The class-specific diminishing-returns ages summarized in Table 5.1 provide empirically grounded reference points for

replacement timing that are used as structural inputs to the optimization models developed in Chapter 6. In the next chapter, these age-based cost inflection points inform scenario construction, penalty calibration, and decision-focused learning objectives, linking observed lifecycle behavior directly to prescriptive fleet policies under uncertainty.

Chapter 6

Hybrid Predict–Then–Optimize Framework

This chapter develops a hybrid predict–then–optimize framework that structurally integrates interpretable machine learning with stochastic fleet planning under uncertainty. The approach follows a modular structure: (i) data integration and preprocessing; (ii) predictive modeling for cost and downtime estimation; (iii) stochastic optimization for replacement and procurement decisions; (iv) hybrid decision-focused learning (DCL) for aligning prediction with economic objectives; and (v) evaluation and sensitivity analysis. Figure 6.1 summarizes these interactions.

6.1 Conceptual Overview

Traditional "predict–then–optimize" approaches train a predictive model using standard statistical losses such as mean squared error and then feed those predictions into an optimizer. However, minimizing prediction error does not guarantee minimal total cost. Decision-focused learning (DFL) modifies the training objective to reflect the downstream optimization loss (Elmachtoub and Grigas, 2022; Ban and Rudin, 2019).

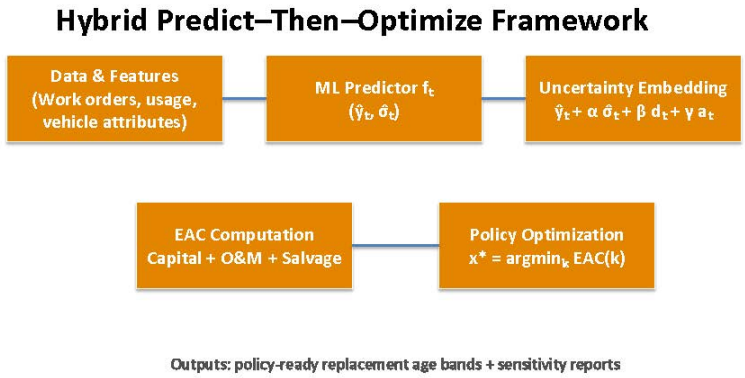


Figure 6.1: Conceptual architecture of the hybrid predict–then–optimize framework linking ML prediction, uncertainty embedding, and policy optimization.

Building on advances in decision-focused learning, the proposed hybrid framework directly couples predictive modeling with stochastic optimization objectives. Rather than minimizing statistical error alone, the proposed approach aligns learning with the ultimate decision cost function, consistent with recent work on decision-focused and task-based learning frameworks (Elmachtoub and Grigas, 2022; Donti et al., 2017). Crucially, this coupling is bidirectional in design: predictive outputs parameterize the stochastic optimization model, while optimization sensitivity provides a principled basis for evaluating and refining the learning objective through decision-focused loss functions.

This integration reflects a broader methodological trend toward unifying machine learning and mathematical optimization, as surveyed by Bengio et al. (Bengio et al., 2021), who synthesize techniques linking combinatorial optimization with modern learning paradigms.

The hybrid framework developed here integrates both paradigms. A transparent, interpretable machine-learning stage produces baseline cost and downtime predictions. These are then refined by an optimization-aware training layer that tunes the model with respect to EAC-based decision loss, balancing accuracy and prescriptive relevance. The framework ensures that auditability and structural interpretability—key requirements for public agencies—are maintained through explainable model structure and auditable optimization logic (Rudin, 2019).

6.2 Data and Feature Preparation

The analysis draws on detailed records from the state's fleet management system, combining cost, utilization, and vehicle attribute data spanning fiscal years 2015–2025. This dataset forms the empirical basis for the hybrid modeling framework introduced in this chapter.

6.2.1 Fleet Dataset Overview

Table 6.1 summarizes the fleet composition by ownership and vehicle category. The dataset covers more than six thousand active and retired units across multiple state agencies. Sedans and pickups together constitute over half of the inventory, reflecting their widespread use in administrative and field operations. Leased assets dominate in light-duty segments (particularly sedans), while heavy-duty and specialized equipment remain almost entirely owned. This heterogeneity in ownership and duty cycle underscores the need for differentiated replacement policies rather than a uniform age or mileage rule. The observed mix also provides sufficient representation of both ownership types for the comparative modeling in this study, supporting robust machine-learning estimation and scenario-based optimization analyses.

The subsequent section outlines the data preparation and preprocessing pipeline used to engineer model-ready features, address missing or inconsistent records,

Table 6.1: Fleet composition by ownership and vehicle category.

Category	Owned (count)	Leased (count)	Total
Sedan	327	1477	1804
SUV	466	262	728
Pickup	1411	406	1817
Van	625	4	629
Police	947	0	947
Heavy Duty	298	0	298
Other	23	0	23
Total	4097	2149	6246

and structure inputs for both the machine learning and stochastic optimization components.

6.3 Data Pipeline and Preprocessing

This section describes the end-to-end data processing workflow used to prepare the fleet dataset for machine learning and optimization analyses.

All preprocessing steps are implemented in Python using `pandas` and `scikit-learn`. Records with missing or inconsistent cost data are removed prior to modeling, and variables are harmonized across source systems to ensure consistent naming conventions and units. The final analytic dataset retains the key explanatory features shown in Table 6.2, including vehicle attributes (*Make, Manufacturer, Asset Type, Vehicle Type*), usage metrics (*Asset Age, Days in Service, Dormant Duration*), organizational factors (*Agency, Using Department, Grand Division*), and ownership descriptors (*Attribute: Ownership Type*).

The target variable, *Total Cost w/o Dep*, represents total lifecycle cost excluding depreciation and serves as the dependent variable in predictive modeling.

The analysis combines administrative cost records, telematics data, and asset master files to construct a multi-year panel of vehicle operations. Key fields include asset identifiers, model year, make, model, agency, utilization measures, downtime

Table 6.2: Key features retained for predictive and optimization modeling.

Feature	Description
Asset Age	Years since vehicle acquisition
Days in Service	Total operational days since deployment
Dormant Duration (Days)	Inactive periods between assignments
Make, Manufacturer	Vehicle identity and production attributes
Asset Type / Vehicle Type	Class and configuration (e.g., sedan, truck, SUV)
Agency / Using Dept	Organizational ownership and user group
Grand Division	Regional administrative classification
Attribute: Ownership Type	Categorical ownership (Owned vs. Leased)

indicators, and detailed cost categories (fuel, repair, preventive maintenance, tires, and labor). These records are merged into a unified analytic dataset using deterministic keys and temporal alignment across fiscal years.

Data cleaning procedures address missingness, outlier detection, and unit standardization to ensure data integrity. Derived features include:

- **Asset Age:** computed using service-entry dates when available, with model-year fallbacks as described in Chapter 5;
- **Utilization Metrics:** annual mileage, operating hours, and stop counts per vehicle;
- **Ownership Type:** categorical indicator distinguishing owned versus leased assets;
- **Downtime Indicators:** total days out of service per year;
- **Regional Controls:** binary indicators for agency and grand division.

Continuous features are standardized using z-score normalization, while categorical features are encoded using one-hot or target encoding depending on cardinality. To avoid temporal leakage, cross-validation follows temporally stratified folds, ensuring that models are trained only on data available prior to the validation period (Arlot and Celisse, 2010; Varma and Simon, 2006; Roberts et al., 2017). This approach preserves the chronological integrity of fleet operations and supports realistic generalization performance estimates.

6.4 Predictive Modeling Layer

This section develops the predictive modeling layer that estimates vehicle lifecycle cost as a function of operational, organizational, and ownership attributes. Each model is trained using the standardized features described in Section 6.3, with the target variable defined as total cost excluding depreciation (*Total Cost w/o Dep*). Model tuning employs grid or Bayesian search over key hyperparameters, using mean absolute error (MAE) as the primary evaluation metric and root mean square error (RMSE) as a secondary criterion.

The objective of this layer is to generate accurate, interpretable cost predictions that can later inform the optimization stage under uncertainty. Tree-based ensemble methods are evaluated, including Random Forests (Breiman, 2001), Gradient Boosting Machines (Friedman, 2001), XGBoost (Chen and Guestrin, 2016), LightGBM (Ke et al., 2017), and CatBoost (Prokhorenkova et al., 2018). These models naturally capture nonlinearities and high-order interactions among vehicle age, utilization, and organizational factors, while accommodating mixed categorical and numerical predictors without restrictive functional assumptions.

Model tuning employs grid or Bayesian search over key hyperparameters, using mean absolute error (MAE) as the primary evaluation metric and root mean square error (RMSE) as a secondary criterion. Temporal cross-validation ensures that model training reflects realistic forecasting conditions, where future costs are never used to predict past outcomes.

Interpretability is addressed through the explainable modeling workflow developed in Chapter 4. In that chapter, model explainability is achieved using SHAP and LIME to quantify both global feature contributions and local prediction behavior (Lundberg and Lee, 2017; Lundberg et al., 2020; Ribeiro et al., 2016; Molnar, 2022). These techniques support visualization of cost sensitivities by asset age, utilization intensity, ownership type, and organizational context, ensuring transparent and auditable interpretation of model outputs.

In the hybrid framework developed here, the stochastic optimization layer consumes the predictive outputs produced by this explainable modeling pipeline, including mean cost estimates and empirically derived uncertainty summaries. No additional explainability computations are required within the optimization routine itself; instead, interpretability is preserved through the structure and diagnostics of the upstream predictive models and the transparent logic of the optimization formulation.

6.5 Stochastic Optimization Model

The second layer of the hybrid framework models the fleet replacement decision under uncertainty in future maintenance cost, downtime, and salvage value. A two-stage stochastic programming formulation is employed to determine the optimal mix of vehicle ownership and replacement actions that minimize expected lifecycle cost:

$$\min_{x,y} \quad \sum_{v\in V} c_v^{\text{own}} x_v + \sum_{s\in S} p_s \sum_{v\in V} \left(c_{v,s}^{\text{maint}} + c_{v,s}^{\text{downtime}}\right) y_{v,s} \tag{6.1}$$

$$\text{s.t.} \quad Ax + By_s \geq d_s, \quad \forall s \in S, \tag{6.2}$$

$$x_v \in \mathcal{X}_v, \quad \forall v \in V, \tag{6.3}$$

$$y_{v,s} \geq 0, \quad \forall v \in V,\ \forall s \in S. \tag{6.4}$$

Here, y_s denotes the vector $(y_{v,s})_{v\in V}$ of recourse decisions under scenario s. The feasible set $\mathcal{X}_v$ reflects vehicle-class–specific decision structure: binary replacement indicators for indivisible assets (e.g., sedans and SUVs), and integer-valued decision variables for pooled or capacity-type assets (e.g., specialized or shared equipment).

The first-stage variables x_v represent ownership and replacement decisions made prior to uncertainty realization, while $y_{v,s}$ denotes second-stage recourse allocations for each scenario $s \in S$. The scenario set S captures stochastic variations in operating

costs, repair needs, and resale outcomes, calibrated from the empirical variance of the machine-learning predictions in Section 6.4. This structure allows the stochastic optimization model to operate on data-driven cost distributions rather than static parameter estimates.

To solve the stochastic replacement problem, the formulation is implemented as a scenario-based mixed-integer optimization model that captures uncertainty in operating costs and capital expenditures. Classical decomposition approaches, including Benders decomposition (Benders, 1962), Lagrangian relaxation (Geoffrion, 1974; Guignard, 2003), and progressive hedging (Watson et al., 2012), provide the theoretical foundation for scalable solution methods in large-scale stochastic programming (Birge and Louveaux, 2011; Shapiro et al., 2009; Rockafellar and Wets, 1991). These methods motivate the structure of the model and inform the design of scenario-based evaluation and sensitivity analysis, even when the full decomposition machinery is not explicitly invoked.

Model calibration uses cost distributions and uncertainty bounds derived from the predictive modeling layer. In this way, the stochastic optimization component is both *data-informed* and *policy-relevant*, translating machine-learning cost predictions into actionable fleet replacement strategies under uncertainty.

6.6 Hybrid Decision-Focused Integration

The hybrid integration layer embeds the optimization objective into the learning process through decision-aware evaluation and policy-level loss metrics. Rather than training the predictive model solely to minimize error between predicted and observed costs, this layer evaluates predictive quality using a surrogate loss that measures deviation in expected annual cost (EAC) from the true or oracle-optimal policy. This approach aligns predictive learning with downstream decision quality rather than pure statistical accuracy.

Formally, let $\boldsymbol{\theta}$ denote the parameters of the predictive model $f_{\boldsymbol{\theta}}(\cdot)$, which produces the cost signal supplied to the stochastic optimization model. Let

$$\mathcal{O}(\cdot)$$

denote the optimization operator that maps predicted cost parameters to an optimal replacement policy under the two-stage fleet model. The induced policy is

$$\pi(f_{\boldsymbol{\theta}}) = \mathcal{O}(f_{\boldsymbol{\theta}})\,.$$

Rather than evaluating predictive performance solely through pointwise statistical loss, the hybrid framework evaluates model quality through *decision regret*:

$$\mathcal{L}_{\mathrm{DCL}}(\boldsymbol{\theta}) = \mathrm{Regret}(\boldsymbol{\theta}) = \mathrm{EAC}\big(\mathcal{O}(f_{\boldsymbol{\theta}})\big) - \mathrm{EAC}\big(\pi_{\mathrm{oracle}}\big),$$

where π_{oracle} is the policy obtained when the optimization model is parameterized by realized costs on the same instance.

In capital-intensive fleet systems, policy regret translates directly into misallocated replacement capital and avoidable operating cost escalation. Under fixed annual budgets, even modest regret compounds across vehicle classes and fiscal cycles. Evaluating predictive models through regret therefore ensures that statistical improvement corresponds to measurable capital efficiency and lifecycle performance gains.

The use of expected annual cost as the decision loss reflects its central role in public-sector capital budgeting and replacement policy, where decisions are evaluated on an annualized, comparable basis across vehicle classes and ownership types.

This alignment ensures that learning updates are most sensitive in regions of the cost surface that influence replacement timing or ownership policy. In contrast to standard mean-squared-error objectives, decision-focused training explicitly reduces *policy regret*—the loss in decision quality due to predictive imperfection—thus

yielding more robust fleet strategies under uncertainty, as measured by reduced policy regret and improved capture of economically dominant replacement decisions (Bertsimas and Kallus, 2020; Bertsimas et al., 2011a; Ban and Rudin, 2019). The resulting model integrates seamlessly with the stochastic optimization layer, enabling iterative improvement of both predictive accuracy and decision performance.

6.6.1 Positioning of the Hybrid Framework within Decision-Focused Learning

Decision-focused learning (DFL) and task-based optimization methods embed optimization objectives directly into predictive model training, often requiring differentiability of the downstream optimization problem (Elmachtoub and Grigas, 2022; Bertsimas and Kallus, 2020). In many applications, the optimization layer is convex and continuously differentiable, enabling end-to-end gradient propagation.

The fleet replacement problem studied here differs structurally. The stochastic planning model contains integer first-stage decisions, scenario-based recourse, and nonanticipativity constraints. These features make full end-to-end differentiation through the mixed-integer program computationally intractable and theoretically non-smooth.

Accordingly, this dissertation adopts a decision-coupled evaluation framework rather than fully differentiable end-to-end training. Predictive models generate cost signals that parameterize the two-stage stochastic optimization model. Model quality is then evaluated through downstream policy regret measured in equivalent annual cost (EAC).

This approach preserves three essential properties:

- **Economic alignment:** predictive performance is judged by its impact on replacement policy quality rather than statistical loss.
- **Computational tractability:** the stochastic program remains solvable using classical decomposition methods without requiring differentiable reformulations.

- **Auditability and transparency:** the optimization model, dual variables, and policy outputs remain interpretable to public-sector stakeholders.

The contribution lies in operationalizing decision-coupled evaluation within a mixed-integer stochastic planning environment characteristic of public-sector fleet systems. This positioning aligns with prescriptive analytics frameworks that emphasize decision impact over pure predictive accuracy (Bertsimas and Kallus, 2020).

6.7 Evaluation Framework

Evaluation of the proposed hybrid framework focuses on decision quality rather than predictive accuracy alone. While standard error metrics are reported as diagnostic checks, the primary emphasis is placed on how predictive uncertainty propagates through the stochastic optimization model and affects replacement policy outcomes.

In the implemented framework, the predictive cost model achieved an out-of-fold mean absolute error (MAE) of approximately $9,700 per vehicle-year under VIN-safe cross-validation. This level of error provides a baseline indication of model fidelity but is not treated as the primary success criterion.

The hybrid framework is evaluated along three complementary dimensions:

1. **Predictive Sanity Checks:** Predictive performance is summarized using out-of-fold mean absolute error (MAE) computed under VIN-safe cross-validation. This metric serves as a diagnostic check to ensure that cost predictions are well-calibrated and free from data leakage, but it is not used as the sole indicator of model quality.
2. **Decision Quality and Policy Regret:** Decision performance is evaluated by embedding predicted costs into a budget-constrained stochastic replacement model and comparing resulting policies against oracle benchmarks constructed using realized costs. Key metrics include total expected policy cost, oracle cost,

absolute regret, and regret expressed as a percentage of oracle-achievable savings. Additional diagnostics include replacement capture rates at varying budget levels and comparisons against top-K heuristic replacement rules.

3. **Policy Diagnostics under Uncertainty:** The stochastic optimization output is evaluated through realized replacement mix, capital expenditure, expected operating cost, and total expected cost. Sensitivity to capital-cost assumptions and vehicle-class constraints is examined to confirm that policy recommendations respond coherently to economic inputs and operational priorities.

Comparative analyses include deterministic ranking heuristics, oracle replacement policies, and the full hybrid ML–stochastic optimization framework. This evaluation structure ensures that improvements in predictive modeling are assessed based on their economic and operational consequences rather than on statistical accuracy alone.

6.7.1 Economic Implications of Decision-Coupled Learning

Traditional predict–then–optimize workflows separate statistical estimation from economic decision-making. Predictive models are trained to minimize average forecast error, and their outputs are subsequently inserted into an optimization model. This separation implicitly assumes that minimizing statistical loss produces economically optimal decisions.

However, in lifecycle replacement settings, the economic objective is nonlinear and threshold-sensitive. Small prediction errors near a replacement boundary can induce materially different retirement ages or capacity decisions. Conversely, larger errors in regions far from the decision threshold may have negligible policy impact.

Decision–Coupled Learning reframes predictive modeling as an input to an economic objective rather than an isolated forecasting task. Instead of minimizing mean squared error (MSE), predictive performance is evaluated through its downstream effect on expected annual cost (EAC), replacement timing stability, and policy regret.

Formally, let $\hat{C}(x)$ denote predicted lifecycle cost and $C(x)$ the realized cost. Traditional learning minimizes:

$$\min \ \mathbb{E}\big[(C - \hat{C})^2\big],$$

whereas decision-focused evaluation emphasizes:

$$\min \ \mathbb{E}\big[\mathrm{EAC}(\hat{x}) - \mathrm{EAC}(x^*)\big],$$

where $\hat{x}$ is the policy induced by predictions and x^* is the policy under perfect information.

In public-sector fleet management, this distinction is consequential. Replacement policies must remain stable across budget cycles and withstand audit scrutiny. Evaluating models by economic regret rather than predictive error ensures that analytical improvement translates directly into improved capital allocation and replacement timing decisions.

Thus, the hybrid framework does not merely couple prediction and optimization computationally; it aligns statistical modeling with institutional economic objectives.

6.8 Sensitivity and Robustness Analyses

Sensitivity analyses systematically vary key economic and policy parameters, including the discount rate, capital replacement costs, and downtime cost multipliers, to assess the stability of replacement decisions under plausible uncertainty. These experiments evaluate how changes in input assumptions propagate through the stochastic optimization model and affect policy outcomes.

Scenario expansion experiments are used to examine performance under higher-uncertainty regimes, consistent with scenario design principles outlined by Rockafellar

and Wets (1991). In these experiments, cost variability is increased to test whether replacement policies remain stable or exhibit sensitivity to tail-risk realizations.

Robustness is further evaluated by adjusting capital-cost assumptions for heavy-duty vehicle classes, including buses and incomplete vehicles. As capital costs increase, the optimization model reallocates replacement effort toward lower-cost vehicle classes, selecting heavy-duty replacements only when required by budgetary or policy constraints. This behavior confirms that the stochastic optimization responds coherently to economic inputs and does not rely on fragile parameter tuning.

Results are reported as ranges of expected annual cost (EAC), replacement counts, and vehicle-class mix. This presentation enables managers to visualize trade-offs between budget allocation, operational reliability, and replacement risk, ensuring that recommended policies remain defensible across a range of uncertainty conditions.

6.8.1 Policy Sensitivity to Capital Cost Assumptions

To assess robustness, the replacement policy was evaluated under alternative capital-cost assumptions for heavy-duty vehicle classes, including buses and incomplete vehicles. As capital costs increased, the optimization naturally shifted investment toward light-duty assets, while heavy-duty replacements were selected only when required by policy constraints. This behavior confirms that the stochastic optimization responds predictably to economic inputs and that resulting policies are stable under plausible parameter variation.

6.9 Outcomes and Delivered Artifacts

The hybrid framework developed in this chapter demonstrates how interpretable machine learning and stochastic optimization can be integrated into a unified decision-support system for public-sector fleet planning. The principal outcomes of this framework include:

- Reduction in expected annual cost (EAC) relative to benchmark replacement policies by aligning predictive modeling with downstream economic objectives;
- Interpretable estimates of cost risk and operational burden that serve as actionable managerial levers in fleet-replacement planning;
- Transparent and reproducible analytics that strengthen accountability and defensibility in public-sector capital decision making.

The implementation of the hybrid framework yields the following delivered artifacts:

1. A documented Python-based analytics pipeline integrating data preprocessing, predictive modeling, uncertainty construction, and stochastic optimization;
2. A calibrated two-stage stochastic replacement model in which scenario-based operating costs are derived from machine-learning residuals and solved under realistic budget and policy constraints;
3. An interpretable machine-learning component with SHAP-based explanations, validation diagnostics, and decision-aware evaluation metrics;
4. A decision-focused evaluation layer that quantifies policy regret, capture rate, and expected cost reduction induced by predictive error;
5. Tabular and graphical policy summaries describing replacement mix, cost savings, and sensitivity to capital-cost and policy assumptions.

All data processing and modeling were conducted locally in Python and maintained in private repositories. Due to confidentiality agreements and data sensitivity, neither raw data nor executable code are released publicly; however, all modeling steps, assumptions, and evaluation procedures are documented to ensure reproducibility and auditability within the sponsoring agency.

While fully differentiable learning-augmented optimization remains an active research direction, the framework developed here demonstrates that economically

aligned hybridization is achievable without sacrificing interpretability or tractability in large public-sector systems.

6.10 Summary and Integration of the Hybrid Framework

This chapter presented a hybrid predict–then–optimize framework that integrates machine learning and stochastic programming into a single, data-driven decision system for fleet replacement planning. The predictive modeling layer produces interpretable estimates of vehicle operating cost that capture nonlinear effects of age, utilization, ownership, and organizational context. These predictions, together with empirically calibrated uncertainty derived from out-of-fold residuals, parameterize a stochastic optimization model that generates replacement policies under budgetary and operational constraints.

Rather than treating predictive accuracy as the sole objective, the framework evaluates learning performance through its impact on downstream decision quality. Decision-aware metrics such as policy regret, replacement capture, and expected annual cost reduction are used to assess how predictive errors translate into economic consequences. This evaluation-driven form of decision-focused learning ensures that model improvements are concentrated in regions of the cost surface that materially affect replacement decisions.

Together, these interconnected components establish a unified analytical foundation that balances interpretability, robustness, and policy relevance. The results demonstrate that aligning machine-learning predictions with stochastic optimization objectives—without requiring reformulation into differentiable optimization structures. As such, the proposed framework provides a defensible and scalable approach for data-informed asset lifecycle planning across public agencies.

Chapter 7

Integration and Conclusion

7.1 Overview

This dissertation develops and evaluates an integrated analytical framework for public-sector fleet management that unifies stochastic optimization, interpretable machine learning, empirical lifecycle analysis, and decision-focused integration. The central objective has been to answer the following question:

> *How can government agencies make transparent, economically defensible fleet replacement and procurement decisions under uncertainty?*

Rather than advancing a single algorithm or prescriptive rule, the dissertation constructs a decision infrastructure that connects empirical cost behavior, uncertainty-aware optimization, and interpretable predictive modeling within a coherent policy framework. The emphasis is not merely on computational optimality, but on defensibility, transparency, and institutional feasibility in real public-sector environments.

Within this context, the dissertation establishes that stochastic programming, interpretable machine learning, empirical lifecycle analysis, and decision-focused integration are not competing paradigms but interlocking components of a unified

analytical architecture. Their integration yields a scalable, auditable, and economically grounded foundation for lifecycle-oriented fleet policy capable of withstanding institutional, fiscal, and political constraints.

7.2 Synthesis of Contributions

The dissertation contributes four complementary methodological components and demonstrates how their integration strengthens both analytical rigor and practical relevance in public-sector capital planning.

7.2.1 Stochastic Optimization and Decomposition

Chapter 3 formulates fleet planning as a two-stage stochastic program that captures the core commitment–recourse trade-off faced by public agencies. A calibrated four-vehicle-type, three-scenario instance is used to compare Benders Decomposition (BD), Lagrangian Relaxation (LR), and Progressive Hedging (PH).

Across methods, the results converge to the same optimal ownership expansion and expected cost, indicating that classical decomposition techniques provide alternative algorithmic perspectives on the same structural decision problem. Importantly, each method offers a distinct form of decision insight: BD produces explicit bounds and cuts, LR provides strong early dual information, and PH offers intuitive scenario-level coordination. This equivalence reinforces that method choice is less about optimality and more about the form of insight and transparency required by the decision context. These properties are particularly relevant in public-sector environments where solution interpretability and documentation are as important as computational performance.

7.2.2 Interpretable Machine Learning for Cost Modeling

Chapter 4 develops an interpretable machine-learning framework for predicting vehicle-year operating cost using ensemble models combined with SHAP-based explanations. Validation is conducted using VIN-safe temporal splits and agency-blocked folds to reflect realistic forecasting constraints.

The analysis shows that lifecycle cost is driven primarily by asset age, utilization, downtime, and organizational context, with substantial variation across vehicle classes and agencies. Once censoring and usage heterogeneity are explicitly accounted for, widely used rules-of-thumb—such as a universal lease–own crossover at mid-life—are not empirically supported.

7.2.3 Empirical Identification of Diminishing Returns and Economic Life

Chapter 5 provides a data-driven bridge between classical replacement theory and observed fleet cost behavior. Using sixteen years of longitudinal operational data and nonparametric smoothing techniques, the chapter estimates class-specific cost–age trajectories and identifies economically meaningful diminishing-returns tipping ages for major vehicle classes.

Rather than imposing fixed age–mileage thresholds, the analysis derives economic-life benchmarks directly from observed cost acceleration patterns. The results demonstrate substantial heterogeneity across sedans, SUVs, and pickup trucks, revealing that uniform replacement cycles obscure meaningful differences in lifecycle dynamics.

This empirical component plays a structural role within the broader framework. It connects descriptive lifecycle modeling to prescriptive decision-making by anchoring both the stochastic optimization model and the hybrid decision-focused integration in transparent, data-derived economic-life estimates. In doing so, it operationalizes

classical economic replacement theory within a modern, data-rich public-sector context.

7.2.4 Hybrid Predict–Then–Optimize Integration

Chapter 6 introduces a Decision–Coupled Learning (DCL) framework that evaluates predictive models based on their downstream economic impact within a stochastic optimization setting. Rather than treating prediction accuracy as an end in itself, the framework measures model quality through expected annual cost (EAC) and policy regret.

By embedding predictive cost signals within the optimization structure, the hybrid approach aligns learning objectives with economic decision performance. The results demonstrate that models calibrated for decision quality produce more stable replacement timing and improved economic relevance relative to sequential predict–then–optimize workflows.

7.3 Managerial and Policy Implications

Beyond methodological contributions, the dissertation speaks directly to the institutional realities of public-sector fleet management. The class-specific diminishing-returns ages identified empirically provide concrete, auditable economic-life benchmarks that anchor both the stochastic and hybrid models. These tipping ages transform descriptive cost behavior into prescriptive guidance, enabling agencies to justify replacement timing using transparent, data-derived evidence rather than heuristic thresholds.

First, the stochastic results show that modest preemptive capacity—owning slightly more than average demand—can be economically rational when emergency procurement is costly or slow. This provides analytical justification for practices that are often intuitively appealing but difficult to defend during budget negotiations.

Second, the integration of SHAP explanations and dual information supports documented, reproducible policy justification. Agencies can explain *why* a vehicle is expensive to retain and *why* additional capacity or earlier replacement is economically justified, using evidence that is numerically explicit and reproducible.

Third, the framework corrects misaligned heuristics. Uniform age-based replacement rules and simplistic lease-versus-own guidelines mask substantial heterogeneity in cost behavior. Replacing these heuristics with data-driven, class-specific, and risk-aware guidance supports more equitable and fiscally efficient policy.

Taken together, the framework enables agencies to move from static replacement rules toward structured lifecycle planning that aligns cost, risk, and service reliability within existing procurement and governance constraints.

7.4 Conceptual and Methodological Integration

While each chapter addresses a distinct analytical gap, the principal contribution of the dissertation lies in demonstrating how stochastic optimization, interpretable learning, and empirical lifecycle estimation can be integrated into a coherent and institutionally viable decision system.

First, the research reframes fleet replacement as a structured economic decision problem under uncertainty rather than a forecasting exercise. Statistical accuracy alone does not guarantee improved lifecycle efficiency. By embedding cost prediction within a stochastic optimization structure and evaluating performance using expected annual cost (EAC) and policy regret, the framework aligns analytical tools with economic decision quality.

Second, interpretability is treated not as an auxiliary feature but as a structural requirement in public-sector governance environments. SHAP explanations, dual-variable interpretations, and scenario comparisons provide economically meaningful diagnostics that support auditability and legislative accountability.

Third, the empirical identification of diminishing returns provides a missing link between descriptive lifecycle cost modeling and prescriptive replacement timing. By estimating class-specific cost inflection points from longitudinal data, the dissertation operationalizes classical economic-life theory within a modern data context.

Together, these elements establish a hybrid analytical paradigm in which:

- stochastic optimization manages commitment under uncertainty,
- interpretable machine learning captures heterogeneous cost dynamics,
- empirical smoothing identifies structural tipping ages, and
- decision-focused evaluation aligns learning with economic performance.

This integration advances both methodological rigor and institutional applicability, particularly in high-stakes public decision systems.

7.5 Scope and Limitations

Several elements of the framework are intentionally scoped to balance realism, interpretability, and computational tractability. The stochastic optimization experiments use a stylized instance to isolate decomposition behavior; the ML models focus on explainable vehicle-year cost prediction rather than failure-time modeling; and the hybrid integration emphasizes decision-aware evaluation rather than fully differentiable end-to-end learning. These choices prioritize auditability and clarity while motivating future extensions. The framework does not require fully differentiable end-to-end optimization through mixed-integer programs; instead, it prioritizes implementability, auditability, and policy stability in public-sector environments.

Additionally, the empirical diminishing-returns analysis relies on historical operating data and assumes that past cost trajectories remain structurally informative for near-term planning. Structural shifts—such as electrification, supply-chain

disruptions, or regulatory change—may alter future cost surfaces, motivating periodic recalibration of lifecycle benchmarks.

7.6 Future Research Directions

The dissertation establishes a foundation for several extensions:

- **Multi-period stochastic replacement**, capturing depreciation, rollover constraints, and dynamic budget interactions over longer horizons.
- **Telematics and condition-based integration**, incorporating mileage, engine hours, and health diagnostics to refine uncertainty representation.
- **Statewide and cross-agency coordination**, modeling shared capacity and geographic constraints.
- **Electrification planning**, extending the framework to battery-electric and hybrid fleets, where maintenance patterns, energy costs, and infrastructure requirements differ substantially from conventional vehicles.
- **Interactive decision-support tools**, translating the framework into dashboards that support scenario exploration and stakeholder engagement.
- **Equity and service-level constraints**, evaluating how replacement policies interact with regional access, response time, and service equity.

7.7 Concluding Remarks

Stochastic optimization and machine learning are often developed and applied in isolation. This dissertation demonstrates that their integration—when structured around interpretability, empirical lifecycle analysis, and decision-aware evaluation—yields a coherent and practically implementable framework for public-sector fleet management.

The research reframes fleet replacement as an economic decision problem under uncertainty rather than a forecasting exercise or a rule-based administrative task. By combining decomposition-based stochastic optimization, interpretable machine-learning models, empirical identification of diminishing-returns replacement ages, and decision-focused integration, the dissertation constructs a unified analytical pipeline that aligns predictive modeling with economic policy quality.

The resulting framework is transparent, reproducible, and uncertainty-aware. It provides economically meaningful diagnostics—shadow prices, scenario sensitivities, SHAP attributions, and policy regret measures—that support auditability and legislative accountability. Importantly, it replaces heuristic age–mileage rules with data-driven, class-specific, and risk-aware replacement guidance grounded in observed lifecycle behavior.

While motivated by a statewide fleet system, the framework generalizes to a broad class of public asset environments characterized by capital irreversibility, demand uncertainty, institutional rigidity, and governance constraints. In such settings, analytical methods must do more than compute optimal solutions—they must produce defensible, interpretable, and economically credible decisions.

By integrating uncertainty modeling, interpretable machine learning, empirical identification of economic life, and decision-coupled evaluation, this dissertation establishes a methodological foundation for transparent and economically grounded capital planning in the public sector. It demonstrates that rigorous analytics and institutional accountability are not in tension; when properly integrated, they are mutually reinforcing.

Bibliography

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79. doi: https://doi.org/10.1214/09-SS054. 55, 82

Ban, G. and Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, 67(1):90–108. https://doi.org/10.1287/opre.2018.1757. 2, 21, 78, 87

Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252. doi: https://doi.org/10.1007/BF01386316. 20, 21, 33, 40, 43, 85

Bengio, Y., Lodi, A., and Prouvost, A. (2021). Machine learning for combinatorial optimization: A methodological tour d'horizon. *European Journal of Operational Research*, 290(2):405–421. doi: https://doi.org/10.1016/j.ejor.2020.07.063. 17, 79

Berman, O. and Larson, R. C. (1985). Optimal fleet planning. *Transportation Science*, 19(2):134–148. 16, 28

Bertsimas, D., Brown, D., and Caramanis, C. (2011a). The price of robustness. *Management Science*, 57(1):35–50. 87

Bertsimas, D., Brown, D. B., and Caramanis, C. (2011b). Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501. doi: https://doi.org/10.1137/080734510. 48

Bertsimas, D. and Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044. doi: https://doi.org/10.1287/mnsc.2018.3253. 1, 21, 23, 87, 88

Birge, J. R. and Louveaux, F. (2011). *Introduction to Stochastic Programming*. Springer Science+Business Media, LLC, New York. doi: https://doi.org/10.1007/978-1-4614-0237-4. 21, 85

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32. doi: https://doi.org/10.1023/A:1010933404324. 21, 83

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. doi: https://doi.org/10.1145/2939672.2939785. 22, 83

Chen, V., Li, J., Kim, J. S., Plumb, G., and Talwalkar, A. (2022). Interpretable machine learning: Moving from mythos to diagnostics. *Communications of the ACM*, 65(8):43–50. doi: https://doi.org/10.1145/3546036. 51, 60

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836. doi: https://doi.org/10.1080/01621459.1979.10481038. 70, 72

Cleveland, W. S. and Devlin, S. J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 74(368):829–836. doi: https://doi.org/10.1080/01621459.1988.10478639. 70, 72

Codato, G. and Fischetti, M. (2006). Combinatorial benders' cuts for mixed-integer linear programming. *Operations Research*, 54(4):756–766. doi: https://doi.org/10.1287/opre.1060.0286. 41

Cook, A. R. (1985). Life cycle costing in the transit industry. *Transportation Research Record*, (1011):1–4. 69

Donti, P. L., Amos, B., and Kolter, J. Z. (2017). Task-based end-to-end model learning in stochastic optimization. In *Advances in Neural Information Processing Systems*, volume 30. https://doi.org/10.48550/arXiv.1703.04529. 16, 79

Elmachtoub, A. N. and Grigas, P. (2022). Smart "predict, then optimize". *Management Science*, 68(1):9–26. doi: https://doi.org/10.1287/mnsc.2020.3922. 1, 2, 10, 16, 21, 23, 67, 78, 79, 87

Ercan, T., Zhao, Y., Tatari, O., and Pazour, J. A. (2015). Optimization of transit bus fleet's life cycle assessment impacts with alternative fuel options. *Energy*, 93:1156–1169. doi: https://doi.org/10.1016/j.energy.2015.09.018. 18, 19, 29, 39, 54, 62

Escudero, L. F., Garín, M. A., Merino, M., and Pérez, G. (2007). The value of the stochastic solution in multistage problems. *TOP*, 15(1):48–64. doi: https://doi.org/10.1007/s11750-007-0005-4. 28, 39

Eskandarpour, M., Dejax, P., Miemczyk, J., and Péton, O. (2015). Sustainable supply chain network design: An optimization-oriented review. *Omega*, 54:11–32. doi: https://doi.org/10.1016/j.omega.2015.01.006. 29

Fan, J. (1996). *Local Polynomial Modelling and Its Applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman and Hall/CRC, New York. doi: https://doi.org/10.1201/9780203748725. 70

Fisher, M. L. (1981). The lagrangian relaxation method for integer programming problems. *Management Science*, 27(1):1–18. doi: https://doi.org/10.1287/mnsc.27.1.1. 40, 43

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232. doi: https://doi.org/10.1214/aos/1013203451. 21, 22, 83

Frieß, N. M. and Pferschy, U. (2024). Planning a zero-emission mixed-fleet public bus system with minimal life cycle cost. *Public Transport*, 16(1):39–79. doi: https://doi.org/10.1007/s12469-023-00345-4. 18, 19, 29, 39

Geoffrion, A. M. (1974). Lagrangean relaxation for integer programming. *Mathematical Programming*, 2(1):82–114. 20, 21, 40, 43, 85

Govindan, K., Fattahi, M., and Keyvanshokooh, E. (2017). Supply chain network design under uncertainty: A comprehensive review and future research directions. *European Journal of Operational Research*. doi: https://doi.org/10.1016/j.ejor.2017.04.009. 29

Guignard, M. (2003). Lagrangean relaxation. *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research*, 11(2):151–200. doi: https://doi.org/10.1007/BF02579036. 20, 40, 85

Hartman, J. (2004). Multiple asset replacement analysis under variable utilization and stochastic demand. *European Journal of Operational Research*, 159(1):145–165. doi: https://doi.org/10.1016/S0377-2217(03)00397-7. 11, 17, 19, 69

Hooker, J. N. (2007). Planning and scheduling by logic-based benders decomposition. *Operations Research*, 55(3):588–602. doi: https://doi.org/10.1287/opre.1060.0371. 41

Hooker, J. N. and Ottosson, G. (2003). Logic-based Benders decomposition. *Mathematical Programming*, 96(1):33–60. doi: https://doi.org/10.1007/s10107-003-0375-9. 41

Jardine, A. K. S. and Tsang, A. H. C. (2021). *Maintenance, Replacement, and Reliability: Theory and Applications*. CRC Press, Boca Raton, FL, 3 edition. doi: https://doi.org/10.1201/9780429021565. 11, 17, 19, 69

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30, pages 3146–3154. Curran Associates, Inc. 21, 22, 83

Khalili-Damghani, K., Tavana, M., and Santos-Arteaga, F. J. (2022). A stochastic bi-objective location–allocation model for disaster relief logistics under uncertainty. *Annals of Operations Research*, 309:373–404. doi: https://doi.org/10.1007/s10479-021-04191-0. 28

King, A. J. and Wallace, S. W. (2012). *Modeling with Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer. doi: https://doi.org/10.1007/978-0-387-87817-1. 16, 28, 32

Kleywegt, A. J., Shapiro, A., and Homem-de Mello, T. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502. doi: https://doi.org/10.1137/S1052623499363220. 16, 20, 32

Laporte, G. and Louveaux, F. V. (1993). The integer l-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, 13(3):133–142. doi: https://doi.org/10.1016/0167-6377(93)90002-X. 20, 40, 41

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to

global understanding with explainable ai for trees. *Nature Machine Intelligence*, 2(1):56–67. doi: https://doi.org/10.1038/s42256-019-0138-9. 21, 22, 51, 60, 83

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, volume 30, pages 4765–4774. Curran Associates, Inc. 21, 22, 60, 83

Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Christoph Molnar. 50, 51, 57, 83

Mulvey, J. M. and Ruszczy'nski, A. (1995). A new scenario decomposition method for large-scale stochastic optimization. *Operations Research*, 43(3):477–490. doi: https://doi.org/10.1287/opre.43.3.477. 39

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B. (2019). Interpretable machine learning: Definitions, methods, and applications. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080. doi: https://doi.org/10.1073/pnas.1900654116. 21, 22, 50, 51, 60

Natekin, A. and Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7:21. doi: https://doi.org/10.3389/fnbot.2013.00021. 22, 53, 56, 57

National Cooperative Highway Research Program (2019). Fleet replacement management practices. Technical Report NCHRP Research Report 917, Transportation Research Board, National Academies of Sciences, Engineering, and Medicine, Washington, D.C. Guidance on state DOT fleet lifecycle modeling and replacement policy. 18, 19, 48, 51

Noyan, N. (2012). Risk-averse two-stage stochastic programming with an application to disaster management. *Computers & Operations Research*, 39(3):541–559. doi: https://doi.org/10.1016/j.cor.2011.03.017. 40, 65

Noyan, N. (2016). A stochastic optimization model for designing last-mile relief supply networks. *Transportation Science*, 50(4):1162–1179. doi: https://doi.org/10.1287/trsc.2015.0621. 40, 65

Pantuso, G. (2016). Uncertainty in fleet renewal: A case from maritime shipping. *Transportation Science*, 50(3):910–925. doi: https://doi.org/10.1287/trsc.2014.0566. 28

Pantuso, G., Fagerholt, K., and Hvattum, L. M. (2014). A survey on maritime fleet size and mix problems. *European Journal of Operational Research*, 235(2):341–349. doi: https://doi.org/10.1016/j.ejor.2013.04.058. 9

Powell, W. B. (1986). A stochastic model of the dynamic vehicle allocation problem. *Transportation Science*, 20(2):117–129. doi: https://doi.org/10.1287/trsc.20.2.117. 20, 48

Powell, W. B. and Topaloglu, H. (2005). Stochastic programming in transportation and logistics. *Transportation Science*, 39(4):377–396. 28, 39

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31. 83

Quarles, N., Gopalakrishnan, B., Machek, D., and Phillips, J. (2020). Costs and benefits of electrifying and automating bus transit fleets. *Sustainability*, 12(10):3977. doi: https://doi.org/10.3390/su12103977. 18, 19, 54

Rahmaniani, R., Crainic, T. G., Gendreau, M., and Rei, W. (2017). The Benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259(3):801–817. doi: https://doi.org/10.1016/j.ejor.2016.12.005. 20, 33, 40, 41, 43

Ramírez-Pico, C., Ljubić, I., and Moreno, E. (2023). A Benders adaptive-cuts method for two-stage stochastic programs. *Transportation Science*, 57(5):1115–1140. doi: https://doi.org/10.1287/trsc.2022.0073. 20

Raposo, H., Farinha, J. T., Ferreira, L., and Galar, D. (2018). Dimensioning reserve bus fleet using life cycle cost models and condition based/predictive maintenance: A case study. *Public Transport*, 10(1):169–190. doi: https://doi.org/10.1007/s12469-017-0167-x. 29, 39

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144. doi: https://doi.org/10.1145/2939672.2939778. 21, 83

Ritari, A., Vepsäläinen, J., Kivekäs, K., Tammi, K., and Laitinen, H. (2020). Energy consumption and lifecycle cost analysis of electric city buses with multispeed gearboxes. *Energies*, 13(8):2117. doi: https://doi.org/10.3390/en13082117. 18

Roberts, D. R., Bahn, V., Ciuti, S., and et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Methods in Ecology and Evolution*, 8(5):628–639. doi: https://doi.org/10.1111/ecog.02881. 22, 55, 82

Rockafellar, R. T. and Wets, R. J.-B. (1991). Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16(1):119–147. doi: https://doi.org/10.1287/moor.16.1.119. 20, 40, 43, 85, 90

Rudin, C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215. doi: https://doi.org/10.1038/s42256-019-0048-x. 16, 22, 23, 80

Santoso, T., Ahmed, S., Goetschalckx, M., and Shapiro, A. (2005). A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*, 167(1):96–115. doi: https://doi.org/10.1016/j.ejor.2004.01.046. 20, 28, 39

Sarangi, S., Sarangi, S., and Sabounchi, N. S. (2023). How managerial perspectives affect the optimal fleet size and mix model: A multi-objective approach. *OPSEARCH*, 60(1):1–23. doi: https://doi.org/10.1007/s12597-022-00603-2. 40

Sarayloo, F., Crainic, T. G., and Rei, W. (2023). An integrated learning and progressive hedging matheuristic for stochastic network design problems. *Computational Optimization and Applications*, 29(4–6):409–434. doi: https://doi.org/10.1007/s10732-023-09515-w. 23, 41, 67

Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2009). *Lectures on Stochastic Programming: Modeling and Theory*. Society for Industrial and Applied Mathematics. doi: https://doi.org/10.1137/1.9780898718751. 16, 20, 21, 28, 32, 85

Triantafyllou, N. and Papathanasiou, M. M. (2024). Deep learning enhanced mixed integer optimization: Learning to reduce model dimensionality. *Computers & Chemical Engineering*, 187:108725. doi:https://doi.org/10.1016/j.compchemeng.2024.108725. 23, 67

Varma, S. and Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7:91. doi: https://doi.org/10.1186/1471-2105-7-91. 22, 55, 82

Wallace, S. W. and Fleten, S.-E. (2003). Stochastic programming models in energy. In Ruszczyński, A. and Shapiro, A., editors, *Handbooks in Operations Research and*

Management Science, volume 10, pages 637–677. Elsevier. doi: https://doi.org/10.1016/S0927-0507(03)10010-2. 39

Wang, J., Shirkoohi, M. G., Akter, R., and Mérida, W. (2025). Urban transit decarbonization via stochastic programming and statistical inference under uncertainty. *Transportation Research Part D: Transport and Environment*, 142(5):104711. doi: https://doi.org/10.1016/j.trd.2025.104711. 48

Wang, Q., Liu, Z., Jiang, P., and Luo, L. (2022). A stochastic programming model for emergency supplies pre-positioning, transshipment and procurement in a regional healthcare coalition. *Socio-Economic Planning Sciences*, 82. doi: https://doi.org/10.1016/j.seps.2022.101279. 20, 29

Wang, X., Fagerholt, K., and Wallace, S. W. (2018). Planning for charters: A stochastic maritime fleet composition and deployment problem. *Omega*, 79:54–66. doi: https://doi.org/10.1016/j.omega.2017.07.007. 28

Watson, J.-P., Woodruff, D. L., and Hart, W. E. (2012). PySP: Modeling and solving stochastic programs in Python. *Mathematical Programming Computation*, 4(2):109–149. doi: https://doi.org/10.1007/s12532-012-0036-1. 20, 21, 39, 40, 43, 85

Zheng, S. and Chen, S. (2016). Fleet replacement decisions under demand and fuel price uncertainties. *Transportation Research Part D: Transport and Environment*. doi: https://doi.org/10.1016/j.trd.2016.09.001. 28, 39

Vita

Lanette Marie Phillips was born in Tennessee. She received the Bachelor of Science degree in Civil Engineering from Tennessee Technological University and the Master of Science degree in Industrial Engineering from the University of Tennessee, Knoxville.

Before pursuing the Ph.D., she worked as a professional engineer in the public sector, focusing on transportation and fleet asset management. Her research at UTK's Department of Industrial and Systems Engineering has centered on stochastic optimization, machine learning, and explainable decision support for public-sector fleets.

She is a licensed Professional Engineer and a member of INFORMS and the Institute of Industrial and Systems Engineers (IISE). His doctoral research integrates predictive analytics and optimization to improve the lifecycle cost management of state and municipal fleets.