

**An Evaluation of NHTSA’s Economics-based Modeling
and Implications for Benefit-Cost Analysis in the NHTSA/EPA
August 24, 2018 Notice of Proposed Rulemaking (NPRM)
[“The Safer Affordable Fuel-Efficient (SAFE) Vehicles Rule for
Model Years 2021–2026 Passenger Cars and Light Trucks”]**

Prepared for the California Air Resources Board
October 24, 2018

by

David S. Bunch, PhD
University of California, Davis

Dr. Bunch is a Professor in the Graduate School of Management at the University of California, Davis, and a faculty associate of the Institute of Transportation Studies since its inception. He is an internationally recognized expert on discrete choice modeling methods, which are used to understand and predict consumer choices when they select one product from a competing set. His research interests include new product development and introduction, travel behavior, and vehicle choice, including market potential for alternative fuel vehicles. He has consulted on transportation policy issues for state and federal agencies, public utilities, and the airline industry. His recent teaching responsibilities have included Marketing Research, Product Management, and the Integrated Management Project Capstone. He received his PhD in Mathematical Sciences from Rice University (1985).

Professor Bunch has specific expertise in developing simulation models of vehicle market behavior for the purpose of evaluating alternative transportation and energy policy scenarios, including new vehicle greenhouse gas regulations, and feebates. He is the designer and creator of three versions of the CARBITS model for the California Air Resources Board, and recently chaired an expert panel advising the California Energy Commission on their ongoing enhancement of DynaSim (their market simulation model for producing transportation fuel forecasts, and evaluating alternative transportation and clean energy policies in California).

In preparing this report I analyzed the NPRM/PRIA materials, CAFE model documentation and peer review, and evaluated the NHTSA economic modeling approach based on theoretical considerations and direct comparison to relevant literature publications. I also performed numerical studies to generate empirical evidence to definitively characterize the quality of the models and their degree of conformance with economic theory.



EXECUTIVE SUMMARY

NHTSA and EPA recently published a Notice of Proposed Rulemaking (NPRM) for fuel economy/GHG regulations in the Federal Register¹, which they denote “Safer Affordable Fuel-Efficient (SAFE) Vehicles Rule for Model Years 2021–2026 Passenger Cars and Light Trucks”. The proposal is to rescind current new vehicle fuel efficiency/greenhouse gas reduction regulations for 2021-2025 (“Existing standards”) and freeze the standards at 2020 levels through 2026. This proposed “Rollback” represents a complete reversal of the previous determination in 2017 by EPA that the Existing standards continue to be appropriate. The 2017 determination was based on an extensive and well-documented midterm review process--see, e.g., the 2016 Technical Assessment Report (“TAR”), EPA et al. (2016).

The Agencies’ summary argument in favor of the Rollback rests on two major claims (plus related measures). From the NPRM (page 42986, column 1):

“Compared to maintaining the post-2020 standards set forth in 2012, current estimates indicate that the proposed SAFE Vehicles Rule [the Rollback] *would save over 500 billion dollars in societal costs and reduce highway fatalities by 12,700 lives* (over the lifetimes of vehicles through MY 2029).” [Emphasis added.]

The two italicized items are outcome measures (“net benefits”) from NHTSA’s economic modeling. When announced, the immediate reaction from a wide range of experts was that they are dubious, and that the claims lack what researchers call *face validity*². This report provides a rigorous analysis and evaluation of key aspects of NHTSA’s economic modeling efforts, and unequivocally confirms what seems so obvious to so many.

More specifically, estimation of net benefits relies critically on developing reasonable projections of how the future vehicle market will behave under alternative regulatory scenarios (a potentially challenging undertaking). In prior rulemakings, NHTSA and EPA carefully crafted an approach that incorporated projections from reputable, third party sources. For this rulemaking, NHTSA elected to instead develop its own in-house vehicle market simulation model. This report identifies and demonstrates multiple shortcomings in key modeling components, and shows how these shortcomings lead to results that are inconsistent with basic economic principles (a violation of NHTSA’s own modeling standards).

¹ **Federal Register**/Vol. 83, No. 165/Friday, August 24, 2018/Proposed Rules, pp. 42986 – 43500.

² “In statistics, etc., the fact of something seeming to be a reasonable or accurate measure of something: If a test has face validity, then it looks like a valid test to those who use it.” [<https://dictionary.cambridge.org/us/dictionary/english/face-validity> (October 18, 2018).]

In more common parlance: *their models have major flaws for multiple reasons, and produce incorrect results that are bad enough to lead to a wrong conclusion, i.e., that a Rollback has positive net benefits versus the Existing standards.*

Moreover, all of these actions and decisions were made within a very short time frame, and apparently without the benefit of any meaningful peer review. A careful reading of the NPRM and related documentation suggests that NHTSA was fully aware of these deficiencies, but chose to proceed anyway. Because economic modeling is a subject requiring significant technical expertise, the remainder of this summary provides a review of background material prior to a more detailed enumeration and explanation of the report's findings.

Background on vehicle market modeling and approaches.

The purpose of rulemaking for new vehicle efficiency/GHG standards is to alter the behavior of a large and important economic market (the vehicle market) in a way that produces outcomes that meet specified policy objectives (greater fuel efficiency and reduced greenhouse gases). However, this is also quite likely to have an economic impact on multiple stakeholders. Different regulatory options will have different impacts, and the benefits and costs must be evaluated. In this case, regulatory analysis relies on quantitative economic models to determine how market behavior is likely to change under alternative regulatory scenarios, so that the impact on benefits and costs can be estimated.

Economic theory plays a major role because it is the accepted framework for developing models of market behavior. The underlying principles take the form of specific behavioral assumptions that determine the decisions of stakeholders (e.g., consumers and producers) when participating in a market that involves multiple interactions. In the case of vehicle markets, manufacturers use technologies (inputs) to produce and sell new vehicles (*Supply*). They decide which designs/features to use for their vehicle offerings, and what prices to charge. The prototypical assumption is that they make these decisions so as to maximize profits subject to constraints (e.g., their technological capabilities, availability of inputs).

Demand arises from an aggregation of individual decisions by consumers (or, households) on how many vehicles to own, which ones (chosen from among a large set of competing options), and how to use them. They get "utility" from the mobility services that vehicles provide, and their specific choices are based on *preferences* for *product attributes* (e.g., new or used, car/SUV/van/truck, size, seating capacity, fuel efficiency) including *purchase price*. Consumers can vary in their preferences and behaviors, which can be represented using the concept of consumer *segments*.

Multiple types of interactions contribute to complexity in the vehicle market. Some consumers buy new vehicles from manufacturers, but in the *used* vehicle market consumers can be both buyers and/or sellers. Near the end of a vehicle's life when it no longer makes economic sense to keep it in operation, it will be sold for scrap.

With all of these interactions, *price changes* occur to “clear the market” so as to create a balance between *supply and demand* across all sub-markets (new, used, scrappage).

To perform analyses that capture these phenomena, theory-based models have been developed that incorporate the fundamental behavioral features and interactions described above. These so-called *structural models* specifically identify how changes in certain variables (e.g. price, fuel economy) *directly* influence an outcome of interest (e.g. purchase behavior), usually in the form of *parameters that have clear economic interpretation* (e.g. preference for an attribute). One specific class, *discrete choice models*, can be used simulate households’ vehicle-related decisions using the features discussed above (e.g., attributes, preferences, and segmentation). (For a more detailed discussion, see section 3.2).

It is clear that implementing a highly detailed system of structural models could be challenging, and model development inevitably involves making simplifying assumptions. It is understandable that analysts would seek simpler modeling approaches that require less detail. Many times these efforts yield models in the form of equations that produce results at a more highly aggregated level, so that potentially important details on the structural, behavioral features are lost: so-called *reduced form* models. If such models are carefully derived with the goal of ensuring consistency with theory, they could produce realistic “behavioral responses” to input changes even when structural features are no longer apparent. While this can be feasible under the right circumstances, there are always risks. Rigorous validation and testing of model behavior are an absolute requirement. Unfortunately, it is very easy for such modeling attempts to go awry when analysts face limitations on data and other resources.

A specific class of reduced form models (used by NHTSA) is autoregressive distributed-lag (ARDL) models, which are used for *aggregate-level time-series forecasting*. In the literature, this type of model represents the polar opposite of structural models. Their major strength is in producing *short-term forecasts*, which are essentially a descriptive extrapolation of *existing trends* in historical data. These can be useful for supporting short-term decisions (by, e.g., industry managers) in situations where it is safe to assume that no meaningful structural changes are occurring in the market. Although there are times when such models can play *some* role in supporting policy analysis, they are particularly unsuitable for capturing behavioral responses of stakeholders to fundamental market changes of the type induced by policy.

Modeling requirements and evaluation standards.

Given the previous discussion, it is clear that developing and using economic models for regulatory analysis requires careful judgment and technical expertise. The Agencies have helpfully provided some (brief) background material on the standards they are required to meet. The Preliminary Regulatory Impact Analysis (or, “PRIA”—see page 937) focuses on the requirement to produce *measures of economic benefits and costs*, indicating that they...

“... are important considerations, because as Office of Management and Budget Circular A-4 states, *benefits and costs* reported in regulatory analyses must be *defined and measured consistently with economic theory*, and should also reflect how alternative regulations are anticipated to change the *behavior of producers and consumers* from a baseline scenario.⁴⁷⁹ In this analysis, those include *vehicle manufacturers, buyers of new cars and light trucks, and owners of used vehicles*, all of whose *behavior* is likely to be affected in complex ways by the proposed action to adopt less strict CAFE and CO₂ emission standards for future years.” [Emphasis added.]

This description, when combined with the previous background discussion, suggests the following are required for success: Choosing a modeling approach/methodology sufficient for capturing behavioral responses to policy options in a manner prescribed by economic theory, availability of required data and related resources, and correct execution (e.g., choice of model specifications, statistical estimation, etc.). Having established this background on modeling and evaluation standards, the remainder of this summary provides more detail on what the Agencies have done, and our evaluation.

Background on the CAFE Model. Over the course of many rulemakings, NHTSA has relied on output from its own modeling system (the CAFE Compliance and Effects Model, also referred to as “the CAFE model” or, prior to the current rulemaking, “the Volpe model”) for producing net benefit analyses. Until the current rulemaking, the CAFE model was focused almost exclusively on modeling *manufacturer* decision-making behavior in response to fuel efficiency standards. Specifically, the CAFE model was primarily concerned with identifying feasible vehicle redesign pathways that could be used by manufacturers to comply with fuel efficiency standards.

Within the CAFE model is a “manufacturer decision module” that simulates each manufacturer’s adoption of new fuel-saving technologies for future model years (relying on an extensive database of technologies and costs, and a complex algorithm). Technologies are added to a base year fleet of over 1600 vehicle offerings to create new vehicles in future model years. This “behavioral model” is a simplification of our earlier description: in the CAFE model, manufacturers do not decide on both technologies and prices to maximize profits. Instead, they can only choose technologies, and they do so on the basis of minimizing their costs. (This aspect of NHTSA’s modeling is outside the scope of this report, but we note that the simplification of excluding pricing, while perhaps reasonable under the

circumstances, could yield simulation results that inappropriately adopt technologies at higher cost versus what would happen in the real world.)

On the consumer side, in prior rulemakings the CAFE model incorporated vehicle market projection data from external sources (as mentioned earlier) as well as other assumptions to compute outcome measures of *economic costs* and *environmental effects*. However, in the most recent rulemaking NHTSA developed and incorporated its own internal vehicle market model into the CAFE model. In addition, the CAFE model has been modified to analyze greenhouse gas regulations as well as fuel economy standards, so that all of the NPRM analyses rely on the CAFE model. In prior rulemakings, EPA independently performed its own analyses using its own models.

Overview of NHTSA’s models. NHTSA’s *vehicle market model* can be viewed as consisting of multiple (sub-)models (or modules). NHTSA added three of these to simulate the following phenomena: *new vehicle sales*, *dynamic fleet share (changes in relative share of cars versus light trucks)*, and *dynamic scrappage (vehicles retired/removed from the fleet)*. The results of these are then combined to simulate the evolution of the US light-duty vehicle fleet. When evaluating model behavior, the focus could be on one of the three individual components, or on the overall vehicle market (model) behavior when they are combined.

All three of the above components are reduced-form models of the type discussed previously: *aggregate-level time-series forecasting models*. For example, their *new vehicle sales* model takes the following form:

$$NewSales(Year\ t) = F_{Sales} \left[\begin{array}{l} NewSales(Year\ t-1),\ NewSales(Year\ t-2), \\ GDP_GrowthRate(Year\ t),\ \Delta_AveComplianceCost(Year\ t), \\ LaborForceParticipation(Year\ t),\ LaborForceParticipation(Year\ t-1) \end{array} \right].$$

There are three types of input variables. Two are macroeconomic indexes (GDP_GrowthRate, and LaborForceParticipation). There is only one variable related to new vehicles: the change (Δ) in the *average compliance cost* (the average cost of adding technology to meet the regulations, taken over all manufacturers). There are many *lag effects* (a defining feature of time-series models), including *NewSales* from the two previous periods.

It is important to understand that this model is used to *directly* compute *total (aggregate) sales forecasts for three types of vehicles: Cars, SUV/Vans, and Trucks*. In other words, although the CAFE model maintains detailed specifications for over 1600 vehicle configurations, and simulates redesign decisions for each year, *no “consumer” ever actually “sees” any of these vehicle options or makes purchase decisions based on preferences for vehicle attributes*. The only vehicle-related variable (average compliance cost for the entire fleet) is used as a proxy for vehicle

price. However, the reason for increased compliance cost is that technology has been added to vehicles to improve their fuel efficiency (another vehicle attribute for which consumers have preferences).

This model lacks both the level of detail and the structural features (e.g., vehicle offerings, attributes, preferences, and segmentation) that would seem to be required for capturing consumer responses to changes in new vehicle market offerings. In this specific implementation, *even though all cost increases are accompanied by improvements in fuel efficiency, this improvement is completely ignored.*

The situation is similar for dynamic scrappage, although in this case the models are much more complex. Some background: A *scrappage model* gives *scrappage rates* for groups of vehicles of the same age. A scrappage rate is defined as: the *probability (P)* that a vehicle in a given *age group* will be scrapped during the current year. (Conversely, the probability of survival is $1-P$). The total number of vehicles (T) in this group will therefore diminish, so that the number of surviving vehicles at the beginning of the next year will be: $T_i(1-P)$.

NHTSA's dynamic scrappage model is the *sole determinant* of what happens in the "used vehicle market" in the CAFE model. However, as in the new vehicle sales model, a key vehicle-related input variable is *average compliance cost for new vehicles* (not, e.g., a more direct measure of used vehicle prices). In fact, there are *no structural connections* to capture behavioral interactions between the new and used vehicle markets (a fact that will turn out to be very important, as discussed later). For reasons that will become apparent later, no attention is devoted to the dynamic fleet share model in this report. The purpose of this discussion was to support the following summary of key findings. We divide these into two categories: evaluation based on theoretical considerations, and evidence based on numerical studies and analysis.

Summary of findings based on theoretical considerations. Recall the Agencies' own requirements: Their models should be able to *simulate the behavioral responses of vehicle manufacturers, buyers of new cars and light trucks, and owners of used vehicles to policy changes, and to produce results that are consistent with economic theory.* A review of the models based on theoretical considerations yields the following findings.

T1. The single-equation aggregate-level, time-series equations used by NHTSA lack sufficient level of detail and *structural features* to adequately capture consumers' *behavioral responses to changes in vehicle prices and attributes* made by manufacturers attempting to comply with CAFE/GHG standards. Vehicle choice options, attributes, and consumer preferences are not adequately represented (i.e., only aggregate new vehicle sales are being forecasted, and they are assumed to respond only to average increases in compliance costs).

T2. There is *no meaningful structural relationship or interaction between the new and used vehicle markets* of the type required by theory. Specifically, there are no linkages between the new vehicle sales model and the dynamic scrappage model ensured to capture theoretical requirements. This deficiency makes the models vulnerable to producing simulation results that are *inconsistent with economic theory*.

NHTSA's dynamic scrappage model (considered in isolation) has comparable analogs in the published literature to support an evaluation. Based on key literature references (and, in turn, other literature cited by them), NHTSA's approach is clearly deficient. In particular:

T3. Other scrappage models in the literature are specified based on well-established theory, and use structural formulations with parameters that have clear economic meaning. NHTSA essentially ignores these and instead opts for a time-series curve-fitting approach with no obviously identifiable behavioral structure, and uninterpretable parameters.

One unfortunate consequence of this approach is that the documentation and results in the PRIA for the scrappage models were generally quite impenetrable to direct interpretation (even by experts), contributing to the need for the numerical studies discussed below. (See section 3.6.2.)

Even if aggregate-level forecasting approaches were appropriate for policy analysis, the large literature on travel demand forecasting suggests that NHTSA's overall approach of forecasting new sales in conjunction with scrappage is inferior to other alternatives. Established aggregate-level approaches start with projections of *growth in the total vehicle fleet (reflecting transport needs), not new vehicle sales*. These typically exclude prices altogether, because experience has shown that prices do not help the accuracy of aggregate forecasting models. (In contrast, approaches using more detailed *discrete choice models for consumers' vehicle-related decisions* make successful use of price effects.)

T4. A key reference, Greenspan and Cohen (1999), solves the *same* forecasting problem as NHTSA, but uses a different approach that is consistent with the travel demand literature. They first forecast the size of the *total vehicle fleet* (based on population and household ownership trends from the Census Bureau, plus economic indexes), and then simulate *scrappage*. New vehicle sales are determined on the basis of these other two forecasts (see section 3.3).

There are also other issues that raised concerns.

T5. In terms of technical execution (statistical estimation), NHTSA’s complex, time-series based models are vulnerable to *over-fitting the historical data*, leading to poor “out of sample properties” (i.e., *poor forecasting behavior*).

T6. Available rulemaking documents suggest that these models have not been peer reviewed (see section 1), and (based on our numerical results—see below) it seems unlikely that they have been adequately tested.

T7. The PRIA sections devoted to modeling make frequent reference to a phenomenon known as the “Gruenspecht effect.” Briefly, this effect relies on an economics-based chain of logic whereby regulations that cause new vehicle prices to increase will result in fewer used vehicles being scrapped (all else equal). Although this subject is a potential source of many discussions, the main consideration here is the role the concept has played in NHTSA’s model development decisions.

A careful reading of the NPRM and PRIA documents suggest that a desire to “mathematically mimic” this specific effect was a guiding motivation behind their modeling decisions, in contrast to, e.g., a focus on adequately capturing a more fundamental underlying behavioral structure that could legitimately allow this effect to occur (section 3.1.2). This further undermines their overall approach to modeling vehicle market behavior, contributing to its ultimate failure to produce results that are consistent with economic theory (as shown by our numerical studies).

Overall, these theory-based observations suggest many potential problems. In our view, the observations offered thus far would ensure that the NHTSA modeling approach would be very unlikely to withstand a peer review if it were, e.g., submitted for publication in an academic journal.

Summary of findings based on numerical studies. To determine whether or not these concerns can be conclusively shown to produce incorrect modeling results, and to also determine any implications for outcome measures and conclusions, we performed a number of numerical studies. These are largely based on actual computer runs from the CAFE model in CO₂ mode, using the assumptions for Table II-27 of the NPRM.

As a starting point, we established procedures for replicating NHTSA’s benefit-cost results. Next, we experimented with the model by changing input assumptions. An initial finding was that turning off the dynamic scrappage model had a major impact on the results, whereas the role of the other two models was much more limited.

For this reason, most of our numerical studies focus on the dynamic scrappage model. (This and the following two points are discussed in section 2.)

N1. Under the reference case, the Rollback costs \$200.7B less than Existing standards. Turning off the dynamic scrappage model more than erases the estimated cost differences between the Existing and Rollback scenarios. Specifically, the Existing standards cost \$14.3B less than the Rollback with the scrappage model turned off. This reverses NHTSA’s conclusion that the Rollback has positive net benefits versus the Existing standards.

The next finding involves the so-called “VMT rebound effect.” NHTSA’s analysis assumes a value that in our judgment is too high (based on recent publications and expert opinion)—see, e.g., Gillingham (2018). NHTSA assumes 20% in the 2018 NPRM, whereas they assumed 10% in the 2016 TAR (which is closer to expert consensus).

N2. Turning off both the dynamic scrappage model and the rebound effect reverses the N1 result even further: The Existing standards cost \$39.2B less than the Rollback.

The result of running the model with no dynamic scrappage and a 10% VMT rebound rate: The Existing standards cost \$27.8B less than the Rollback.

N1 and N2 suggest the importance of carefully evaluating the dynamic scrappage model in terms of its actual behavior. We performed a detailed numerical study of the dynamic scrappage model behavior for passenger cars. We started by comparing scrappage curves from multiple sources: a recent reference in the literature that performed a very similar modeling exercise (Bento et al. 2018), two different CAFE curves used by NHTSA in the 2016 CAFE model, a “No Gruenspecht” sensitivity case from the PRIA, and curves for the Existing and Rollback scenarios. These results, and results for the following points (N3 and N4) are in section 4.

N3. A high-level summary of conclusions:

N3a. The dynamic scrappage model implicitly projects “durability improvements” for recent, current and future model years that seem overly optimistic, leading to systematically lower future scrappage rates.

N3b. The model demonstrates inappropriately high sensitivity to new vehicle price increases, creating unreasonably large gaps between the scrappage rate curves for the Existing and Rollback

scenarios. *It is these gaps that drive the difference in results reported in the 2018 NPRM (both economic costs and fatalities).*

N3c. The “gap” between the Existing and Rollback scrappage curves occurs primarily between the ages of 18 and 22, a region where data have been deemed “too noisy to use” by some academic researchers, raising further questions about the model’s validity. (The implications of this are discussed below.)

These observations are consistent with our earlier concerns in T3 and T5 (that high levels of “over-fitting” lead to poor out-of-sample behavior).

N4. We devised additional numerical tests to address the noise issue. We found that:

N4a. Even though the gap between Existing and Rollback scrappage curves is inappropriately large, the *size of this gap is dwarfed by the amount of statistical error (noise) in the predicted scrappage rates.* In other words, the gap that is ultimately responsible for NHTSA’s modeling results is not *statistically meaningful* (see section 4).

N4b. These scrappage rate differences are the ultimate source of the benefit-cost differences between the Existing and-Rollback scenarios in the Agencies’ analysis. When these dynamic scrappage model rates are replaced with the scrap rates most recently developed and vetted by the Agencies, the Existing standards have positive net benefits versus the Rollback (not the negative net benefits reported by the Agencies). (See section 2.)

N4c. Additional numerical tests suggest that this high level of noise in the scrappage model’s predicted scrappage rates propagates through the rest of the CAFE model, creating large differences in the final results that compromise their validity. The failure of NHTSA to identify and test these behaviors speaks to the lack of peer review, testing, and validation that *they* should have performed to comply with their rulemaking requirements. (See Appendix A.)

These serious problems with the dynamic scrappage model raise a larger question: *What is the impact of these deficiencies on the overall economic modeling of vehicle fleet behavior?*

We conducted a series of numerical studies to answer this question (see section 5). For purposes of comparison, we also replicated all calculations on corresponding results from the U.S. Department of Energy-Energy Information Agency’s NEMS model, published in their 2018 Annual Energy Outlook (AEO). NEMS is a well-

established model that is frequently used in energy-related policy analyses (see section 3.4 and Appendix B).

N5. Numerical results highlight major deficiencies of NHTSA’s economic modeling of vehicle market behavior (in contrast to NEMS):

As discussed in more detail below, our numerical results conclusively demonstrate that the current CAFE model produces results that are wildly inconsistent with economic theory on multiple measures, whereas NEMS is consistent with economic theory on the same measures.

We summarize two specific examples. In a CAFE-model-based comparison of Existing and Rollback scenarios, the *only difference* in terms of “economic modeling inputs” for the two scenarios is: the *overall* cost of “driving vehicles” is *systematically higher* under the Existing standards, due to its pattern of increased stringency. According to economic theory, if the overall cost of a good (“driving vehicles”) *increases*, demand for the good *decreases*. For this specific analysis, the total vehicle fleet size should *decrease*. However:

N5a. The CAFE model simulates systematic *increases* in the size of the total vehicle fleet under the Existing standards (versus the Rollback) when it should produce systematic *decreases*. This violates economic theory. (See section 5.1.)

It is important to understand that, not only does this result violate economic theory, it does so via a specific mechanism: keeping a large number of very old used vehicles on the road. This is a direct result of the dynamic scrappage model behavior, which predicts very large retention of 18-22 year-old vehicles.

N5b. This problematic result is the main source of *both* of NHTSA’s claimed advantages for the Rollback: lower economic costs (higher net benefits), and lower fatality rates.

As another test, the available numerical output allowed us to compute estimates of a highly relevant economic measure for both the CAFE model and NEMS: *elasticity of scrappage with respect to new vehicle prices*.

The definition of this elasticity is: the percentage *change* in scrap rates for a one percent *increase* in new vehicle prices. For example, a value of *-1* indicates that, if new vehicle prices were to *increase* by 1% (all else equal), scrap rates would *decrease* by one percent.

Theory suggests that *scrap rates should go down when vehicle prices increase*, i.e., that the elasticity should be *negative*. The recent literature provides elasticity estimates for *used* vehicle prices (not new vehicle prices) in the range of *-0.4* to *-3*. Although it would be preferable to have them for new vehicle prices, these provide a

reasonable baseline for comparison. For our purposes we use the range 0 to -3.5. (The fact these elasticities are generally unavailable is additional evidence suggesting the inappropriateness of NHTSA's approach—see section 5.2 for more discussion).

N5c. The vast majority of elasticity estimates from NEMS falls into a reasonable range based on theory (0 to -3). In contrast, the CAFE model elasticity estimates show wild variation, taking on very large values (both positive and negative) with about half of the values being *positive* (a violation of theory). In other words, the CAFE model elasticity measures (and therefore the underlying results) are *inconsistent with economic theory*.

The reasons for these violations are easily traced to previously identified problems based on theoretical considerations: as a “system,” NHTSA's economic models consist of single-equation aggregate-level time-series projections with almost no structural/behavior-related factors suggested by theory or more detailed structural modeling approaches (T1). The new vehicle sales and scrappage model equations themselves have no structural connections that capture interactions between, e.g., the new and used vehicle markets (T2). For more details, see section 3.1.2.

Why does the NEMS model produce superior results that are consistent with economic theory? It adopts various aspects of the modeling principles reviewed earlier—see section 3.4 and Appendix B for details.

Concluding Comments.

The analysis and evaluation in this report rigorously establish the deficiencies in the Agencies' economic modeling approach, including the fact that their results violate the OMB requirement that regulatory analyses *must be based on measures that are consistent with economic theory*.

We also provide details on the multiple reasons for these failures, which can be largely traced to the dynamic scrappage model. This is very important because removing the scrappage model and replacing it with scrappage curves developed for the 2016 analysis *yield benefit-cost results that reverse the conclusions in the 2018 NPRM* (even with a 20% VMT rebound rate).

We also emphasize that our findings are *narrowly limited in scope*, and are not intended to imply that all other aspects of the CAFE model are problem-free.

Finally, there is one other aspect of the NPRM/PRIA that bears mentioning. A careful reading of the NPRM and PRIA suggests that NHTSA was fully aware of many of the problems and deficiencies associated with their modeling approach, yet they decided to proceed anyway. This raises serious questions about why this course of action was taken (see section 3.6).

1. Introduction

On August 24, 2018 NHTSA and EPA officially published their Notice of Proposed Rulemaking (NPRM) for fuel economy/GHG regulations in the Federal Register. The NPRM proposes to rescind the previously established regulations for 2022-2025 (the so-called *Existing standards*) in favor of a policy that requires no further improvements after 2020 (*Rollback*).

The NPRM, as well as a Preliminary Regulatory Impact Analysis (PRIA) and other documents, have been made available at <https://www.regulations.gov/docket?D=NHTSA-2017-0069>. These lengthy documents include an agency analysis to support and justify the proposed Rollback. They include, in part, a quantitative analysis of benefits and costs that purports to provide a meaningful comparison of the Existing versus Rollback scenarios, and which alleges the Rollback scenario to be superior. Portions of this analysis rely on modeling economic effects related to the future behavior of the vehicle market in response to the policies. The introduction to Chapter 8 of the PRIA (page 937) indicates that measuring economic benefits and costs...

“... are important considerations, because as Office of Management and Budget Circular A-4 states, benefits and costs reported in regulatory analyses must be *defined and measured consistently with economic theory*, and should also reflect how alternative regulations are anticipated to change the *behavior of producers and consumers* from a baseline scenario.⁴⁷⁹ In this analysis, those include *vehicle manufacturers, buyers of new cars and light trucks, and owners of used vehicles*, all of whose behavior is likely to be affected in complex ways by the proposed action to adopt less strict CAFE and CO₂ emission standards for future years.” [Emphasis added.]

The purpose of this report is to review and evaluate the economic models used by NHTSA to produce the PRIA/NPRM analyses and results that claim to satisfy these requirements.

The reasons for doing this require additional context. Over multiple cycles of rulemakings, NHTSA has used its own modeling system (the CAFE Compliance and Effects Model, also referred to as “the CAFE model” or “the Volpe model”) for performing a variety of quantitative analyses.³ In prior rulemakings, the CAFE model was largely limited to projecting how manufacturers might be able to comply with CAFE standards, given the feasibility and costs of using various technologies. In the past, the approach to addressing economics-based factors could be characterized as “conservative,” given the notable difficulty of modeling and predicting the behavior of a complex market with many stakeholders.

³ See <https://www.nhtsa.gov/corporate-average-fuel-economy/compliance-and-effects-modeling-system>

However, as will be discussed in more detail below, NHTSA has made a dramatic shift in its approach since the midterm review in 2016. In a very short time period NHTSA made major changes to its CAFE model, incorporating never-before-used models for forecasting the future behavior of the vehicle market, and with no (readily apparent) peer review.

Specifically, see USDOT-NHTSA (2018) for the most recent peer review of the CAFE model. The focus of the review appears to have been primarily technology issues, which also was reflected in the expertise of the panel. Although the economic modeling did not receive much attention, in response to one reviewer's question about the likely impact on new vehicle sales of technology cost increases, the following response was (page 303):

“The model has been updated to including [sic] procedures to estimate impacts on new vehicle sales, and on older vehicle scrappage. Model documentation will be revised to document these new methods, and a new Regulatory Impact Analysis will discuss the development of corresponding model inputs.”

This peer review was published in July 2018, not long before the rulemaking. This response suggests that updating the documentation and providing the materials in the PRIA is the first opportunity for any actual external review.⁴

An important consequence is that the results and conclusions developed during the previous rulemaking have been dramatically reversed. Further exploration reveals that this reversal is not due to some important new finding or additional data, but is a direct result of NHTSA changing its analysis approach. Because of the direct impact on results and conclusions, this report begins by reviewing relevant details of the benefit-cost measurements involved (section 2). Section 2 shows that turning off the dynamic scrappage model and replacing its results with the scrappage schedules recently developed by NHTSA completely *reverses* the net-benefit result reported in the NPRM, i.e. results from dropping the scrappage model show that *the Existing standards have higher net benefit than the Rollback*.

This result by itself would not be consequential if, as the Agencies might contend, the new economic modeling approach were an “improvement” over their previous analysis methods. However, this report provides a thorough review of these models and finds that NHTSA's newly introduced models are wholly inadequate for their intended purpose, falling well below the standard articulated above. All of the models rely on the application of time-series-based approaches that are appropriate for short-term forecasting of trends under “stable market conditions,” but are unlikely to satisfy the requirements for policy analysis.

⁴ A global document search for the word “scrap” located only the cited response, i.e., there was no other discussion in the peer review related to scrappage, which is a major change in modeling approach, and the primary subject of this report.

To be more specific, NHTSA has added three new economic models to project the following three behaviors: future new vehicle sales, high-level shifts in market share between cars and light-duty trucks (“fleet share”), and scrappage of used vehicles. These are combined to simulate evolution of the future US light-duty vehicle fleet. Although all three models have clear deficiencies, our analysis concludes that the one with the largest impact is the “dynamic scrappage model” [DSM], and it will therefore be singled out for the most scrutiny.⁵

Section 3 begins with a more detailed discussion of these models, and also highlights the fact that NHTSA’s development of its models seems primarily oriented toward ensuring that they “mathematically mimic” a phenomenon known as the “Gruenspecht effect.” The models are designed to create a type of “correlation” between new vehicle sales and used vehicle scrappage, rather than capture meaningful behavioral effects and market structure based on theory. To provide a well-documented baseline for what such theory-based models could look like, we review theory and methods for economic modeling of vehicle markets in section 3.2. The remainder of section 3 reviews important, relevant references in the literature and provides an evaluation of the dynamic scrappage model based on theoretical considerations. Troubling aspects of the model development process are also discussed.

Section 4 evaluates NHTSA’s scrappage model in more detail by using numerical results to directly demonstrate its multiple deficiencies. Section 5 evaluates the behavior of their total “system” when simulating the future behavior of the vehicle market, based on a comparison of numerical results from the CAFE model and the National Energy Modeling System (or “NEMS,” described in section 3.4 and Appendix B). This comparison reinforces and extends insights from sections 3 and 4. Section 6 summarizes the report’s conclusions.

The critically important findings are: the economic modeling approach produces results that are highly inconstant with the requirements of economic theory (violating Agencies’ requirements for regulatory analysis). Although the overall approach is highly flawed, the specific failings of the dynamic scrappage model appear to be the primarily source of the errors in the regulatory analyses that wrongly conclude that the Rollout has higher net benefit than the Existing standards.

Finally, although it requires some careful reading between the lines, the lengthy discussions and documentation provided in the PRIA/NPRM reveal that, in many

⁵ However, it is important to note that there are a variety of other economics-related issues. For example, NHTSA has opted to use an inappropriately high estimate of the “VMT rebound effect” that is at odds with the most current literature, and the judgment of most experts. These and other issues are outside the scope of this report, so any lack of discussion on other economics-related issues should not be interpreted as a judgment of acceptability.

instances, agency staff understands and recognizes that the methods employed (and, in particular, the data) have major shortcomings that render them inadequate for their intended purpose. The primary justification offered is typically based on practicality: limited access to data and other resource constraints could only support the type of methods that were employed. This, and other aspects of the supporting material contained in the NPRM/PRIA are also discussed. (See section 3.6.)

2. Review of Benefit-Cost Measurements

As part of the earlier 2016 midterm review, EPA produced its Technical Assessment Report (TAR), which includes a chapter (Chapter 13) “Analysis of Augural CAFE Standards” performed by NHTSA using the version of the CAFE/Volpe Model available at that time. The chapter’s introduction includes a review of the CAFE model’s history, as well as a concise and informative description of its functionality:

“...NHTSA designed the model with a view toward (a) detailed simulation of manufacturers’ potential actions given a defined set of standards, followed by (b) calculation of resultant impacts and economic costs and benefits. The model is intended to describe actions manufacturers could take in light of defined standards, estimated production constraints, and other input assumptions and estimates, not to predict actions manufacturers will take.”
(page 13-2)

The results from item (b) were used to produce the benefit-cost comparison of the Augural and Rollback scenarios reproduced in Table 2-1 (Table 13-25, 2016 TAR, page 13-103). The result: The Augural standards yield \$85B in net benefits *over* the Rollback scenario. Table 2-1 shows one way that benefits and costs can be separated into different categories (although the sign convention can be confusing). The Augural standards result in higher costs (primarily technology costs), but also higher benefits (primarily fuel savings). Importantly, estimated benefits outweigh costs under the Augural standards.

Table 13.25 Estimated Present Value of Costs, Benefits and Net Benefits (\$b) Over the Lifetimes of MYs 2016-2028 Vehicles Using 3 Percent Discount Rate (2013\$)

	MY 2022 - MY 2025 Augural Standards		
	Light Truck	Passenger Cars	Total
Social Costs			
Technology Cost	42	45	88
Maintenance Cost	2	2	5
Crashes, Fatalities, Congestion, Noise	-3	9	6
Social Benefits			
Pre Tax Fuel Savings	64	56	122
Refueling Time Savings	3	3	6
Energy Security	5	4	9
Social Cost of Carbon Emissions ¹	14	12	27
Increased Mobility	5	4	9
Conventional Pollutants	6	5	11
Net Benefits	55	28	85

¹ [Social cost of carbon to be added]

Table 2-1. Excerpt from 2016 TAR: Net-Benefit Analysis

However, the net-benefit analyses in the 2018 NPRM/PRIA lead to dramatically different results, and an *opposite* conclusion (i.e., that the Rollback produces larger net benefits). To demonstrate this, and to provide background on the main assumptions used for analyses in this report, we compute net-benefit results using the 2018 CAFE model. Our analyses use the case from Table II-27 of the NPRM (Federal Register, page 43065, not shown here). Economic costs and benefits are computed using the 2018 CAFE model in “CO₂ mode” with a discount rate of 3% (plus some minor spreadsheet calculations). In the remainder of this report, results from our CAFE model-based calculations will typically be reported for two scenarios: Existing standards, and Rollback.

See the first column in Table 2-2. First, the reporting convention is different from Table 2-1: All figures are reported as net benefits for the Rollback versus the Existing standards (equivalent to subtracting the costs for Existing standards from Rollback costs), i.e., a positive number indicates that the Rollback is “better.”⁶ Second, we needed to confirm that we could reproduce the figures the NPRM. After some experimentation, we were able to do so.⁷

⁶ The NPRM explains in detail that the Existing standards (not the Rollback) are used to define the policy “baseline” (see, e.g., NPRM page 43003).

⁷ One challenge in performing this analysis was a lack of clarity in the NPRM and PRIA regarding how the results in many tables were calculated. In this case, the

	CAFE Model	2016 TAR
Tech Cost	259.8	88
Maintenance Cost		5
Crashes, Fatalities, Congestion, Noise		6
Reduction in external costs from lower vehicle use	62.4	
Pre-tax fuel savings	-143.8	-122
Mobility Benefit	-69.5	-9
Refueling Benefit	-9.4	-6
GHG emissions	-4.7	-27
Criteria pollutants	-0.8	-11
Petroleum externality	-11.9	-9
Total [excluding accident]	82.1	-85
Total Fatal Crash Costs	94.0*	
Total Non-Fatal Crash Costs	147.0*	
Non-Rebound Fatal Crash Costs	46.2*	
Non-Rebound Non-Fatal Crash Costs	72.3*	
Net Benefits	200.7	-85

* Only the non-rebound crash-related costs are used to compute net benefits—see the text.

Table 2-2. Comparison of 2016 TAR and 2018 NPRM Analyses: Estimated Net Benefits from Adopting the Rollback Proposal

Although some basic input assumptions are not strictly the same as in the 2016 TAR⁸, we compare (to the degree possible) results from Table 2-1 (with adjusted signs) to corresponding measures from the NPRM *for purposes of illustration only*. There are some notable differences. The net benefit for Tech Cost (positive in both analyses) is much larger in the NPRM. Although this is an important difference, it is outside the scope of this report: We do not evaluate the portion of the CAFE model that simulates manufacturers’ technology choices. All of the remaining differences are due to changes in NHTSA’s approach to modeling vehicle-related market behavior.

CAFE model was run with the “Fleet Analysis” option turned on, and results from the Annual Societal Costs output file were filtered to include MY1977-2029, and Calendar Year less than 2070 (corresponding to a 40 year lifetime for MY2029 vehicles). This allowed us to reproduce the figures published in the NPRM for the case we are using. Our result (\$200.7B) matches the NPRM result in Table II-27. However, NPRM Table VII-51 (the same results using a different format) reports \$200.8B.

⁸ The base year and vehicle fleet in the 2016 TAR are 2015, versus 2016 in the current CAFE model. Some economic projections (e.g., GDP growth, fuel price projections) are not going to be exactly the same. As noted in the text, our computer runs use CO₂ mode rather than CAFE mode.

Other notable differences between the NPRM and 2016 TAR are the crash-related costs. In the 2016 TAR, costs due to crashes, congestion, and noise are combined, resulting in a relatively small number. However, in the NPRM crash-related costs are much larger. The reasons for this will be clearly established later in the report.

However, the crash-related costs raise an issue related to an economic “parameter” that that must be discussed for purposes of completeness: the *rebound effect*. According to economic theory, if the cost of a good (or service) goes *down* (all else equal), then demand for it will go *up*. If fuel economy for new vehicles goes up (e.g., due to the regulation), then the per-mile cost of driving goes down. Theory therefore suggests that new vehicle purchasers would be expected *to drive more miles* (than they would have otherwise). This increase in driving (were it to occur) is the rebound effect.

Assuming that the rebound effect produces increased driving for certain individuals, this means that their expected number of crashes goes up. However, it must be recognized that this increase in crashes will have occurred because drivers *decided* they would be *better off* by driving more. Therefore, any crashes associated with these extra miles are due to consumer choice. Consumers are assumed to take this risk into account when making their decision to drive more miles. Therefore, any costs associated with these crashes are not directly attributable to the regulation that caused the increase in fuel economy. For this reason, only “non-rebound crash costs” are included in the regulatory analyses.

One final remark about the rebound effect: The Agencies’ analysis in the NPRM assumes a value that in our judgment is too high (based on recent publications and expert opinion). For example, see Gillingham (2018). He concludes that the range of central estimates is 8.1% to 14.1%, where the 8.1% is arguably preferred because it is based on a methodology that uses two odometer readings. A reasonable estimate would therefore be 10% (which happens to be the value used in the 2016 TAR).

With this as background, we now explore the role NHTSA’s new economic modeling approach plays in determining the NPRM net-benefit results. The CAFE model includes options for turning off various modules, and, in addition, values in the parameter input file can be changed to produce alternative sets of results.

Table 2-3 shows results from different CAFE model runs with two key effects turned on and off: the rebound effect, and the dynamic scrappage model. When the rebound effect is turned off there is some reduction in net benefit for the Rollback standard (but not especially large).

However, when the dynamic scrappage model is turned off, *net benefit switches from positive to negative*, i.e., the Existing standards provide more net benefit than the Rollback (\$14.3B). This is true even though the 20% rebound has been left in place. When both are turned off, the net benefit of the Existing standards increases by

another \$25B (roughly what would be expected from the difference in results between the first two columns). Using our preferred rebound effect of 10%, the Existing standards yield a net benefit of \$27.8B over the Rollback (full figures not shown).

	Reference Case	No Rebound	No Dynamic Scrap	No Rebound + No DS
Tech Cost	259.8	259.8	259.8	259.8
Pre-tax fuel savings	-143.8	-194.0	-185.1	-230.7
Mobility Benefit	-69.5	0.0	-63.7	0.0
Refueling Benefit	-9.4	-11.9	-9.9	-12.2
GHG emissions	-4.7	-6.3	-6.1	-7.6
Criteria pollutants	-0.8	-4.1	-7.1	-10.2
Petroleum externality	-11.9	-16.0	-15.2	-18.9
Reduction in external costs from lower veh use	62.4	27.9	25.7	-6.8
Total [excluding accident]	82.1	55.4	-1.5	-26.4
Total Fatal Crash Costs	94.0	46.2	40.1	-5.0
Total Non-Fatal Crash Costs	147.0	72.3	62.7	-7.8
Non-Rebound Fatal Crash Costs	46.2	46.2	-5.0	-5.0
Non-Rebound Non-Fatal Crash Costs	72.3	72.3	-7.8	-7.8
Net Benefits	200.7	174.0	-14.3	-39.2

* Only the non-rebound crash-related costs are used to compute net benefits—see the text.

Table 2-3. The Impact of Rebound and Dynamic Scrapage Effects on NPRM Net Benefit Results From CAFÉ Model

Similar explorations for the other two components of NHTSA’s new economic modeling approach (new vehicle sales, and dynamic fleet share) revealed that (in contrast to dynamic scrappage) they have relatively little impact on the bottom-line net-benefit results. For this reason, the dynamic scrappage model receives most of the attention in our evaluation.

To summarize:

This exercise demonstrates the critical role played by the dynamic scrappage model in determining the outcome and conclusions of the NPRM regulatory analysis⁹.

⁹ The difference in Tech Costs between the Existing and Rollback scenarios are much larger than in the 2016 TAR, and (according to Table 2-3) are unaffected by the behavior of the vehicle market. If these turn out to be too large, then the net benefit

To elaborate further: When the dynamic scrappage model is turned off, the CAFE model reverts to using scrappage projections developed by NHTSA very recently. More specifically, these were used in the 2016 TAR only a short time ago, so they represent their most recent approach to determining scrappage rates prior the current rulemaking. Because the impact of this modeling change is not only large but also consequential, it is clear that the decision to make such an important change must be the correct one. But, how can this be evaluated?

The Agencies' requirements for performing regulatory analysis were reviewed in the introduction, and are useful for identifying evaluation criteria related to this type of economic modeling, which is technically challenging, and requires expertise in multiple areas. Within this context, considering the range of issues involved suggests the following initial questions:

1. How suitable is NHTSA's dynamic scrappage model when evaluated from the perspective of *economic theory* and/or *methods that have been previously established in the literature*?
2. Do the methods, data, and other factors employed by NHTSA yield a dynamic scrappage model that produces *credible and reasonable results*?
3. Given that the scrappage model is a *statistical* model whose parameters are *estimated* based on *historical data*, what are the implications for this model's performance with respect to the *precision* and *accuracy* of its forecasts?

Specifically, even under ideal conditions, such models produce predictions that are subject to a certain amount of *statistical error*. How large is this error? And, when these errors are propagated through the rest of the CAFE model, what is the impact on the final results?

This report provides answers to the first two questions, as well as most of the third. For the third question, section 4 investigates the issue of statistical error, and the precision of *predictions* from the scrappage model.

The very last question in item 3 (error propagation) raises the concern of how the *final* cost estimates produced by the CAFE model will be affected when a statistical model with prediction error (such as the dynamic scrappage model) is embedded within it. This issue is a relatively complex one, and during our investigation we produced some preliminary results on this subject. These are discussed separately in Appendix A to emphasize that they are in no way required to support the main conclusions in this report. At the same time, we wanted to document the potential impact of this issue, and highlight the fact that it is one of many that *should have*

differences in favor of the Existing standards would be even larger. However, this issue is outside the scope of this report.

been the subject of rigorous testing and validation before deciding to use this model for an important regulatory analysis.

3. Economic Modeling of Consumer Vehicle Markets for Policy Analysis

This section includes background material on theory and methods for economic modeling of consumer vehicle markets in the context of performing policy analysis. The purpose is to establish a well-documented basis for evaluating NHTSA's new models. However, before presenting this material, sub-section (3.1) reviews additional details of NHTSA's models so they are available for consideration while reading the remainder of the section. In addition, an economic argument known as the "Gruenspecht effect" has had a major influence on NHTSA's modeling decisions, so it is also reviewed.

Section 3.2 reviews a modeling framework due to Berkovec (1985), which provides a representative example of an approach with features suitable for policy analysis. It consists of an integrated system of models based on sound behavioral theory and economic principles. In fact, it was specifically created to address the problem considered here: simulating the response of vehicle markets to policy interventions such as fuel efficiency standards.

However, fully implementing this type of modeling system might be impractical for many analysts. Model development almost always requires making decisions on which simplifying assumptions to adopt, where the usual tradeoff is "improved behavioral realism" versus "reduced data and computational requirements." Analysts frequently adopt simpler (less realistic) models when faced with practical limitations. However, correctly evaluating which simplifying assumptions are acceptable requires a clear understanding and application of underlying theory. When faced with practical obstacles, one possible approach would be to start with an established and well-understood modeling framework with good theory-based properties (like Berkovec's), and develop alternative options by applying simplifying assumptions to the framework. This would clarify the theoretical implications of making tradeoffs, including what might be unacceptable.

In this way, section 3.2 also serves to provide a baseline for comparing models in this report. The remainder of the section is devoted to reviewing specific models and research results from the literature that are directly relevant to various aspects of NHTSA's modeling approach (particularly with respect to scrappage). Section 3.3 reviews the aggregate-level vehicle market forecasting approach of Greenspan and Cohen (1999). This is highly relevant, since it is directly comparable to what NHTSA is attempting to do with its approach. Section 3.4 provides background on NEMS from the EIA. [Add more here.]

Finally, section 3.5 discusses aggregate scrappage models in more detail, and summarizes research results from two recently published articles: Bento, et al.

(2018), and Jacobsen and van Benthem (2015). Both references specifically comment on the Gruenspecht effect. Notably, the NPRM and PRIA explicitly cite all three references reviewed here, indicating at least some level of awareness of the modeling approaches and research findings in these references.

3.1. NHTSA's Economic Models and the Gruenspecht Effect

As discussed in section 2, previous versions of the CAFE model prior to the current rulemaking were primarily focused on manufacturers' decisions on which fuel-saving technologies to add to their vehicles at the time they are redesigned. Because the simplest version of NHTSA's responsibility is to set fuel economy standards to levels that are "maximum technically feasible," it makes sense that developing technology databases and algorithms for simulating future compliance pathways for manufacturers in response to regulations would be of paramount importance. Accordingly, NHTSA has developed a highly detailed representation of this process, which we will refer to as the Manufacturer Decision Model (or MDM). This is a significant model of *economic behavior* in its own right.

However, recall that the CAFE model's *other* main function is to compute the estimated impact of these decisions on economic costs and benefits (to produce analyses shown in, e.g., Table 2-3). Computing these measures requires some representation of future consumer behavior in the vehicle market. For example, computing pre-tax fuel savings requires an estimate of how many vehicles are on the road, the distribution of ages, fuel efficiencies and fuel types within the fleet, how far they are driven, and the cost of fuel. There is a wide range of options for how these "consumer demands" might be modeled, which is the subject of later sections.

However, we first discuss the vehicle "supply side" in more detail. The vehicle market is highly differentiated, i.e., it provides a large number of offerings to consumers (who vary widely in their needs and tastes). The CAFE model maintains a representation of new vehicle offerings at a relatively complete level of detail. The model is initialized using an observed vehicle fleet for a specific *base year*. Model Year 2016 is the current base year with over 1600 vehicle offerings, fully characterized in terms of their prices, technologies, attributes, and specifications.

An individual consumer's decision options for participating in the new vehicle market are represented in Figure 3-1. The top-level decision is whether or not to "Buy a New Vehicle." The "No Buy" option subsumes all other options, e.g., participating in the used vehicle market, or keeping the current household fleet. A more detailed treatment would explicitly represent both the new and used vehicle markets (as discussed in the next section), and additional details on transaction option (e.g., replace a currently held vehicle, add a vehicle, dispose of a currently held vehicle).

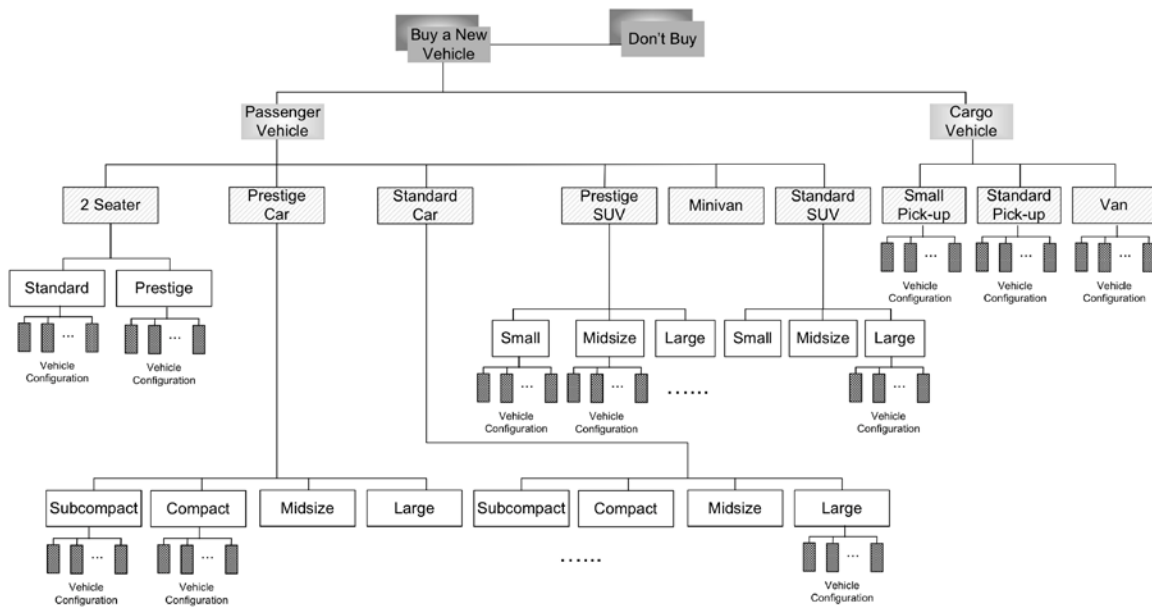


Figure 3-1. Consumer Decision Options in the New Vehicle Market

The entire section of the tree under the “Buy a New Vehicle” branch is how the supply side of the market is represented *inside* the MDM. During a CAFE model run, the MDM simulates manufacturers’ choices about what new technologies to add to their vehicles in each future model year over a specified range (e.g., from 2017 to 2050). Each specific vehicle can only be redesigned in certain years. A vehicle’s current “price” is estimated as the base year price plus all accumulated incremental costs from adding technology (i.e., costs are passed on to the consumer). Technology is added to keep the manufacturer in compliance with fuel efficiency/emissions regulations, so these are referred to as *compliance costs*.

For each year, an algorithm determines what technologies to add (and to which vehicles) so that each manufacturer’s compliance cost increases are minimized¹⁰. This means that both the selling price and the fuel efficiency of each redesigned vehicle will change from the previous year. In the real world this would be expected to cause *sales shifts*, because consumers would change their purchase decisions in response to the changes in the vehicles offered in the new vehicle market. However, this is not addressed by the MDM algorithm due to the complexity it would introduce.

Another consideration is that, in the real world, manufacturers do not necessarily need to add technology to achieve compliance: they can use *pricing strategies* to create sales shifts *among* their vehicles, i.e., they can use *cross-subsidization*. In this case, because they can decide on both redesign and pricing, they could make these decisions so as to maximize profits (rather than minimize costs) subject to

¹⁰ This is actually an oversimplification, but is sufficient for illustrative purposes.

compliance with the regulation. As before, this would require the MDM to anticipate what sales shifts *would* occur in response to changes in vehicle offerings by the entire industry.

However, this another way of saying: *NHTSA does not have the capability of modeling consumer response to changes in vehicles at the individual vehicle level.* This is the type of consumer choice modeling discussed in section 3.2.

In fact, in previous versions of the CAFE model there were no attempts to directly simulate consumer response from *within* the CAFE model at all. Instead, NHTSA relied on fixed projections of future vehicle market behavior from multiple sources for the purpose of performing the required economic cost and benefit calculations. While this might possibly be less than ideal, this approach is only a problem if, in the real world, there notable *differences* in future market behavior occur under different regulation scenarios, and, moreover, that these differences would be large enough to compromise the validity of the net benefit comparisons.

However, for the current NPRM, NHTSA abandoned its previous approach in favor of a new approach that uses the three models to be discussed section 3.1.2. As will be discussed, these three models are extraordinarily limited in terms of their aspirations to model consumer response. For example, they have no capability of modeling the types of sales shifts discussed above. So, why were they introduced?

One early finding of our review was that, when developing these models, NHTSA's primary motivations appears to have been finding a way to introduce the so-called "Gruenspecht effect" into its CAFE model. Numerous references to, and discussions of, this effect appear throughout the NPRM and PRIA documentation. Given this finding, we define and explain this effect before moving to other material.

3.1.1 The Gruenspecht Effect

When economists think about the potential impact of a public policy, they are often on the lookout for possible "unintended negative consequences," whereby a policy might "backfire" in some way so as to be (at least partially) self-defeating (e.g., the rebound effect). The *existence* of such effects are usually easily justified on the basis of high-level theoretical arguments, although empirically *proving* their existence and measuring their size are much more difficult (and frequently the source of controversy).

One such effect that may arise in discussions of new vehicle emissions regulations¹¹ is often attributed to Gruenspecht (1982, 1983), which follows this change of logic:

¹¹ These are an example of so-called "differentiated regulation," because, e.g., used vehicles are not subject to the same regulations as new vehicles.

Regulation of new vehicle emissions increases their manufacturing cost, pushing new vehicle prices *higher*. Higher new vehicle prices causes lower demand for new vehicles, and therefore higher demand for used vehicles. Higher used vehicle demand pushes *used vehicle prices higher*, i.e., they have higher market value. When used vehicles are worth more, they stay on the road longer and are scrapped at lower rates. This causes an *increase* in the emissions that were originally targeted by the regulation.

There are a number of things to consider here. First, this is a *ceteris paribus* argument, i.e., it requires every other factor to remain unchanged. Even if regulation adds costs to new vehicles, ongoing economic growth and other technological advances could cause new vehicle demand to keep rising.

Second, and more important in our view: The above argument is frequently applied to the case of fuel economy/greenhouse gas regulations, and ignores the fact that vehicle costs are increasing due to a required improvement in a vehicle attribute that has value to consumers. One possible reason for this error is a failure to recognize that Gruenspecht's original argument was formulated in the context of *criteria pollutants, not fuel economy/GHG!*

To clarify: Starting in the 1970's federal emissions standards for criteria pollutants (e.g., hydrocarbons, CO, and NOx) were established for new vehicles sold in the United States. Manufacturers were required to add emissions mitigation technologies to their vehicles (e.g., catalytic converters) that provided only *public benefits* while imposing multiple types of *private costs*. This requirement (1) drove up manufacturing costs that (to some degree) were passed on to consumers in the form of higher prices, (2) compromised the performance of some vehicles, and (3) increased average maintenance costs.

However, the case of *fuel economy/GHG emissions* has the important difference noted above. Nevertheless, the Gruenspecht effect argument is frequently repeated almost *verbatim* as though the two contexts were *identical*. There are multiple problematic references of this type in the NPRM and PRIA (to be discussed later). But more concerning is that this same error has effectively been incorporated into one of NHTSA's economic models, as discussed in the next section.

3.1.2 More on NHTSA's 2018 Economic Modeling

Recall that for the current rulemaking NHTSA added three new (sub-)models to its CAFE model: *new vehicle sales*, *dynamic fleet share*, and *dynamic scrappage*. All three are aggregate-level time-series forecasting models. For reasons described in section 2, this report focuses primarily on the dynamic scrappage model, with some discussion of new vehicle sales, and no further exploration of dynamic fleet share.

There are actually three different versions of the model to provide aggregate-level scrappage rates for each of three vehicle types: Car, SUV/Van, and Truck. Although

these are technically three different models, we will refer to “the dynamic scrappage model” for ease of presentation. Similarly, there are separate vehicle-type-specific models for new vehicle sales.

One shorthand depiction of the two models is as follows:

$$Sales_t = F_{Sales}(\overline{MSRP}_{2016} + \overline{\Delta C}_t, \overline{GDP}_t, \overline{LFP}_t) \quad (1)$$

$$ScrapRate_t = F_{ScrapRate}(\overline{MSRP}_{2016} + \overline{\Delta C}_t, \overline{CPM}_t, \overline{GDP}_t, \overline{Int}_t) \quad (2)$$

where *MSRPs* are manufacturer suggested retail prices for the 1600+ vehicles in the 2016 base year fleet, ΔC is the incremental compliance cost for all technologies added to the 2016 vehicle, and *CPM* is a measure of fuel cost per mile. The remaining three variables are macroeconomic indexes: *GDP* (gross domestic product), *LFP* (labor force participation), and *Int* (a measure related to interest rates). The subscript *t* denotes a time period (i.e., year). As indicated above, *average* measures across all the vehicles in a given fleet are used as explanatory variables.

The technical limitations of these models of these models should be obvious when contrasted with, e.g., the idea of forecasting sales shifts for *individual* vehicles discussed earlier in this section. Specifically, the highly aggregated nature of these models is clearly a potential limitation (e.g., not enough detail and/or structure to capture realistic behavior). Another concern is that the time-series approaches employed by NHTSA, while perhaps suitable for short-term extrapolation of existing trends under stable market conditions, are inappropriate for policy analysis (for reasons to be discussed in later sections).

However, to conclude here, we focus on NHTSA’s decision to use these models, and how its apparent relationship to the Gruenspecht effect. Recall that the Gruenspecht effect is only concerned with high-level aggregate behavior as described in the previous section. The “chain of logic” begins with the hypothesis that regulations create “price increases” that dampen new vehicle sales. Recall also our admonition that the Gruenspecht effect (as frequently stated) ignores the accompanying improvement in fuel economy. It so happens that the *only* vehicle attribute included in the aggregate new sales model is a proxy measure for “new vehicle price increase.” As in the misapplication of the Gruenspecht effect argument, fuel economy improvement is ignored in their model¹².

¹² To be fair, NHTSA would argue that it could not find a “statistically significant effect” associated with fuel economy improvement. However, this fact is primarily consistent with the many problems associated with their approach, not evidence that consumers do not care about fuel economy (which is known to be false based on the huge literature on this topic).

With regard to scrappage: Recall that the complete chain of logic suggests that higher new vehicle prices (due to tighter regulations) *eventually* result in *lower scrappage rates*. NHTSA attempts to reproduce this effect by *directly* including “new vehicle prices” in their scrappage rate equations. But, this is *not* an actual implementation of the behavioral effects enumerated in the statement of the Gruenspecht effect. This appears to be an attempt to ensure that, taken together, the two models might “mathematically mimic” the overall behavior hypothesized by the Gruenspecht effect.

To be clear: The best that can be said is that, because the two models *share* a new vehicle price-related variable *in common*, the outputs of these models might somehow be *correlated*. In this regard, a widely known admonishment in introductory quantitative analysis courses is: “Correlation is not causality.” As will be shown in section 5, this “stitching together” of two models via a correlation mechanism is insufficient to produce vehicle market behavior that is consistent with economic theory.

The material in the next section will help clarify how behavioral effects can be correctly captured using theory-based modeling approaches.

3.2. Economics-based Modeling of Vehicle Market Behavior: Berkovec (1985)

The Berkovec modeling framework is reviewed here because it is a good example of a theory-based structural modeling approach for the specific problem we are addressing: Economic modeling of vehicle market behavior. His motivation (as described in the paper’s introduction) is to evaluate policies that

“...work in similar ways on the automobile market in that they modify the attributes (including prices) of the new vehicles available to consumers, thereby leading to different consumer purchases of new vehicles.

The complexity of the automobile market makes it difficult to evaluate the effects of these policies, especially in the short run. Automobiles are highly differentiated durable goods with variable lifetimes. *If an ‘improvement’ (e.g., fuel efficiency) is mandated in the offerings of new cars **at sufficiently high cost**, it will induce a demand shift away from new vehicles and cause existing vehicles to be more highly valued and longer lived. This may cause the fuel efficiency of the vehicle stock to fall in the short run **if older cars are sufficiently less efficient than new cars**.* Empirical estimates of market response are needed to evaluate the effectiveness of regulations.” [Emphasis added.] (Berkovec 1985, pp. 195-196.)

Italics were added to identify his concern about the possibility of a type of Gruenspecht effect. However, we further emphasize in bold some phrases to highlight the *care* with which this statement was made. Note that his version does not assume that increases in manufacturing costs (due to regulation) will

automatically cause used vehicle scrappage rates to decline. The outcome depends on the details of other factors.

Proceeding to his framework: The two main economic actors are *Households* and *Manufacturers*¹³.

The specific structure and features are as follows¹⁴:

- Overall automobile market behavior is captured for a sequence of interconnected time periods.
- Vehicles enter the market via annual offerings from Manufacturers.
- Vehicles exit the market by being scrapped.
- In each period, Households decide how many vehicles to own, and which ones (e.g., which vehicle classes, whether they are new or used, and if used, what vintage(s)).
- Changes in Household vehicle holdings are captured when going from one period to the next.
- Households make decisions based on their vehicle preferences.
 - Vehicle preferences are based on vehicle attributes (e.g., capital cost, fuel operating cost, vehicle type and size)
 - They can vary across household types due to differences in household characteristics (e.g., income, household size, age, education, employment status, residential location).
- An economic equilibrium occurs that balances supply and demand for vehicles in each period.
- Equilibrium is achieved through the setting of market-clearing prices for all vehicles.

There are T time periods, indexed by $t = 1, \dots, T$.

In every period t there are N vehicle ‘types’ available. Vehicles vary by vehicle class, indexed by $j = 1, \dots, J$, and vehicle age, indexed by $a = 0, \dots, A$ (i.e., vehicles of age $a = 0$ are new vehicles). For simplicity, we assume that all classes are available in all periods, and that vehicles of age A in period t are retired (scrapped) in period $t+1$. This means that in any period the total number of classes $N = J*(A+1)$, and that a vehicle type can be alternatively represented by $n = aj$.

¹³ In the real world, many businesses and governmental entities also purchase light-duty vehicles. However, when discussing economic modeling, these are frequently ignored. However, see section 3.3.

¹⁴ Selection of which modeling features to include is a matter of professional judgment, determined by factors such as the purpose of the model, data availability, etc. For example, this framework excludes details such as decisions on the use of public transportation, and how far to drive each vehicle.

For each time period t , define the following:

- P^t = the vector of prices for the N vehicles;
- X^t = a matrix of vehicle characteristics;
- Z^t = a matrix of household characteristics.

The key (aggregate) market quantities are defined as follows¹⁵:

- $R_{ajt}(P^t, X^t)$ = number of vehicles of age a and class j being scrapped (or retired) during period t ;
- $D_{ajt}(P^t, X^t, Z^t)$ = consumer demand for vehicle aj in period t ;
- $S_{ajt}(P^t, X^t, Z^t)$ = production quantity of vehicle aj in period t ;
- Q_{aj}^t = existing stock of vehicle aj in period t .

The general equation balancing supply and demand is [Berkovec (1985, equ. 1)]:

$$S_{ajt}(P^t, X^t) + Q_{aj}^t = R_{ajt}(P^t, X^t) + D_{ajt}(P^t, X^t, Z^t) \text{ for all } a, j. \quad (3)$$

However, additional considerations place some limitations on what values are allowed. During period t the *existing stock* of new vehicles is 0, i.e., $Q_{0jt} = 0$. Similarly, the only non-zero values for $S_{ajt}(P^t, X^t, Z^t)$ occur when $a = 0$. Finally, for our purposes we also assume that only vehicles in existing stock can be scrapped during period t , i.e., $R_{ajt}(P^t, X^t) = 0$. Given these restrictions, (3) can be rewritten as:

$$S_{0jt}(P^t, X^t) = D_{0jt}(P^t, X^t, Z^t) \quad \text{for all } j, \quad (4a)$$

$$Q_{aj}^t = R_{ajt}(P^t, X^t) + D_{ajt}(P^t, X^t, Z^t) \quad \text{for all } a > 0, j. \quad (4b)$$

The aggregate demand from households during period t becomes the total vehicle stock for period $t+1$, i.e., $Q_{(a+1)j}^{t+1} = D_{ajt}(P^t, X^t, Z^t)$. This aggregate demand is determined by adding up results from more *detailed* behavioral models that “simulate” vehicle-related choices by many different household types. This can be shown by decomposing the matrix (Z^t) into individual vectors of characteristics for S household segment (i.e., Z_s^t for $s = 1, \dots, S$), so that aggregate demand is given by:

$$D_{ajt}(P^t, X^t, Z^t) = \sum_{s=1}^S D_{ajt}(P^t, X^t, Z_s^t) \quad \text{for all } a, j. \quad (5)$$

The features discussed thus far are depicted in Figure 3-2.

¹⁵ Berkovec’s framework also includes a numeraire good (“money”) that represents the value of all other goods in the economy (denominated in dollars). This can be used to represent income for households, household expenditure on non-vehicle goods, inputs (investment) to pay for materials used in producing vehicles, and manufacturer profits. In addition to vehicle supply and demand, total money supply is balanced in the system. We omit this feature for simplicity.

Manufacturers and Vehicles. The tree structure inside the Manufacturers box represents the detailed vehicle offerings corresponding to Figure 3-1. The multiple arrows leaving the right-hand boundary denote that all of these individual vehicle offerings are made available to Households. The flow of supply to the Households is denoted by the second set of arrows, accompanied by $S_{ajt}(P^t, X^t)$. Households have full information on prices and vehicle attributes.

Vehicle Stock. Although Households own previously purchased vehicles, Vehicle Stock is represented separately to model the used vehicle market. The multiple trees depict used vehicles of previous model years. In addition to new vehicle supply $S_{ajt}(P^t, X^t)$, current used vehicles Q_{aj}^t ($a = 1, \dots, A$) can be bought, sold, or kept during period t .

Household Demand and Scrapage. Figure 3-2 shows S Household segments, each with different preferences (assumed to be determined by their demographic vector Z_s^t). Each Household type has a demand for all vehicle classes and ages (with j suppressed), and total demand is determined by equation (5). When Household segments determine their vehicle ownership, some vehicles are purchased new, some are purchased used, some used vehicles are sold, and some are scrapped ($R_{ajt}(P^t, X^t)$). At the end of period t , the Vehicle Stock is updated for period $t+1$ (Q_{aj}^{t+1}).

In going from period t to $t+1$, Manufacturers redesign their vehicles for the next model year, and Households are “aged” based on demographic and economic trends. As already described, in Berkovec’s framework a vector of prices is identified that clears the market. Demand shifts occur in each period due changes in many possible factors: the distribution of household types, the vehicle attributes of new vehicle offerings introduced by manufacturers, fuel prices, the characteristics of the remaining used vehicle fleet, etc.

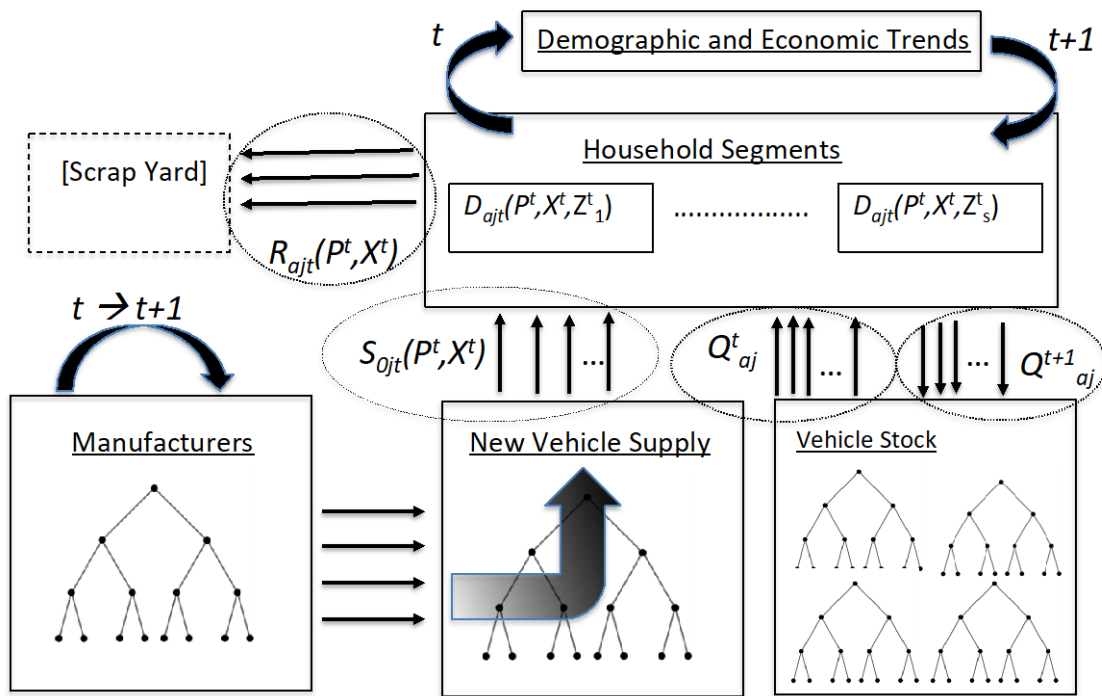


Figure 3-2. Berkovec (1985) Vehicle Market Modeling Framework

There are many modeling options for determining the various quantities represented in Figure 3-2, some of which we discuss now.

Household Choices. Household demand ($D_{ajt}(P^t, X^t, Z_s^t)$) can be simulated using *discrete choice models*. There are many possible approaches. One approach (called a *holdings model*) is to treat households as “re-deciding” what vehicles to own in each year. Multiple decisions can be depicted using a tree structure.

See Figure 3-3: The top-level choice is whether to own 0, 1, or 2 vehicles¹⁶. Under the branches for 1 or 2 vehicles, additional choice options are enumerated as vehicle “portfolios”. Under the branch for 1-vehicle, one could imagine two additional branches: new or used. Under “new” would be an additional tree structure (as in Figure 3-1). Under “used” there would be A additional trees (for ages $a = 1, \dots, A$). Alternatively, all possible combinations of vehicle options could be depicted (as in Figure 3-3, which provides two example portfolios under each branch for 1 and 2, respectively.)

Discrete choice models are estimated based on, e.g., data on actual household vehicle ownership from a survey. They compute the probabilities for all branches at the “bottom” of the tree. Berkovec (1985) uses this approach, and can be consulted

¹⁶ Although 0-vehicle households clearly exist, they are frequently omitted for practical reasons.

for more details. The literature on discrete choice models is voluminous, but one reference devoted to this type of vehicle choice modeling is Train (1986).

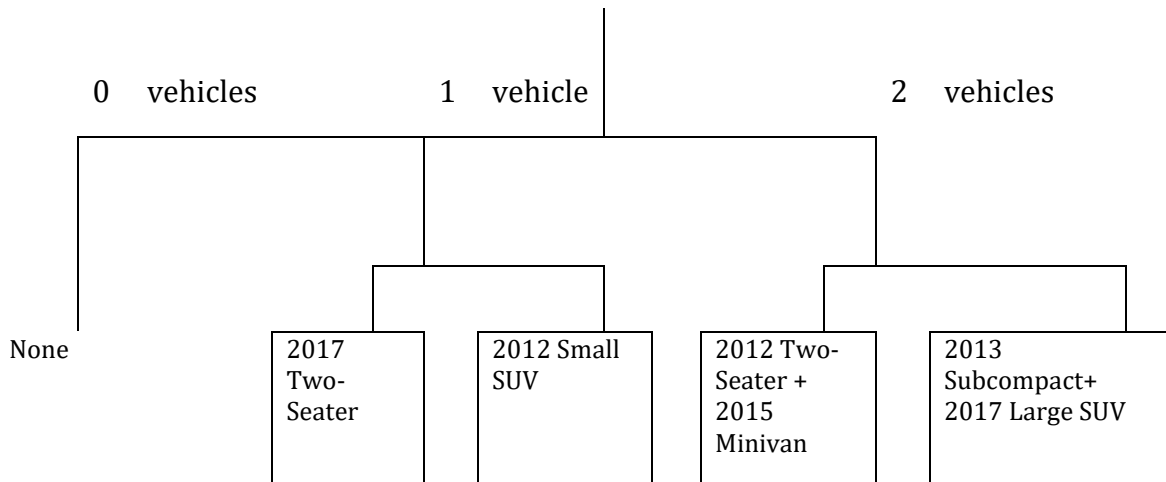


Figure 3-3. Nested Decision Structure for a Household Holdings Model

As has been described previously, preferences are a function of household characteristics and vehicle attributes. These are typically represented by a “utility function” with weights to represent their relative importance. For example, in Berkovec (1985) he estimates a vehicle type choice model for 1-vehicle households. The coefficient for expected vehicle capital cost (in thousands) in the case of low-income households is -2.24, and for high-income households is -0.653 (indicating they are less price sensitive). The coefficient of fuel operating cost (in cents per mile) is -0.199. All of these coefficients are negative, indicating that higher values lower the utility of the vehicle option. Similarly, the coefficient on age is negative, and the coefficient on a “seat space” variable is positive, with a higher coefficient for larger households (5 or more members).

An important point to emphasize about this approach is: Demand for new and used vehicles is explicitly addressed in terms of bottom-up household choices that include their decisions about owning new versus used vehicles (which are substitutes), and these decisions are determined on the basis of full information on prices and vehicle attributes for all possible available options.

Scrappage. As shown in equation (3) and Figure 3-2, a complete model for the vehicle market takes into account how used vehicles are eventually scrapped/retired. One approach is to ensure that scrappage is somehow inferred as part of the Household decision process (e.g., when demand for old vehicle type is less than supply, then the remainder are scrapped). Another approach is to explicitly model scrappage (a Berkovec does). The subject of scrappage models is obviously a critical one for this report, and a more detailed discussion of scrappage models appears in section 3.5.

Manufacturer Decisions. For completeness, we review here the range of possible manufacturer decisions (see section 3.1) that could be incorporated into a behavioral model. The CAFE model focuses on redesign decisions, where manufacturers add fuel saving technologies to comply with emissions standards. Pricing strategies are also a possibility, but these are not included in the CAFE model.

In the above framework the equilibrium process ultimately determines the prices, and these could be viewed as *conditional* on manufacturers' design decisions. However, in theory, manufacturers would have full information about the existence of both the equilibrium mechanism and household choice models, so it could determine exactly what demand shifts *would* occur in response to redesign decisions. This would allow manufacturers to evaluate whether or not their decisions would result in compliance, and also allow them to optimize their decisions on the basis of some specified criterion (e.g., profit maximization or cost minimization). For an example of a modeling approach that includes both design and pricing decisions by manufacturers, see Liu, Greene, and Bunch (2014).

However, such models are rather complicated. In fact, the system actually implemented by Berkovec assumes a short-run planning horizon, i.e., manufacturers' design decisions are assumed to have already been made: production quantities and prices are therefore determined by the supply and demand process alone. However, the framework itself does not preclude more complex versions of the manufacturer decision model.

Having commented on this issue, there are practical alternatives for evaluating policies that do not require a complex manufacturing decision model. A demand-oriented model using the Berkovec framework can be used to perform iterative scenario analyses, where a user supplies possible design decisions and then simulates the outcomes. This is the approach used with the CARBITS model developed for the California Air Resources Board, and the DynaSim model used by the California Energy Commission for its biannual transportation demand energy forecast—see, e.g., Bunch (2009).

Remarks. The purpose of providing this background is to demonstrate what a theory-based modeling approach for analyzing regulatory options that meets the OMB requirements in the Introduction would look like. There is large and well-established literature on this subject.

While we recognize that such a model could be difficult to implement for some researchers, *note that most aspects of the framework in this section are incorporated into the models mentioned above (CARBITS and DynaSim).* At the start of this section, we quoted Berkovec's goal of addressing the need simulate whether or not a "Gruenspecht effect" might occur in response to vehicle emissions regulations as a motivation for this framework.

From reading Chapter 8 in the PRIA, the Agencies are very forthcoming about their goal of producing such an effect in their analyses. In an effort to support and justify these efforts, the NPRM contains a section entitled “Models of the Gruenspecht Effect Used in Other Policy Analyses” (NPRM, page 43094, column 3). Here is an excerpt:

“This is not the first estimation of the ‘Gruenspecht Effect’ for policy considerations. In their Technical Support Document (TSD) for the 2004 proposal to reduce greenhouse gas emissions from motor vehicles, California Air Resources Board (CARB) outlines how they utilized the CARBITS vehicle transaction choice model in an attempt to capture the effect of increasing new vehicle prices on vehicle replacement rates.”

The Agencies do not identify any other models in this category (only CARBITS). Their apparent purpose in citing CARBITS is to provide some justification and support for their own attempt to create a model that might produce a “Gruenspecht Effect”:

“The CARB study captures the effect on new vehicle prices on the fleet replacement rates and offers some precedence for including some estimate of the Gruenspecht Effect.”

However, their description of CARBITS (while correct in some respects, but incorrect in others) is one that most readers will not understand the implications of: CARBITS is a bottom-up structural model of the type discussed *in this section*, capable of capturing household behavior in response to regulations in a manner that is consistent with economic theory. It was specifically designed to do a good job of analyzing alternative regulations by adhering to sound, theory-based modeling principles.

Because I am the designer of CARBITS, I can definitively say that the model was not specifically designed or intended to “capture the effect of increasing new vehicle prices on vehicle replacement rates.” The intent of CARBITS was to simulate the behavior of the new and used vehicle market in California under alternative scenarios. Because new vehicle price increases would also be accompanied by other attribute improvements (fuel economy, but perhaps others, depending on technology forecasts), and because of the large amount of preference heterogeneity across households captured by the model, there would be no way to know in advance exactly what the market response would be.

In contrast, the NHTSA models described in section 3.1 *do not adhere to these same principles*, so the existence of CARBITS is in no way a justification of their approach (in fact, just the opposite). Before moving to a more detailed evaluation of NHTSA’s models based on the material in this section, we review other relevant references in the literature.

3.3. Aggregate-level Vehicle Market Forecasting: Greenspan-Cohen (1999)

The modeling approach proposed by Greenspan-Cohen (1999) [hereafter, GC96] provides a useful example for comparison with NHTSA's modeling approach. This is because, in contrast to models of the type discussed in the previous section (e.g., CARBITS), GC96 is pursuing a very limited form of modeling that is highly similar to NHTSA's: Aggregate (macroeconomic) forecasting of motor vehicle stocks, scrappage, and new vehicle sales.

Recall that NHTSA's approach is based on forecasting new vehicle sales, and then predicting scrappage rates of used vehicles in order to make vehicle market projections.

GC96 takes a different approach. They also model scrappage of used vehicles. However, instead of modeling new vehicle sales, they develop projections of *total fleet size* (or "aggregate vehicle stock"). Aggregate vehicle stock can be combined with scrappage estimates to *infer* new vehicle sales.

Although projecting either one (total fleet size or new vehicle sales) can be challenging, well-known insights from the travel demand forecasting literature suggest why projecting total fleet size is more tractable. Households own vehicles because of the accessibility and mobility services they provide. Different household types might have different ownership levels (vehicles per household), e.g., larger households typically have more vehicles. Rural households typically have more vehicles, and have a higher percentage of trucks (versus cars). Higher income households have more vehicles per household member. Estimates of future ownership levels for different household types can be combined with population growth, demographic projections, and economic trends to obtain an overall forecast.

Note that insights into useful variables for forecasting vehicle ownership levels can be obtained from considering more detailed bottom-up models of the type described in section 3.2. For example, a household's choice of how many vehicles to own is known to be a function of: household size, number of adults, number of children, number of workers, household income, and possibly access to good public transportation.

One specific finding in the travel demand literature is that vehicle prices are not usually useful in aggregate-level forecasting. For example, the experience of one highly regarded expert has led him to conclude:

"It has proved generally difficult to introduce *price terms* into the models. Although on a priori grounds one would obviously expect them to influence demand for car ownership, it is difficult to find suitable datasets in which adequate variation in prices over time exists. It can certainly be said that there is no correlation between the unexplained growth over time and the movement of any general price indices relating to motoring. Thus it does not

appear that the temporal stability would be improved by the inclusion of price effects.” (Emphasis added.) [Bates (2013, page 25)]

However, what immediately follows is:

“The only way in which it has been possible to develop *price effects* on car ownership is by means of so-called ‘*car type models*’ (see Chapter 28)¹⁷, where the basic concept of ‘car’ is greatly expanded to consider engine size, age and other essential characteristics; an example is the work by Train (1986)...” [ibid]

Put another way: When using aggregate-level models to project car ownership levels (which are then translated into projections of vehicle stock) price effects are not helpful. However, when using more detailed (Berkovec-type) systems that include household discrete choice models based on vehicle attributes, price effects can then have an effect on the results.

This means that, even given NHTSA’s perception that they had limited options available, *their choice to adopt an approach using a “new vehicle sales model” that relies heavily on (average) new vehicle price is at odds with the perceived wisdom of the travel demand forecasting literature, and is essentially the opposite of GC96’s approach.*

To review: GC96 starts with forecasting the change in vehicle stock (not new vehicle sales). They use an approach that relies on Census data projections of various aggregate-level household statistics and trends on vehicle ownership, e.g., average number of vehicles per household, as well as breakdowns involving fractions of household holdings of trucks versus cars. These can be combined to produce estimates of future vehicle stocks of cars and trucks.

Finally, note that GC96 specifically discusses why this approach is better than, e.g., *time series methods* of the type used by NHTSA (GC96, page 137).

3.4 National Energy Modeling System (NEMS)

EIA’s NEMS is widely used in policy analysis involving energy-related issues, and is used by EIA to produce its Annual Energy Outlook. Appendix B provides a description of NEMS. The 2018 AEO provides projections of total fleet size and new vehicle sales for both the Existing standards and the proposed Rollback that are *directly comparable* to CAFE model projections, which is the subject of section 5.

The purpose of this section is to briefly review salient features of NEMS in relation to the material in sections 3.1-3.3. First, NEMS is a large-scale model of the economy that includes *price equilibration* of the type described in section 3.2. This is a feature

¹⁷ See Bunch and Chen (2008).

that very few modeling systems have. Second, NEMS generally follows the recommended approach discussed in the previous section: Getting a high-level estimate of overall “transportation needs” for the economy as a first step.

However, after that, NEMS represents yet another approach that is different from the ones discussed thus far. Although it does not have a detailed model of household new and used vehicle markets of the type discussed in section 3.2, it does have a *structural model of new vehicle sales shares* that uses a discrete choice model preferences for vehicle attributes, and a relatively large number of vehicle classes (including fuel types required analyzing scenarios with, e.g., electric vehicles, plug-in hybrid electric, etc.)¹⁸. Recall from section 3.3, Bates (2008) indicates that price can be introduced as an explanatory variable if ‘car type’ (choice) models are used.

Finally, NEMS (like the CAFE model) has its own manufacturer decision model with information on technologies and costs, and it makes vehicle redesign decisions in response to fuel efficiency/GHG standards.

3.5 Background on Aggregate Scrappage Models

The previous background sections establish that scrappage behavior is frequently a key component in determining the overall behavior of the vehicle market. In addition, our analysis reveals that NHTSA’s scrappage model has a much larger impact on CAFE model output and net-benefit analyses than the other two models (section 2). We therefore explore scrappage models in more detail.

As previously discussed, both Berkovec and GC96 directly employ scrappage models. Their models, as well many others in the literature, share many similarities based on a behavioral theory of decision making at the individual consumer level. These can be summarized as follows:

- Once purchased, a household uses a vehicle over some period of time for the purpose of consuming its “mobility services.”
- Vehicles are durable goods that age and deteriorate over time.
- As a vehicle ages/deteriorates, its market value decreases.
- Deterioration also gives rise to an increasing need for maintenance, which can be viewed as occurring in the form of discrete “events”.
- Both the frequency of these events, and the associated costs, can modeled using probability distributions, and which take into account that

¹⁸ It uses a Nested Logit model that can be viewed as taking the tree in Figure 3-1 and making a horizontal “cut” so as to leave the top layers of “vehicle classes”. operating cost, and other vehicle attributes (similar to the description of household choice models in section 3.2).

- probabilities can change over time (e.g., more frequent events, with higher costs).
- As this process continues, it will eventually trigger a decision by the consumer to either replace or dispose of the vehicle.

The details of the consumer's decision options will vary, depending on the current value of the vehicle and the preferences of the consumer. Early in a vehicle's life the consumer could trade in the vehicle as part of a replacement transaction, or simply sell it in the used vehicle market. Later in the vehicle's life when its market value gets lower, the need for a repair could trigger a different type of decision: spend money on a repair, or scrap the vehicle.¹⁹ In this case, the decision can be modeled as being governed by the following "parameters":

- C_n = the repair cost required to return the vehicle to good operating condition,
- δ_n = the scrap value of the vehicle,
- P_n = the market value of the vehicle (when in good operating condition)

The vehicle will be scrapped if:

$$P_n - \delta_n < C_n. \quad (3)$$

It will be repaired otherwise.

Specific models are formulated by including additional details and assumptions. For example, implementing a model requires some assumptions on the probability distributions for frequency of repair events and the level of cost.

One general factor affecting the behavior of a scrappage model is the inherent durability of the vehicle. Another factor might be the rate of driving: although these can clearly vary by consumer and specific vehicle, scrappage models typically assume that any given vehicle *type* (or possibly *class*) will follow a similar pattern of driving.²⁰

However, the vehicle's market value (price) that appears in the above scrappage rule ensures that the decision is determined at least in part by economic factors that may be independent of the vehicle's durability.

¹⁹ Note that a similar decision could arise earlier in the life of a vehicle in the event of an accident.

²⁰ Another factor occurring in the real world is the driving style of the driver, which can be correlated with vehicle choice. For example, a young male purchasing the "muscle car" version of a particular vehicle model might have a higher probability of accidents than another driver that has purchased a more "mainstream" version of an otherwise identical vehicle. These complications are frequently assumed to not play a role.

Greenspan and Cohen (1999) view these two factors (“engineering” and “cyclical”) as “separable,” and employ a two-step model based on Walker (1968). Another reference using this approach is Bento et al. (2018). Alternatively, both effects can be captured using single-equation (“one step”) models. Examples include Parks (1977) and Jacobsen and van Benthem (2015). (All of these references were cited in the PRIA.)

Because of their direct relevance to evaluating NHTSA’s scrapage model, we now review some of these references in more detail, and discuss their implications.

3.5.1 Bento et al. (2018) Scrapage Model

In what follows, we will refer to “Bento” for ease of presentation. Similarly, we will at times use the abbreviations “DS” or “DSM” in reference to NHTSA’s dynamic scrapage model. Bento provides a useful point of comparison to DSM because (1) it is based on recent data that overlaps with DSM, and (2) it models scrapage at an aggregate level (as in the DSM), using separate models for a small number of vehicle “types.”

Specifically, Bento estimates models for two types: Cars and Trucks. The DSM uses three types: Cars, SUV/Vans, and Trucks. However, the CAFE model only reports *results* at the level of Cars and Trucks, so all of our comparisons will be made on this basis. To begin, we have replicated Table 1 of Bento below (Table 3-1), which reports average scrapage rates (as a function of vehicle age) for three different groups of vehicles (grouped according to range of model years).

A scrapage rate can be defined as follows: Assume that a vehicle has survived to be a years old. The scrapage rate is the *probability* that this vehicle will be scrapped during the next year (when it would otherwise attain an age of $a+1$). Another statistic of potential interest is the *survival probability*, which is $1 - \text{ScrapRate}$.

New vehicles have age $a = 0$. There are a number of complications with computing scrapage rates in the first few years, so rates for the first year or two are frequently not reported.²¹

²¹ The complication arises because of the timing of how vehicles of a particular model year (MY) are introduced into the market. Using the most simple equation for scrap rate, this could give a negative scrap rate in year 1. Greenspan and Cohen discuss this problem in some detail, and provide a procedure for processing data based on registration data. NHTSA follows these procedures (see PRIA, pp. xxx).

Table 1: Percent Scrappage Rate by Age for Passenger Cars and Light Trucks

Model Year	Passenger Cars			Light Trucks		
	1969–1979	1980–1987 ^a	1987–2014 ^b	1969–1979	1980–1987 ^a	1987–2014 ^b
Vehicle Age						
2	1.51%	0.61%	0.87%	0.78%	0.99%	0.64%
3	1.84%	1.45%	0.92%	1.32%	0.65%	0.69%
4	2.03%	1.15%	1.30%	1.27%	1.10%	0.96%
5	2.56%	1.74%	1.62%	1.61%	1.27%	1.13%
6	3.79%	3.28%	2.11%	2.20%	1.99%	1.37%
7	5.30%	3.94%	2.56%	2.99%	2.23%	1.56%
8	7.17%	5.32%	3.38%	3.75%	2.73%	2.06%
9	9.40%	6.86%	4.24%	4.40%	3.71%	2.74%
10	11.75%	8.57%	5.59%	5.85%	4.99%	4.08%
11	13.84%	10.13%	6.97%	5.33%	4.79%	5.09%
12	15.65%	11.77%	8.84%	6.57%	6.86%	5.99%
13	17.18%	13.13%	10.64%	7.08%	6.54%	6.94%
14	18.29%	14.56%	12.78%	7.53%	9.16%	8.07%

Notes: The table presents average scrappage rates from passenger cars and light trucks in several time periods at ages 2 to 14 years old. Because inventories are not entirely sold until the second year the true base of cars is established at age 1 rather than age 0.

^a Uses data exclusively from Wards Automotive Yearbooks.

^b Uses data exclusively from Polk/HIS, after 1987.

Table 3-1. Scrappage Rates from Bento et al. (2018)

Table 5 illustrates a number of features. First, average rates for a given age are based on data across multiple model years (which are linked to calendar years). So, calendar-year-specific effects should be averaged out, removing cyclical/economic factors. What remains should represent the average of engineering/durability effects for the vehicle cohort. The average scrap rate for any particular age appears to decline as a vehicle cohort gets “newer” (with some reversals for older years of trucks). Second, Cars seem to be scrapped at a systematically higher rate than Trucks as a function of age. This could be correlated with faster driving rates.

One problem with scrappage data is that vehicle counts can become very small as age increases, so that estimated rates are increasingly volatile at higher ages. For this reason, *tables of rates are frequently truncated (e.g., at 14 years)*. This becomes a potentially important issue when estimating models.

We highlight this distinction between “engineering” and “cyclical” scrappage effects because it is related to the two-step modeling approach used by both GC96 and Bento. In the first stage a model is estimated that captures the average pattern of scrappage due to engineering durability effects. Bento estimates an engineering scrappage rate model for each of the three vehicle groups using the following 3-parameter logistic model:

$$R_{am} = \frac{1}{L + B * e^{-k*a}} + \varepsilon_{am} \quad (4)$$

where R_{am} is the scrappage rate for model year m at age a .²² This function can capture lower and upper bounds, and the overall shape of the curve.

Note that the only explanatory variable here is *vehicle age*. Bento maintains this level of simplicity, requiring only three parameters for each group of vehicles. For example, his model estimates for the 1987-2014 passenger car cohort are: $L = 2.724$, $B = 314.030$, and $k = 0.275$.

Other researchers have used slightly different logit-related functions with similar properties, often with more complex equations in the exponent. However, two-step models should always involve functions of age and/or model year in the first stage to capture the desired effect, i.e., average engineering scrap rates related to durability.

Once the stage-one model is estimated, the residuals (differences between observed values and averages) can be computed. As noted, these are effects that are “left over” after removing engineering/durability effect, so these deviations should be due to cyclical/economic factors. Bento estimates a variety of stage 2 models (using a similarly simple form) beginning with one based on the original theory of Walker (1986).

According to Walker’s theory, the economic factors affecting scrappage can be represented by two explanatory variables:

- Car ownership turnover rate
- Used Vehicle Price Index = Used vehicle price CPI/maintenance-repair cost CPI

where CPI denotes a consumer price index developed by the Bureau of Labor Statistics (BLS). Car ownership turnover rate is not generally available, so Walker suggests approximating it by the ratio of new car registrations to total car registrations (a measure of the overall fleet turnover rate).

The other variable (used vehicle price index, defined above) is worthy of discussion, because it has potentially important modeling implications. First, recall that the behavioral rule for deciding whether to scrap involves both *used vehicle price* and *maintenance cost* (see equation 3), so this is an *economics-based choice*.

²² Because of the assumption of this form and disturbance term, Bento estimates the model using nonlinear least-squares.

However, Walker's approach indicates the importance of using *quality-adjusted* prices (not just market transaction prices). Over time, new vehicles are constantly being improved through addition of new features and technologies (e.g., safety, fuel efficiency, etc.). Because of this, the average used vehicle price for, e.g., a given age group may not have the same meaning at two different points in time. Prices need to be quality adjusted so that they are comparable when estimating models using longitudinal data. This theoretical principle has generally been adopted in the literature.

Bento uses Walker's stage-2 model as the default model, which (similar to stage 1) requires only *three parameters*. One of these parameters is a measure of the *price elasticity of scrappage* with respect to *used vehicle price (index) change*, which in this modeling approach is an overall measure that captures systematic shifts relative to the (baseline) engineering scrappage rates. Their reported representative measure of scrappage elasticity with respect to used vehicle prices is -0.4, which they indicate is "substantially different than values adopted in simulation models."

Note that, if this type of model is available, it represents one specific "link" in the chain of logic of the Gruenspecht effect (section 3.1.1): "when used vehicle prices increase, scrappage rates go down." Modeling this effect directly as part of a larger modeling system (e.g., Berkovec) would be the preferred approach (if possible). However, Bento goes beyond this and performs a specific model-based "test" that provides evidence for the *existence* of the Gruenspecht effect. Specifically, there is evidence of a relationship between increasing CAFE standards and changes in used vehicle prices and scrappage in the data they were analyzing. This subject is also addressed by the next reference.

3.5.2 Jacobsen and van Benthem (2015)

Another relevant reference is Jacobsen and van Benthem (2015) [JvB15]. In part of their study they estimate the relationship between used vehicle price changes and and scrappage rates (similar to Bento). However, their models are estimated using a large database of used vehicle transactions prices for vehicles with highly detailed descriptions (make/model/configuration). (In other words, they are using highly disaggregated data rather than aggregated data.) Their approach exploits the fact that, at this level of detail, price adjustments of used vehicles in response to changes in gasoline prices, and, simultaneously, changes in scrappage rates, can be used to "identify" the effect(s) of interest (again, similar to Bento). However, in this type of approach the goal is to statistically identify specific "effects," not to create a formal scrappage model of the type found in the other references.

For example, they are able to determine that used vehicle prices *change* in response to gasoline price changes, and that the amount of change is different for vehicles with different levels of fuel economy. A similar effect is found for changes in scrappage rates. They emphasize that these effects can also cause vehicles in different vehicle groups to be driven differently over time: Specifically, high

efficiency vehicles will be driven at a higher rate. *The fact that vehicle miles traveled (VMT) can be different for vehicles with different fuel efficiencies is an affect that has not been taken into account in net-benefit analyses. Currently, the only effects that are taken into account are the age and type of vehicle.*

At the same time, JvB15's approach (which relies on a large number of fixed effects) does not incorporate macroeconomic indexes of the type that might be useful for forecasting. Given their data and modeling approach, it is valid for them to base their analysis on actual *observed* used prices without any quality adjustment for different model years. Their central estimate of scrappage elasticity with respect to used vehicle price change is -0.7. This is a bit larger in magnitude than Bento's (-0.4), but still much smaller than values that have been used in simulation studies in the past (e.g., -1 to -3).

JvB15 also addresses the Gruenspecht effect in a very direct way. They employ a full simulation system that formally includes sub-models to represent all of the features found in the Berkovec framework: a manufacturer decision model, a discrete choice model for consumer vehicle choices based on attribute preferences (price, fuel economy, etc.), a scrappage model, and a price equilibration procedure.

The manufacturer decision model is more realistic than the one used by NHTSA, incorporating both the decision to add fuel economy as well as the option for pricing strategies. However, it is much less detailed and more highly stylized (relying on cost curves). Manufacturers are assumed to maximize profits, which is included in the model. The consumer discrete choice model treats all households as one-vehicle households, but capture choices of vehicles by type (car or truck), size (small or large), age (0-18), and manufacturer (7 firms).

They perform simulations on this highly stylized system and demonstrate the existence of a Gruenspecht effect in response to increasing CAFE levels, and estimate fuel savings "leakage" due to higher numbers of used vehicles.

Both this reference and Bento, et al. (2018) were published only recently. They are generally regarded as providing some of the only limited, reliable evidence of the *existence* of the Gruenspecht effect. In addition JvB15 provides evidence that this effect could be important to consider when making policy decisions. However, neither paper provides results that are sufficiently detailed for direct inclusion into a policy analysis. Moreover, NHTSA's scrappage model bears little resemblance to the approaches used by these researchers in terms of theoretical support and rigor.

3.5.3 Greenspan-Cohen (1999) Scrappage Model

As described in section 3.3, GC96 uses a two-stage modeling approach for scrappage (similar to Bento). A stage-one engineering scrappage curve is estimated as a function of vehicle age and model year (albeit more complex than Bento's). In their second-stage model, their explanatory variables are: civilian unemployment rate,

gasoline prices, and a *new vehicle price index* using the new-price version of the Walker (1968) definition for used vehicle price index. Specifically, they use a ratio of the BLS cost of repair index and the BLS new vehicle price index. Moreover, they cite Parks (1977) who finds that the inverse of this ratio is “highly significant in explaining total scrappage.”

The main point in mentioning these references is that they do lend *some* support to the notion of using new-vehicle-price-related explanatory variables in a scrappage model. However, these still maintain the theoretical requirement of using quality-adjusted prices rather than market transaction prices.

Chapter 8 of the PRIA reveals that NHTSA is aware of this theoretical consideration, but takes pains to argue in favor of using *unadjusted* new vehicle prices instead. We postpone any further discussion until later.

3.6 A Theory-Based Evaluation of NHTSA’s Economic Models

With sections 3.1-3.5 as background, we now evaluate NHTSA’s economic modeling approach based on theoretical considerations. First, we summarize some key points that were established in the previous sections.

- Reading the relevant material on their models in the NPRM and PRIA make it clear that their primary motivation was to produce a modeling approach that would somehow produce the Gruenspecht effect. However, rather than develop models that are based on solid behavioral theory to capture the structural causes of what could be a legitimate effect, they produced a pair of models (new vehicle sales, and dynamic scrappage) that were “stitched together” by virtue of sharing a common explanatory variable: average compliance cost.
- It is important to understand that the two models in equations (1)-(2) eschew what is known about how this type of aggregate-level modeling should be done from reviewing the literature, which is further evidence of NHTSA’s motivations.

The material in sections 3.2-3.5 was provided to demonstrate this. Section 3.2 shows what an approach with behavioral content, integration between new and used vehicle markets, etc., looks like. However, anyone reading Chapter 8 of the PRIA (“Economic Analysis of Regulatory Alternatives”) could certainly be forgiven if they were under the impression that the CAFE model included models with behavioral content.

Figure 8-1 from the PRIA has been reproduced in Figure 3-4 as an illustration.

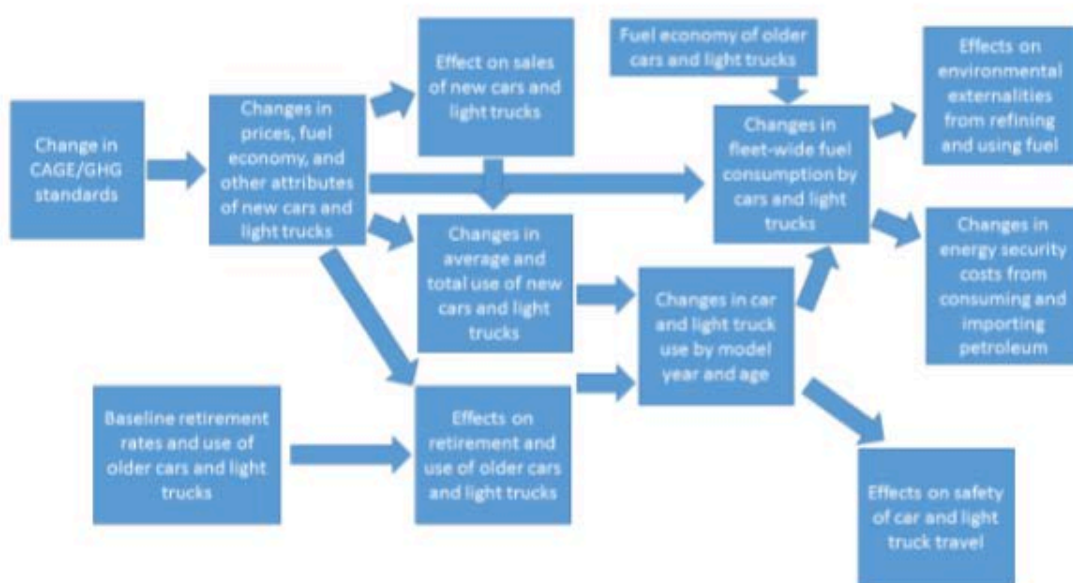


Figure 8-1 - Overview of Economic Effects of Changing Fuel Economy and CO₂ Standards

Figure 3-4. Figure 8-1 from the PRIA

From the PRIA (page 938):

“As the figure indicates, the resulting changes in the *fuel economy, other features, and prices of new vehicles will affect their sales*, although the direction in which they do so is difficult to anticipate. This is because the change depends on how potential buyers value the future savings or increase in fuel costs that result from changing vehicles’ fuel economy, as well as how they value any accompanying changes in other attributes that affect their utility. Modifying vehicles’ fuel economy also changes their operating costs (by changing the amount of fuel consumed in driving each mile), which as the figure also shows, affects how much they are likely to be driven each year and throughout their lifetimes.” (Emphasis added.)

This seems to imply that NHTSA is going to be providing a behavioral model of new vehicle choice (similar to the NEMS description in section 3.4), as well as a VMT model that is a function of fuel operating costs. Continuing on:

“At the same time, changes in the prices, fuel economy, and other features of new cars and light trucks will alter some potential buyers’ choices between new and used models because used vehicles often represent a close substitute for new models. The direction of this effect again depends on the magnitude of changes in new vehicles’ prices and on how buyers value the changes in new vehicles’ fuel economy relative to any accompanying changes in their other features. If on balance fewer buyers elect to purchase new cars

or light trucks, some who would otherwise have purchased a new model may decide to buy a used model instead, while others will continue to drive a vehicle they already own. Conversely, if buyers find the combination of changes in new vehicles' prices, fuel economy, and other attributes attractive, some will respond by purchasing new vehicles instead of buying used models or by replacing one on they already own."

This effect is shown in Figure 8-1 as a change in the demand for used vehicles. ..."

The paragraph above seems to imply that the behavioral interactions between the new and used markets are also being captured, and that, again, these are being modeled on the basis of consumers preferences for "prices, fuel economy, and other features of new cars and light trucks" versus preferences for used versions of these vehicles."

This type of narrative continues to flow through the remainder of the initial sections of Chapter 8, with occasional factually correct statements about what has actually been done (and why) interwoven into the narrative.

3.6.1 New Vehicle Sales

Section 8.6.2 of the PRIA includes a discussion of some of the actual details regarding modeling decisions and approaches, which in this case is focused on "changes in new vehicle sales." The first part describes the challenges they encountered when attempting to estimate a new vehicle sales model based on only aggregate-level data.

"The analysis explored various approaches to predict the response of new vehicle sales to the changes in prices, fuel economy, and other features in an attempt to overcome analytic challenges. This included treating new vehicle demand as the result of changes in total demand for vehicle ownership and demand necessary to replace used vehicles that are retired, analyzing total expenditures to purchase new cars and light trucks in conjunction with the total number sold, and other approaches. However, none of these methods offered a significant improvement over estimating the total number of vehicles sold directly from its historical relationship to directly measurable factors such as their average sales price, macroeconomic variables such as GDP or Personal Disposable Income, and regularly published surveys of consumer sentiment or confidence." (PRIA, page 956)

Parsing this paragraph is revealing. Recall that section 3.3 establishes that a better approach to aggregate-level modeling would be to model "total demand for vehicle ownership and demand necessary to replace used vehicles that are retired," which the above paragraph claims they tried. Recall also that the suggested approach

would *not* involve new vehicle prices. However, it is clear from the first sentence above that including new vehicle “prices” was a requirement for NHTSA.

The above paragraph also says that “none of these models offered a *significant* improvement” over the adopted approach (emphasis added). This implies that one of the other methods *was* an “improvement,” just not a “significant” (enough) improvement to alter their decision. Based on our review (which indicates that prices are not helpful in this type of analysis) it seems possible that, e.g., the superiority of using an approach like GC96’s was discovered, but that it may have been unacceptable because new vehicle prices did not play a role.

Also on page 956 of the PRIA is a description of the time-series econometric approach to estimation, which uses quarterly sales in order to “more accurately capture the causal effects of individual explanatory variables,” which (except for change in average compliance costs) consists entirely of lagged dependent variables and macroeconomic indicators.

Based on this material, our judgment is that the time-series modeling approach applied to quarterly data would be highly vulnerable to over-fitting, in addition to being inappropriate for policy analysis.

On page 957 of the PRIA they indicate that “The results of this approach are encouraging...” *We find that this is an inadequate standard for deciding to continue with this approach.*

Concerns about over-fitting were then confirmed based on pages 959-959 of the PRIA.

Recall that a model of new vehicle sales would be expected, in some sense, to reflect the market’s *preferences for new vehicle attributes other than price, and, in particular, fuel economy*. Page 957 of the PRIA indicates: “The model did not incorporate any measure of new car and light truck fuel economy that added to its ability to explain historical variation in sales...”

Similarly, a portion of the NPRM is revealing on this same point, and also that NHTSA staff understood that this high level of aggregation was yielding a poor modeling approach (yet they proceeded with it anyway).

*“Despite the evidence in the literature, summarized above, that consumers value most, if not all, of the fuel economy improvements when purchasing new vehicles, **the model described here operates at too high a level of aggregation to capture these preferences.** By modeling the total number of new vehicles sold in a given year, it is necessary to quantify important measures, like sales price or fuel economy, by averages. **Our model operates at a high level of aggregation, where the average fuel economy represents an average across many vehicle types, usage profiles, and fuel***

economy levels. In this context, the average fuel economy was not a meaningful value with respect to its influence on the total number of new vehicles sold. A number of recent studies have indeed shown that consumers value fuel savings (almost) fully. Those studies are frequently based on large datasets that are able to control for all other vehicle attributes through a variety of econometric techniques. They represent micro-level decisions, where a buyer is (at least theoretically) choosing between a more or less efficient version of a pickup truck (for example) that is otherwise identical. In an aggregate sense, the average is not comparable to the decision an individual consumer faces.

Estimating the sales response at the level of total new vehicle sales likely fails to address valid concerns about changes to the quality or attributes of new vehicles sold—both over time and in response to price increases resulting from CAFE standards. However, attempts to address such concerns would require significant additional data, new statistical approaches, and structural changes to the CAFE model over several years. [Emphasis added.]

We have highlighted in bold and italic items that are immediately obvious from the material provided in sections 3.1-3.5. These and other statements throughout the NPRM and PRIA indicate that NHTSA staff was painfully aware of the inadequacies of their modeling approach. While the above statements were oriented toward modeling new vehicle demand, similar statements are available in the area of scrappage.

3.6.2 Dynamic Scrappage

Section 3.5 reviewed scrappage models from the literature, and already highlighted a number of items that NHTSA staff chose to ignore. This section provides a more detailed and formal comparison.

The Bento model in section 3.5.1 provides a good option for comparison. Bento and NHTSA analyze similar types of aggregated vehicle count data from multiple model years. Bento analyzes relatively recent data that overlaps with NHTSA. In addition, NHTSA claims to have followed a two-step modeling approach and makes reference to the “engineering” and “cyclical” scrappage concepts applied in Bento (as well as GC96).

Recall that Bento’s two-step approach yields two models with three parameters each, and that the meaning of both models and their parameters are highly interpretable. We reported the stage-1 parameter estimates for passenger cars, as well as the estimated scrap elasticity.

By way of contrast, see Table 3-2 for an example of NHTSA’s first stage “engineering scrappage” model estimates for cars. For an excerpt of some “cyclical scrappage”

model estimates, see Table 3-3. For the final scrappage models used by the CAFE Model, see Table 3-4.

These tables highlight some of potential concerns about NHTSA's approach in comparison to the literature. Although the PRIA includes some discussion on their choice of variables, the actual model estimates defy any realistic possibility of meaningful behavioral interpretation. This is due to the extreme time-series-style approach of including large numbers of lagged effects (including for the dependent variable!). There is a clear risk that over-fitting is occurring.

This highlights further the problems with NHTSA's approach. The approach does not conform to the literature, and relies on reduced form time-series models with no direct interpretation. These factors are the reason for the numerical study in section 4.

In addition to the clearly obvious deficiencies on technical grounds, similar to the previous discussion on the new sales model, NHTSA staff clearly understood that they were miss-specifying their models in ways that run counter to the literature. For example, as discussed in section 3.1, it is clear that they wanted to have new vehicle costs as an attribute in the dynamic scrappage models to support a Gruenspecht effect. Although used prices are more appropriate, they needed new "prices."

In this regard, section 3.5.3 indicates that GC96 and Parks (1977) provide precedents for incorporating new vehicle prices, but only in a form consistent with theory. However, NHTSA rejected this requirement. On page 1018 of the PRIA, they actually make an argument that the reason for their approach was that "it is the purpose of *this study* to measure whether or not this is true" (emphasis added), i.e., whether or not a "quality adjustment" was actually required.

We would submit that this was a regulatory analysis and not a "study." If they believe this was a "study," they need to justify the use of their results for regulatory analysis. The subject of whether the above item in question was "true" or not was never revisited in the PRIA, nor could it have been given the modeling approach they decided to use.

Table 8-10 - Alternative Car Engineering Scrappage Specifications
Variable

cal significance of 5%

1024

Table 3-2. "Engineering Scrappage" Models for Cars in PRIA

Table 8-14 - Alternative Car Cyclical Scrappage Specifications

Variable

0653671 0.01012622

1037

Table 3-3. Excerpt of “Cyclical Scrappage Models’ For Cars from PRIA.

Parameters for the Scrapage Model

Parameter	Cars	Vans/SUVs	Pickups
Estimate Scrapage	TRUE	TRUE	TRUE
Beta Coefs			
Age	0.616047051	-0.473441117	-1.119398279
Age^2	-0.057406753	0.032324147	0.037890057
Age^3	0.001582126	-0.000301894	0
ln(MY-1959)	-1.608885894	-3.946616362	-3.364968508
ln(MY-1959)*Age	0.213582275	0.504803381	0.34204715
ln(MY-1959)*Age^2	-0.006715995	-0.015159639	-0.008384946
ln(MY-1959)*Age^3	0	0	0
New Price	-0.000161276	0.000371589	-0.000303124
New Price*Age	7.84025E-06	-2.88675E-05	2.83304E-05
New Price*Age^2	1.00488E-07	5.91183E-07	-9.62014E-07
New Price*Age^3	-1.212E-08	0	0
Lag New Price	9.47558E-05	-0.000185344	0.000460661
Lag New Price*Age	-2.34041E-06	1.08866E-06	-2.76789E-05
Lag New Price*Age^2	-7.40388E-07	6.36025E-09	6.4343E-07
Lag New Price*Age^3	2.60286E-08	0	0
Lag2 New Price	-0.000335592	-3.4816E-05	-0.000514968
Lag2 New Price*Age	5.20348E-05	1.745E-05	4.61463E-05
Lag2 New Price*Age^2	-2.21832E-06	-6.69202E-07	-1.27972E-06
Lag2 New Price*Age^3	1.84799E-08	0	0
Lag3 New Price	0.000493728	0.00013793	0.000430244
Lag3 New Price*Age	-9.12445E-05	-2.50298E-05	-4.29461E-05
Lag3 New Price*Age^2	5.12464E-06	8.77884E-07	1.18069E-06
Lag3 New Price*Age^3	-8.16078E-08	0	0
CPM	-0.085883784	-0.138151017	0.015197004
Lag CPM	0.085462805	0.044603678	-0.038813225
New CPM	0.080297688	0.273072429	-0.05654879
Lag New CPM	-0.1011716	-0.135346955	0.046611305
GDP Growth Rate	0.047918699	0.06054021	0.029019379
Lag GDP Growth Rate	0.024761655	0.072544513	-0.012424423
Lag2 GDP Growth Rate	0.011013988	0	0
Lag Scrapage	0.202961972	0.189616269	0.146198548
Lag2 Scrapage	0.150871596	0.005029224	0.056999227
Lag3 Scrapage	0.03576158	-0.066280435	0.112202637
Intercept	-0.691368454	-0.233821851	6.459967525
Decay Age	21	21	21
Final Survival Rate	0.01	0.025	0.025

Table 3-4. Parameters Used in CAFE Dynamic Scrapage Models

Moreover, NHTSA staff also clearly recognizes that the approach they are using is hampered by inadequate data. On page 1017 of the PRIA:

“While ideal data would represent individual vehicles, unfortunately the data is only available in aggregate for historical model years. The models are thus unable to be trained on model-specific data and must rely on average measures. This decision is further justified by the fact that the CAFE model does not capture any cross subsidization of technology costs that occurs between vehicles in an OEM’s fleet. Because it is likely manufacturers will cross-subsidize costs, the aggregate measure of average new vehicle price may be the best measure of the general price trend of the new vehicle market under different fuel economy standards, even if disaggregated data were available.”

First, the initial sentence about unavailable data is simply *not true*. Second, the remainder adds insult to injury: It justifies this approach by highlighting another deficiency in the CAFE model that we have already mentioned multiple times: the inability of the manufacturer decision model to take into account cross-subsidization. There are many other examples of this type that can be found in the NPRM and PRIA that call into question the decisions made by NHTSA.

3.6.3 Final Remarks on Economic Modeling

In this section we have shown based on theoretical considerations that NHTSA's approach is not only inadequate but clearly unacceptable for the purposes of a regulatory analysis this important. Based on our assessment, NHTSA's models would be highly unlikely to be accepted for publication if they were submitted to an academic journal. However, in order to conclusively demonstrate that our assessment is accurate, we provide numerical studies in the next two sections to demonstrate this.

What would a better approach look like? Section 3.2 reviews a modeling framework that supports development of modeling approaches that capture the necessary market structure and theory-based behavioral choices required for analyzing the effect of policy changes on vehicle markets. Part of that discussion highlighted the specific role that discrete choice models could play, and mentions two such models that have been used in California. In the 2016 TAR, EPA describes their recent investigation of consumer choice models, and although they recognize the challenges, in our view their findings are promising.

However, the NPRM (page 43076) includes a section entitled "Vehicle Choice Models as an Alternative Method To Estimate New Vehicle Sales," seemingly in anticipation that the flaws in their current modeling approach would be identified and exposed, and that choice models would be recommended as an alternative. The discussion provided in the NPRM is an exercise in suggesting poor ideas for how choice models might be used as "straw men" for the purpose of criticizing them. There is definitely a middle ground where structural models like discrete choice models could be used to support a proper analysis, rather than the two extremes represented in this NPRM: poorly executed aggregate-level reduced form models that do not conform to economic theory versus extraordinarily detailed structural models that would be difficult to implement and overly sensitive to small changes in input values.

4. Evaluation of NHTSA's Dynamic Scrappage Model

Because it is so difficult to evaluate the models by direct inspection of their coefficient estimates, we must rely on the creation of plots to reveal how they behave. We begin with some initial plots to provide context. Recall that Bento provides engineering scrappage models for MY1987-2015. NHTSA uses a similar,

overlapping data set from the same source (HIS/Polk) spanning a wider range of model years (MY 1975-2015).

Before looking at the current models, we begin by reviewing some of the available information included in the CAFE Model Parameters input file. Worksheets include “Vehicle Age Data” and “Historical Fleet Data.” Our understanding is that the Vehicle Age Data are, in effect, a set of static scrappage curves that were developed for use in the previous rulemaking. Moreover, these are the static curves that are used if the DSM is turned “off” during a CAFE model run.

For an initial comparison, see Figure 4-1. Bento-Car is the engineering scrappage curve for the MY1987-2014 cohort. As such, it should represent the effect of average durability for the cohort. The other two curves are based on data collection and analysis by NHTSA. The curves overlap completely, but are not smooth. They bear some similarity to the Bento-Car curve. After 7 years the CAFE curve lies to the right of Bento-Car, indicating higher durability. This might be expected if the curves were generated to avoid downward bias from historical data. However, before six years, the CAFÉ curve implies higher scrappage rate than Bento-Car.

We intentionally limited the initial plot to 14 years, because Bento reports that the data source “reports vehicle counts by age up to 15.” However, CAFE-Static reports values for 30 years, and CAFE-MY2014 reports values for 37 years. (But, as noted, these CAFE figures are identical for the first 30 years). Figure 3-2 extends the age range to 37 years. On this scale all three curves track fairly closely for about 21 years. Interestingly, the CAFE model documentation describes that, although the DSM uses a logit form similar to Bento, it has a trigger that causes it to switch to an alternative function (called a decay function), and the trigger parameter is set to 21 years. The curve is created to follow a decay rate ensuring that the final fleet survival percentage (another parameter) occurs at age 40.

To proceed logically, we would next like to understand how the DSM would behave it were not “dynamic.” That is, if the DSM is turned off, the CAFE model reverts to using a survival curve created for the previous rulemaking. However, what type of “static curve” would be produced by NHTSA’s *current* scrappage modeling approach if the dynamic feature were “turned off”? The dynamic effects in the current model are a function of average compliance costs (and perhaps CPM). In any case, we discovered that the PRIA includes a “sensitivity case” that seems to address this question. It provides a set of parameters that will “turn off” the part of the DSM that has been added for the purpose of creating a Gruenspecht effect. (See PRIA, page 1059).

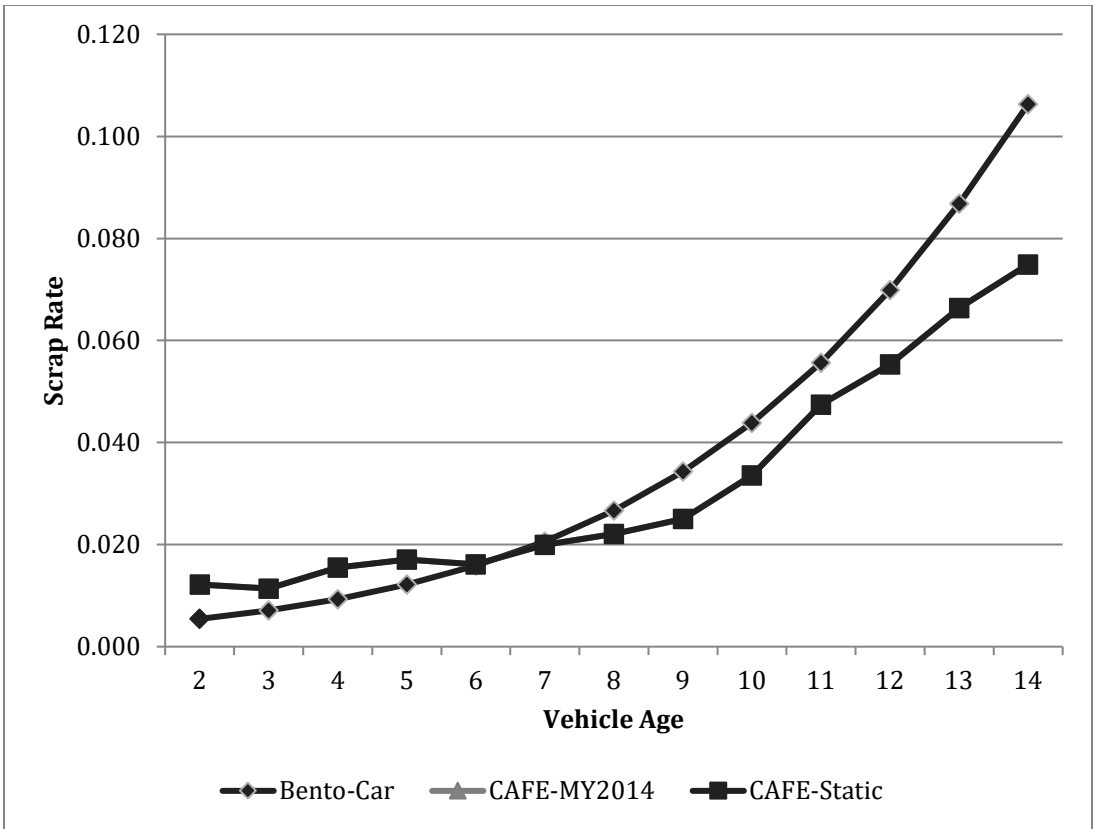


Figure 4-1. Car Scrap Rates: Bento (MY1987-2914 Car) vs. CAFE-Static/Data

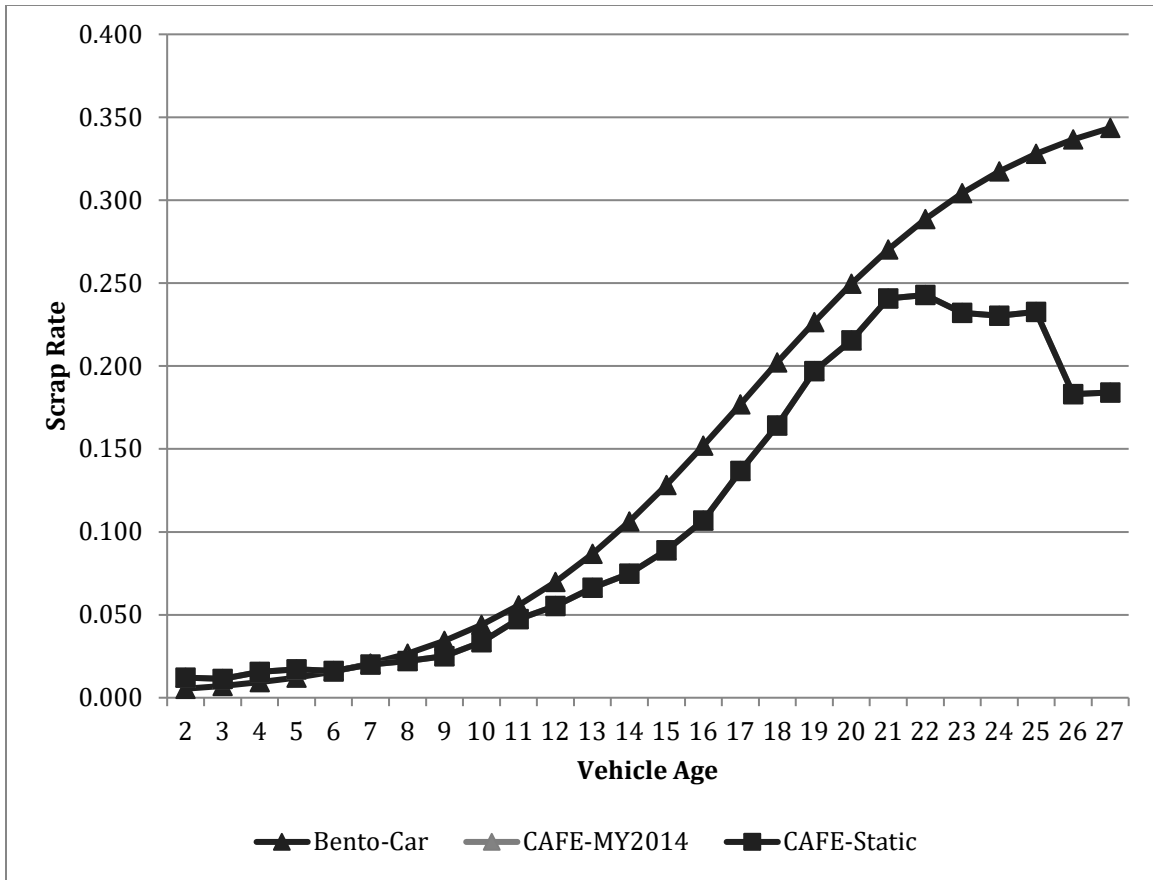


Figure 4-2. Car Scrap Rates from Figure 1 - Extended to 27 Years

We used these parameters to generate another set of “static” results. The “NoGruen” scrapage curve for MY2014 is provided for comparison in Figure 4-3. Note that, for completeness, we have included two NoGruen curves: one for Existing, and one for Rollback. As per NHTSA’s intent, these two curves are essentially the same (although not identical).

These curves raise some serious concerns about the behavior of the DSM. The Bento curve comes from MY1987-2014 data using (up to) 14-15 years for each model year. One argument for not using this curve for regulatory analysis is that the older vehicles in this group might be assumed to have less durability than newer vehicles, so the curve could have scrapage rates that are too high. Indeed, the Bento-Car curve lies just to the left of the CAFE-Static/MY2014 curves. One expects that these curves were constructed to ensure a higher level of durability than one based on averages of older vehicles. However, how does one actually obtain projections of durability levels for recently manufactured vehicles, much less future vehicles, without data?

The PRIA seems to suggest that one rationale for estimating such a complex scrapage model was that it could somehow provide such “projections” of future

durability levels. In this regard, both the CAFE curves and the DSM curves lie to the right of Bento-Car (after about year 8).

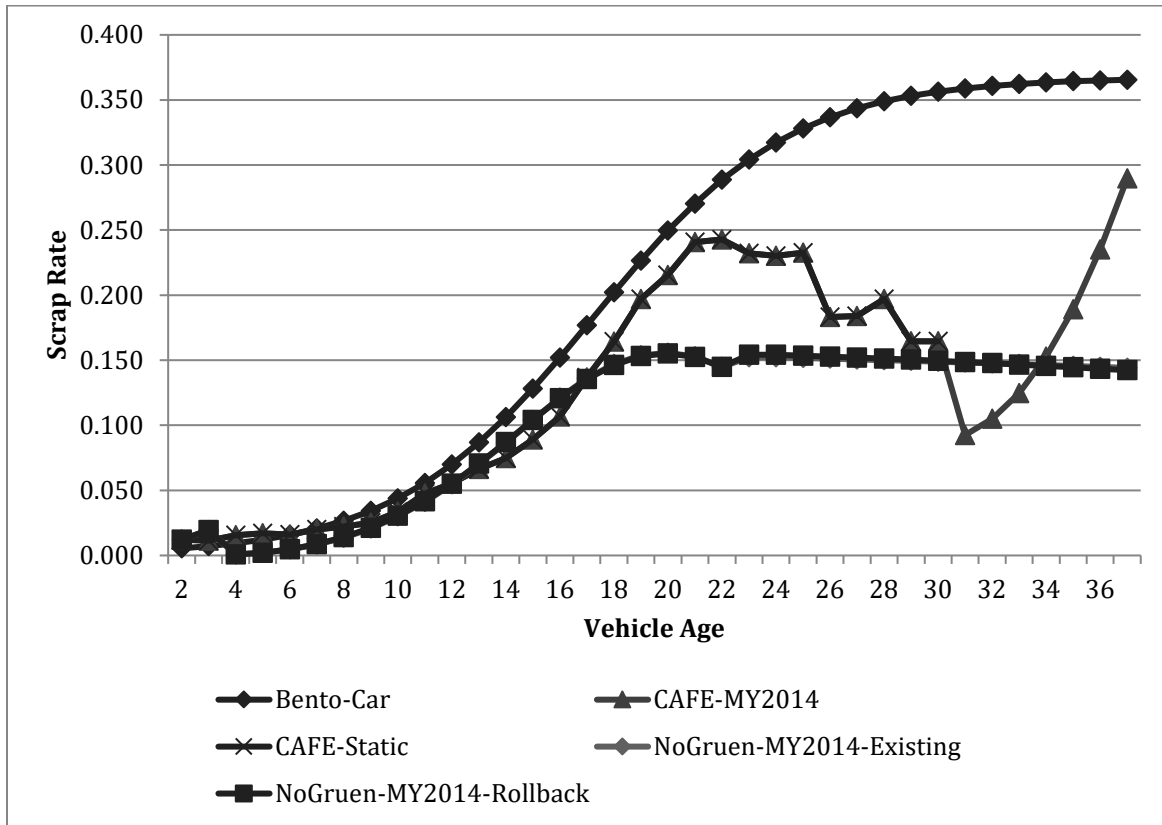


Figure 4-3. Car Scrap Rates: Figure 4-2 plus DSM Non-Gruenspect MY2014

However, there is no good explanation for why the DSM scrapage rate for a MY2014 Car would level off at 0.15 at 18 years. Even though Bento-Car is estimated on 14 years of data, it is not unreasonable for the curve to keep increasing for a number of years, and, indeed, the Bento-Car and CAFE-Static/Data curves track each other fairly closely until the “magic” 21-year mark.

The PRIA includes a very similar discussion for MY2016 (pp. 1055-1056). However, this is accompanied by *survival curves* rather than the scrap rates we have shown here. The survival curves do not seem to demonstrate the size of the effects we have detected here in the same way. Still, even in that discussion the results seem a bit odd, and the explanation offered is that an “optimistic” GDP growth rate is the cause of apparent anomalies. For future reference Figure 8-35 in the PRIA shows that in the 20-22 year age range only about 20% of the original MY fleet would be remaining. (This will be important later.)

To establish some additional points of reference, we generated scrap rate curves for two more model years: 2021 and 2028. For an initial comparison, see Figure 4-4. (To reduce clutter, we include the Existing scenario NoGruen results only, which are essentially the same as the Rollback results). On this scale there are differences in

scrap rates as a function of model year. One might expect these to represent increasing durability. However, see Figure 4-5 that includes only the DSM results.

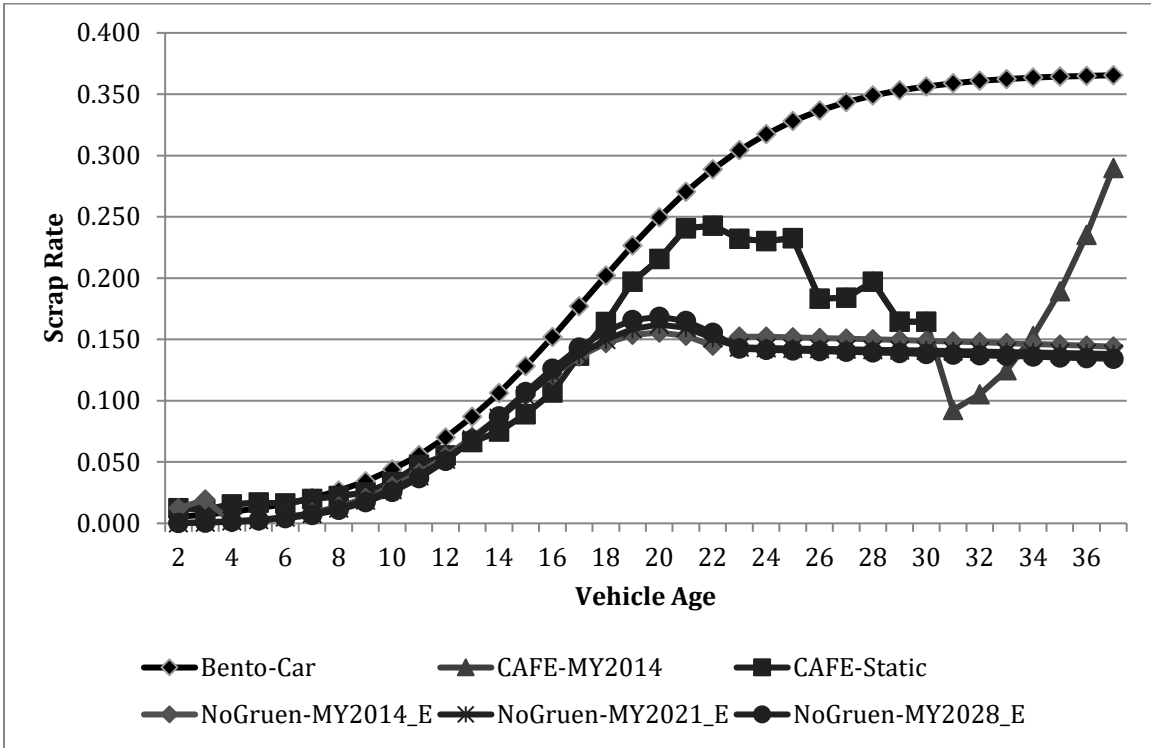


Figure 4-4. Car Scrap Rates: Additional MYs with No Gruenspecht Effect.

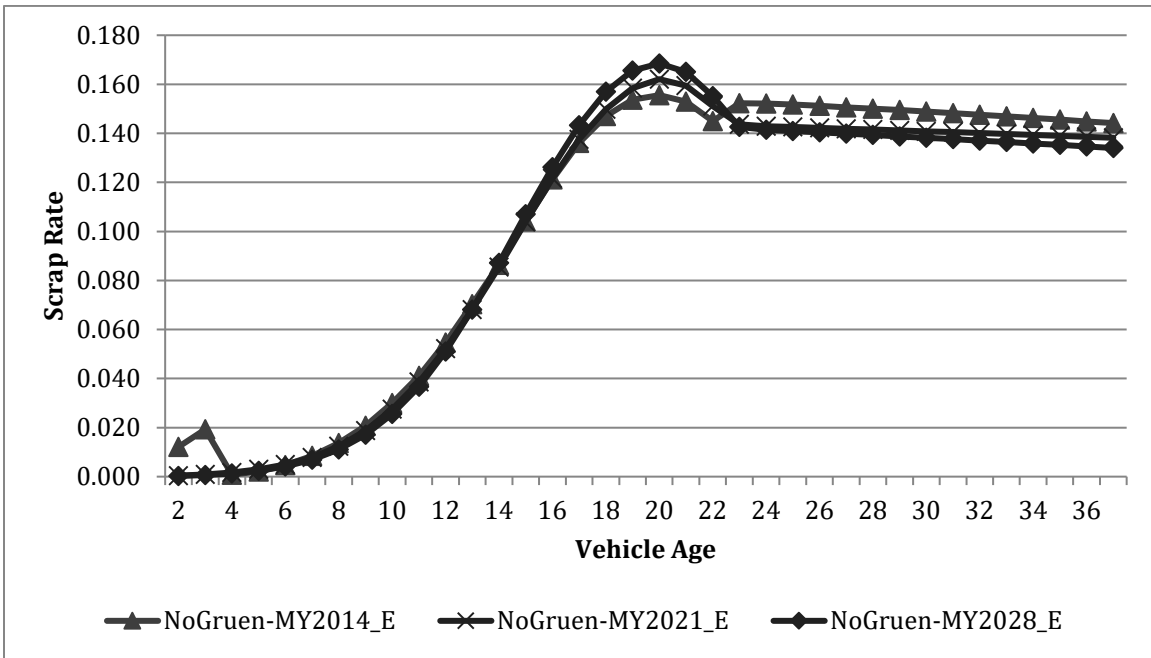


Figure 4-5. Car Scrap Rates: Three MYs with No Gruenspecht (Isolated)

A closer look at Figure 4-5 shows that there is a pattern of differences from increasing model year starting after age 14 or 15. For some reason, scrap rates for cars *increase* with increasing model year from age 16 until reaching the ‘trigger’ at age 21, and then the ordering of scrap rates reverses. Again, these differences are rather small, but they do demonstrate that most of the “action” with this model only seems to occur rather late (e.g., 16 years or later), but, then the trigger is reached at 21 years and an ad hoc model takes over. When action does occur, it seems counterintuitive. Finally, the general leveling off in the 0.14 to 0.17 range for all vehicles represents a type of “regularity” in future durability of vehicles that is questionable in origin.

With the above results for context, it is now time to “reactivate” the Gruenspecht effect built into the DSM. For results corresponding to those from Figure 4-3, see Figure 4-6. When the full DSM is operating, there is a small change in the scrap rate under the Rollback scenario, but a relatively large change under the Existing standards. And, consistent with previous patterns, the changes are concentrated in 16-21 age range. For a collection of the DMS results for the three model years, considered previously, see Figure 4-7.

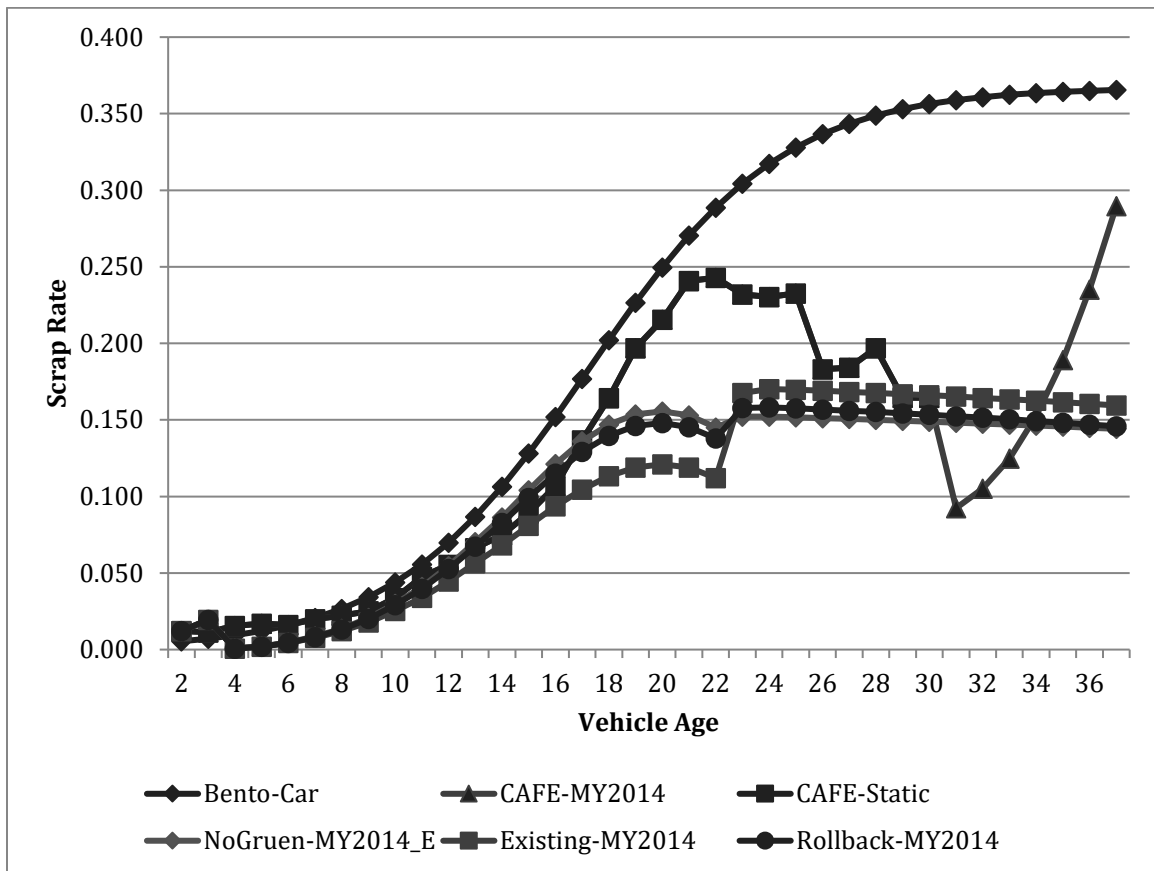


Figure 4-6. Car Scrap Rates: Dynamic Scrappage for MY2014.

The Figure 4-7 results are expected as a logical extrapolation from Figure 4-6, although on this scale the patterns are a bit clearer. The separation in scrap rates between the Existing and Rollback scenarios start to become visible by age 12 or 13, although it is still true that most of the differences occur between ages 16 and 21. However, by now it should be clear that the reversals that occur starting at age 21 are inconsistent with any theory that is based on the concept of durability versus cyclical effects.

How would one evaluate whether the large differences between Existing and Rollback scrappage rates are consistent with theory? The PRIA (page 1003) offers a theory-based discussion that was intended to support the development of these models:

“The effects of this process on prices and the number of vehicles in use are likely to vary significantly among those of different ages and accumulated mileage (a measure of their cumulative lifetime use). Figure 8-17 through - Figure 8-18 illustrate the likely differences. As Figure 8-17 and Figure 8-17 show, the supply of both nearly-new vehicles (say, those less than five years old) and very old vehicles (more than 15 years) is likely to be very unresponsive to changes in their price.” (Emphasis added.)

And, from page 1005:

“Shifts in demand for used cars and light trucks of different ages in response to changes in the prices and attributes of new models are likely to mirror how closely they substitute for their new counterparts. Nearly-new vehicles offer the closest substitutes for new ones, so their demand is likely to be most responsive to changes in prices and other characteristics of new ones, while the outdated features and accumulated usage of older vehicles make them less satisfactory substitutes.”

So, the observed behavior of the dynamic scrappage models appear to be the exact opposite of what is predicted in the PRIA: Extremely old vehicles appear to be the most sensitive to new vehicle “price” changes.

In looking for an actual computational test, the notion of elasticity will be useful. The differences in these curves are being driven by the difference in average compliance costs for new vehicles being introduced into the market. These, of course, don’t actually represent meaningful price changes, for multiple reasons. First, consumers actually select from among individual vehicles (not one “average vehicle”), and second, the prices for these vehicles are determined by market equilibrium (which does not exist in the CAFE model). Finally, Cars actually do compete with vehicles in the other two classes.

However, suppose that the new vehicle market average price increase actually had an effect on scrappage of new vehicles due to a Gruenspecht effect. What is the likely size of the “elasticity”?

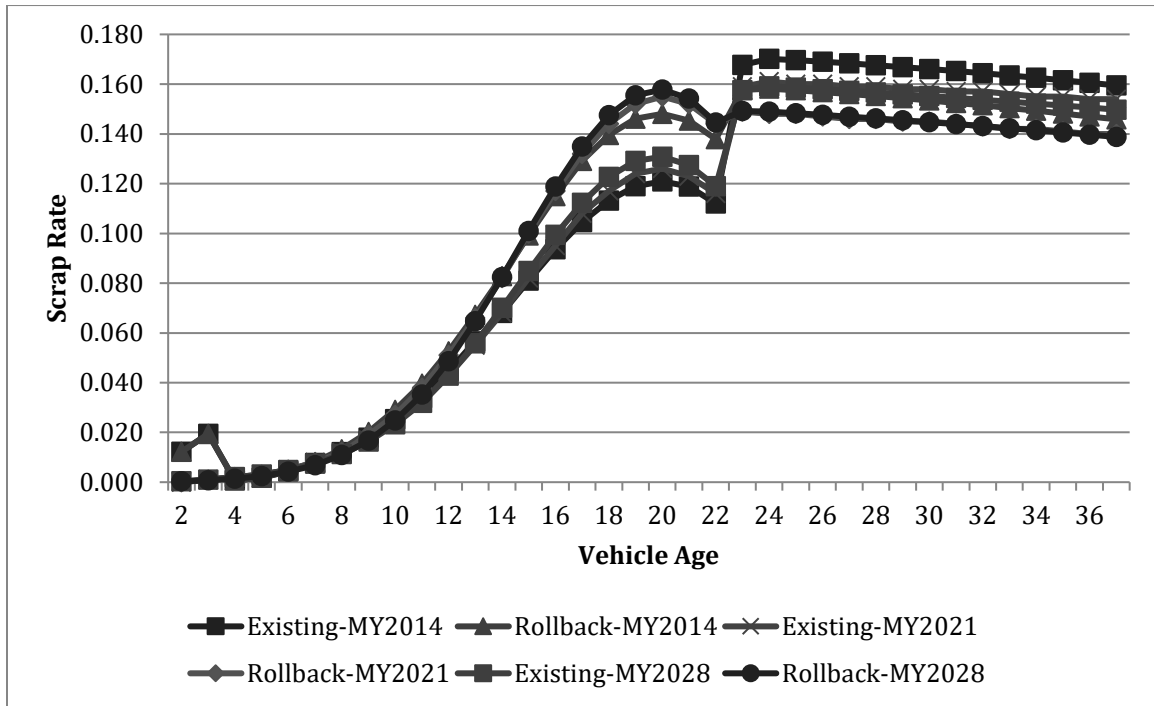


Figure 4-7. Car Scrap Rates: Dynamic Scrapage for Three MYs.

The complexity of the models estimated by NHTSA makes it difficult to develop an interpretation in terms of elasticity by direct inspection of its coefficients. However, we do have the Bento model to use as a reference point. (Recall that Bento uses an equation with a parameter that can be directly interpreted as an elasticity.)

See Figure 4-8. Using the elasticity from Bento, we have created a plot for the effect on scrappage rates *if used prices were to be increased by 20%*. This is a very large percentage increase, and it has been directly applied to all used cars. As might be expected, the impact on scrappage rate increases to its maximum effect for the oldest cars left in the fleet. However, for vehicles of age 16 through 20, the impact of this effect is much *smaller* than the effect generated by the DSM on the basis of *new car price increases*, even though used cars are much closer substitutes for one another than are new cars.

This is a clear indication that the DSM is likely to be too sensitive to average compliance cost increases for new cars.

However, there is yet one more aspect of the DSM models to consider: The difficulty associated with estimating a scrappage model that will produce accurate projections for what goes on in the “tails” (as vehicles are getting older). The model estimation results in the PRIA report root mean square errors (RMSE) for models that are

estimated in the “logit space.” This RMSE gives a measurement of the error for predicted values of the dependent variable (the logit) when using the model. The value reported for the Car model is about 0.15 (PRIA, Table 8-17, page 1044).

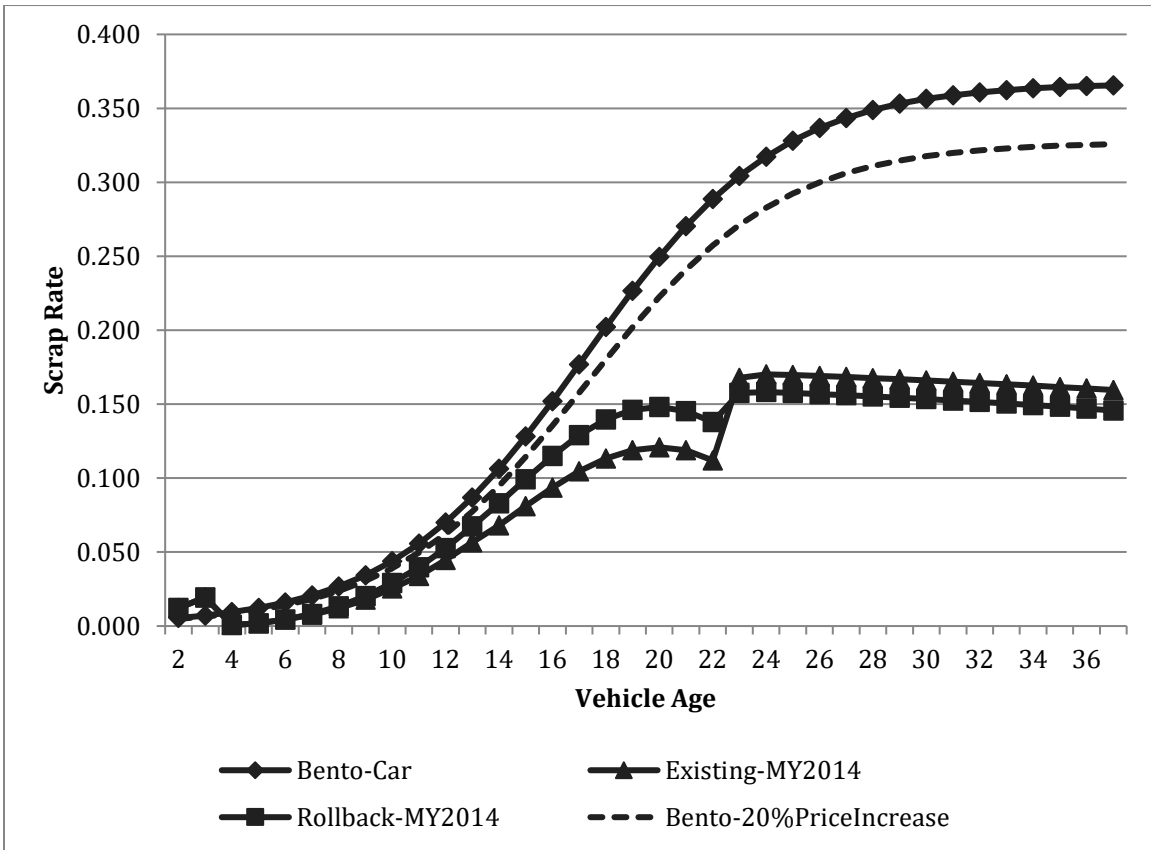


Figure 4-8. Car Scrap Rates: Bento Model Response to 20% Used Vehicle Price Increase

Using this value, we performed additional CAFE model runs that could be used to produce a 95% confidence interval (technically, a “prediction interval”) around the predicted scrappage rates. (The technical details are in Appendix C.) Observed historical data on scrappage rates are used to estimate the model’s parameters, and random variation in the data has an effect on the model’s predictive ability. The noisier the data, the less able the model will be to provide a useful prediction. A prediction interval provides a representation of this impact, producing a range of values where the “true” scrappage rate could be expected to lie with a specified rate of “confidence” (in this case, 95%)²³. Specifically, the CAFE model runs are used to produce an upper bound and a lower bound that define the endpoints of this interval.

²³ Absolutely correct interpretation of confidence and prediction intervals can be rather technical, as discussed in Appendix C. However, for our purposes, these intervals provide a useful representation how random variation in the data affects the model’s predictive ability.

See Figures 4-9a and 4-9b. One issue with this approach is: Two sets of upper and lower bounds are available from these runs (one set for Existing, one set for Rollback). Figures 4-9a and 4-9b use the Existing and Rollback bounds, respectively.

First, consider Figure 4-9a (which uses the 95% bounds from the Existing scenario). There are four lines. The two middle lines are the predicted scrap rates for the Existing and Rollback scenarios. The top dotted line is the upper bound of the 95% prediction interval using Existing runs, and the lower dotted line is the lower bound. It is notable how the uncertainty in predicted scrap rates increases with increasing age. The maximum width occurs at around age 20 (just before the model reaches a trigger at age 21 and switches “modes”). One possible reason for this is the effect of the modeling equation (logit); however, another possible reason is the increased volatility in the observed data as vehicles get older and the population of vehicles declines. (Recall that Bento uses no data past age 14 or 15).

Suppose we want a 95% interval for the predicted scrap rate when a vehicle is 16 years old. Drawing a vertical line at age = 16 yields (approximately) these values: lower bound = 0.063, and upper bound = 0.150. This interval can be represented as: (0.0625, 0.150). The predicted scrap rate for Existing is 0.085 (which must lie between the two bounds). The size and location of this interval represent the uncertainty in the model’s prediction.

Now, consider the predicted scrap rate for the Rollback scenario: It is about 0.113. The values 0.085 and 0.113 both lay within the interval (0.0625, 0.150). The predicted scrap rate for the Rollback lies roughly halfway between 0.085 and the upper bound. This suggests that the difference between the Existing and Rollback predicted scrap rates are small relative to amount of statistical uncertainty.

However, as already noted, it is also possible to compute a prediction interval by using the Rollback scenario results. See Figure 4-9b. These results look very much like those in Figure 4-9a: The width and shape of the 95% prediction intervals from these plots are similar. They are slightly shifted because they are based on Rollback results rather than Existing. However, the basic finding is still the same: The two curves for predicted scrap rates from Existing and Rollback both lay well inside the prediction interval. To allow direct comparison, Figure 4-9c includes both sets of results in the same plot.

So, despite the earlier concerns about the DSM’s unrealistic response to incremental average compliance costs, as well as the general shape of the scrappage curves relative to observed data, these problems may in a sense be moot:

The relatively large uncertainty in the predicted values from the dynamic scrappage model is so large that *the observed differences between the Existing and Rollback predicted scrap rates are not statistically meaningful.*

This is an extremely important observation, because *it is these specific differences that are the ultimate source of the benefit-cost differences between the Existing and Rollback scenarios produced in the Agencies' analysis.* When these scrap rates (and their non-meaningful differences) are replaced with the most recently developed scrap rates available, the Existing standards have positive net benefits versus the Rollback (which reverses the results and conclusions in the NPRM). (See section 2).

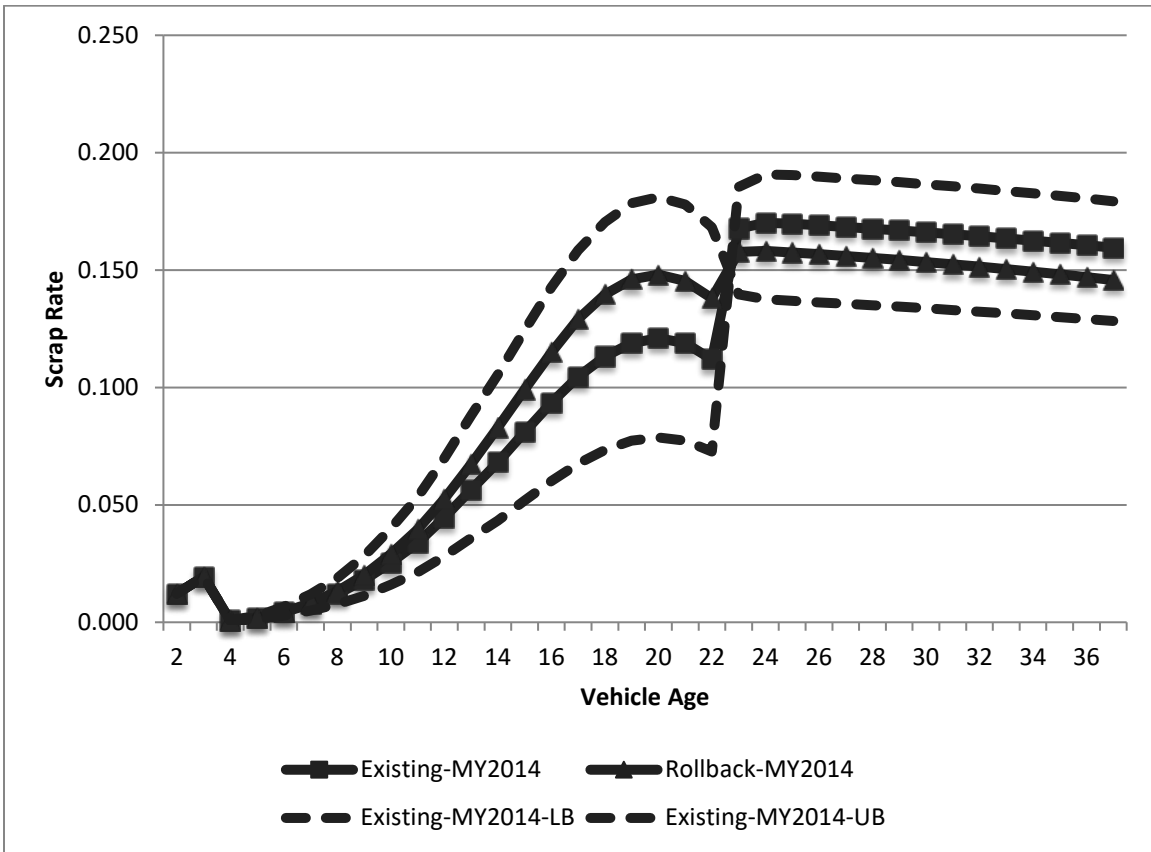


Figure 4-9a. Effect of Statistical Prediction Error on Dynamic Scrap Rates: 95% Confidence Bounds from Existing Standards CAFE Run

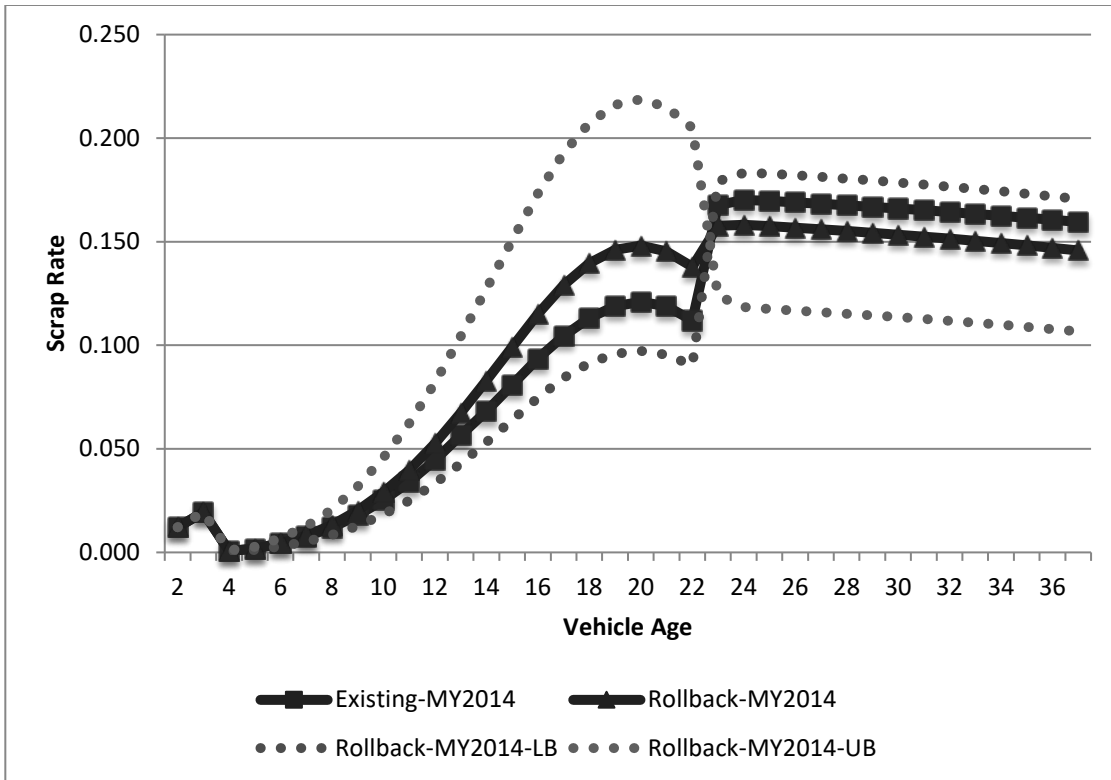


Figure 4-9b. Effect of Statistical Prediction Error on Dynamic Scrap Rates: 95% Confidence Bounds from Rollback CAFE Model Run

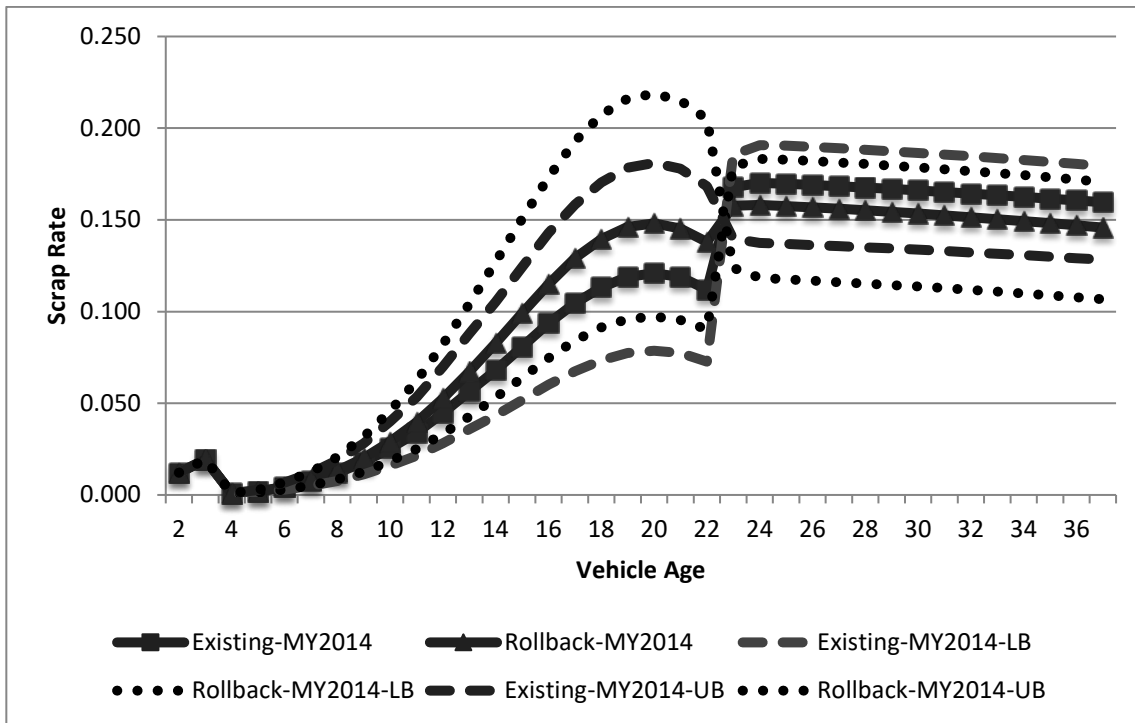


Figure 4-9c. Effect of Statistical Prediction Error on Dynamic Scrap Rates: 95% Confidence Bounds from Both CAFE Model Ru

5. A Comparison of Results from the CAFE Model Results and NEMS

The previous section revealed clear deficiencies in NHTSA's dynamic scrappage model based on isolating its main output (scrap rates). In this section we take a closer look at how the model behaves as a whole, with regard to forecasting the evolution of the entire future light-duty fleet.

We begin in the next section by comparing CAFE model results to the corresponding results from EIA's Annual Energy Outlook (AEO) projections, which are produced by NEMS. This comparison demonstrates that the CAFE model (in contrast to the NEMS model) produces results that are generally inconsistent with economic theory. In succeeding sections, we provide additional results to gain additional insight into reasons for the CAFE model's failure.

5.1 Comparison of Results from the CAFE Model and NEMS

In the past, NHTSA has at times relied on output from NEMS, which has a high level of credibility in policy circles. Figure 5-1 gives projections of total light duty vehicles for 2017-2050, using figures from AEO 2016-2018, CAFE_Existing, and CAFE_Rollback. Note that CAFE uses AEO fuel projections, which is an important factor to have in common.

These plots yield some initial observations about the challenges of producing market projections/forecasts. First, even AEO projections can change quickly in the space of a single year. AEO-2016 and AEO-2017 overlap almost completely, but then there is a major shift in 2018. One reason for including the AEO-2018 projections is that AEO-2018 contains a scenario option for generating results under the Rollback scenario. On the scale used in this figure, the results from Existing versus Rollback differ only slightly. (This will be considered in more detail below.)

Second, the CAFE model projections increase at a much faster rate in the early years than AEO's, and vehicle counts for 2016 are different than the AEO figures (all four of which converge)²⁴. The CAFE results demonstrate notable differences for Existing versus Rollback (which is one reason for performing this evaluation): Fleet sizes get larger under the Existing scenario. At the same time, it could be noteworthy that these differences are of comparable size to the shift in going from AEO 2016/2017 to AEO 2018. In other words, movements of this size can easily occur for a variety of reasons when producing projections.

²⁴ This could be something worth investigating, but we have not done so.

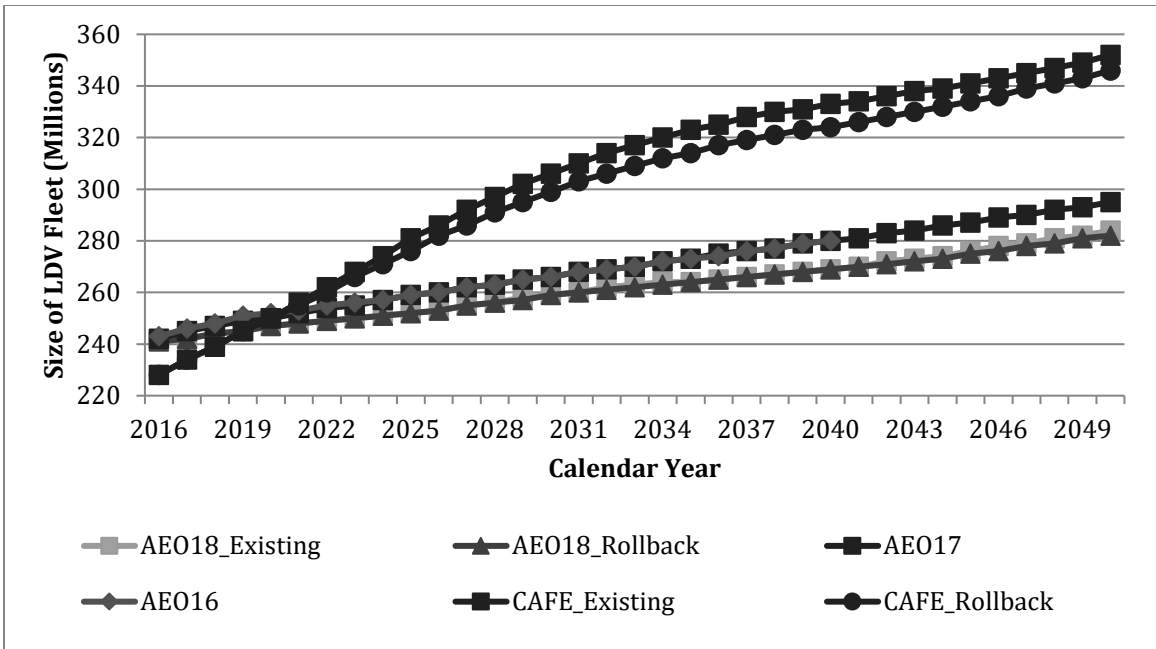


Figure 5-1. Total Light Duty Vehicle Projections (NEMS and CAFE Model)

For the AEO18 results, the fleet size for AEO18_Existing is a little bit larger than for AEO18_Rollback. However, the difference is very small and, and does not begin growing until about 2029. To show this, we plot the difference in fleet sizes (AEO18_Existing - AEO18_Rollback)—see Figure 5-2. The differences are initially about 100K, increasing linearly from 2031 from 200K to 1.8M in 2050. Because even the Existing standards remain at the same level after 2025, this would seem to represent a very different effect from what might be going on in the CAFE model results. This would be consistent with a lower scrappage rate due to the higher value of vehicles that have been produced with greater fuel efficiency.

Next, consider new vehicle sales in Figure 5-3. New vehicle sales for both CAFE results are notably higher than any of the AEO projections (which is perhaps consistent with the total fleet size results). However, the AEO 2016/2017 sales levels are below AEO 2018 levels, while the AEO 2016/2017 fleet sizes were larger. The main difference between AEO 2018 and CAFE is that CAFE has a very steep increase from 2016 to 2021. After 2021, the lines are roughly parallel.

One interesting observation from Figure 5-3 is that CAFE_Rollback sales appear to be slightly larger than CAFE_Existing, whereas for AEO 2018 the reverse is true. These differences can be seen a bit more clearly in Figure 5-4. In the CAFE results, the difference in sales levels begins in 2022 (the first year the policies are different) and the gap stays roughly the same over the entire period. In AEO the differences grow slowly over time, starting in 2027. It is clear that vehicles, by definition, are likely to be more fuel efficient under the Existing standards, and therefore more attractive on this attribute. Fuel prices are the same for both, so there may be some differences, e.g., in assumptions about technological learning.

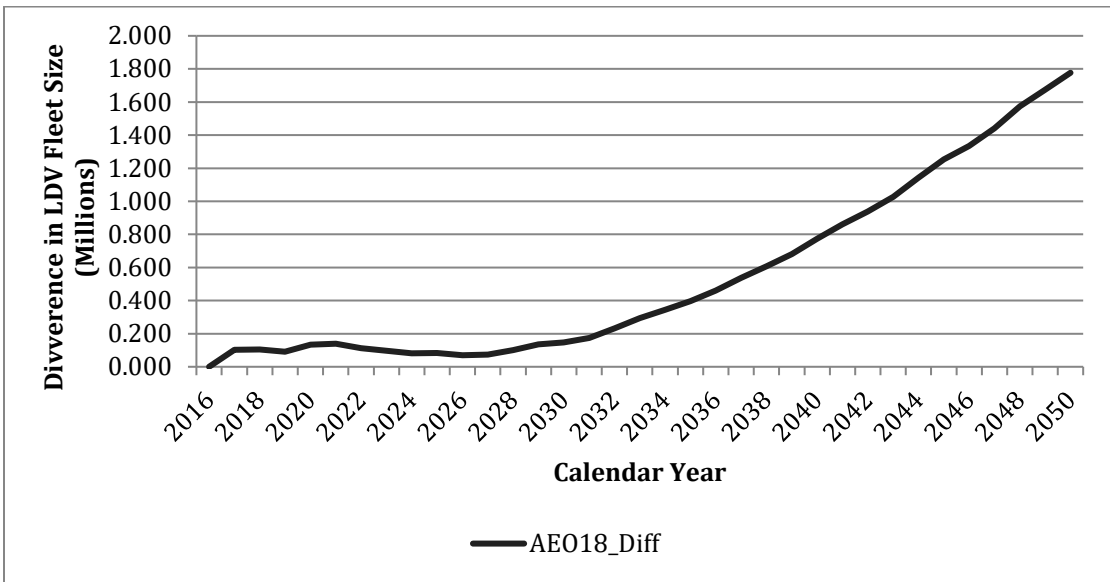


Figure 5-2. Difference in Fleet Size (millions) for AEO18_Existing - AEO18_Rollback

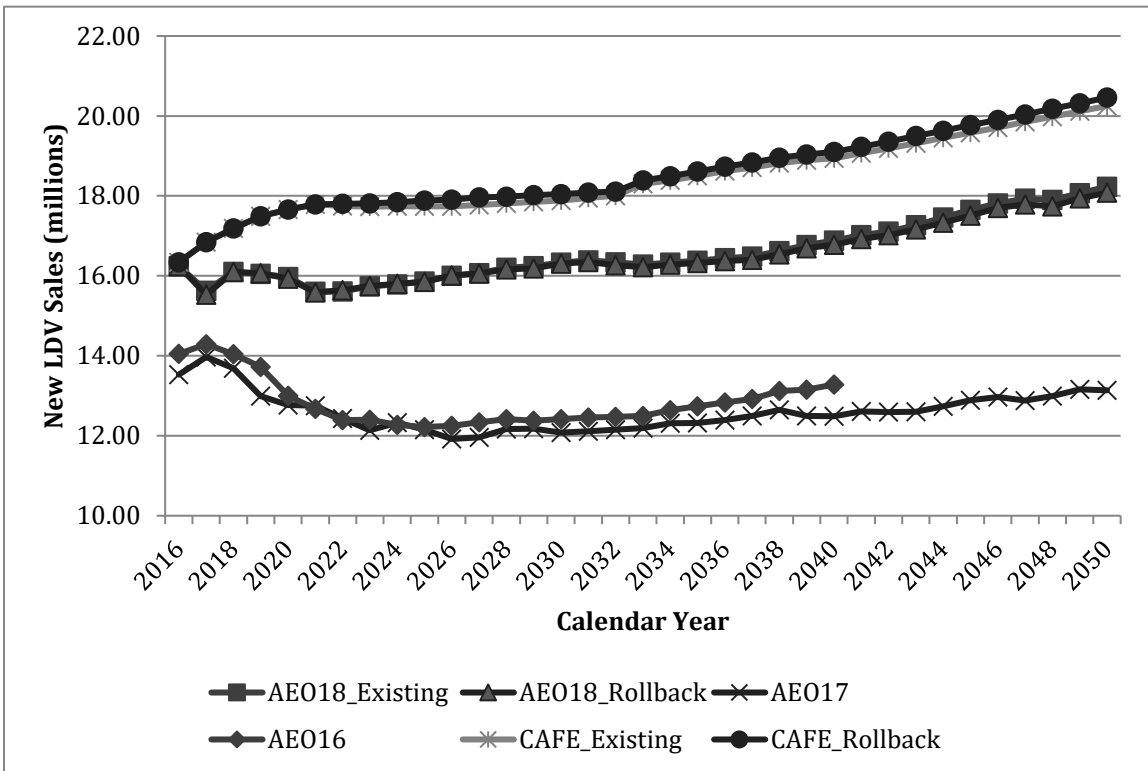


Figure 5-3. New Vehicle Sales (NEMS and CAFE Model)

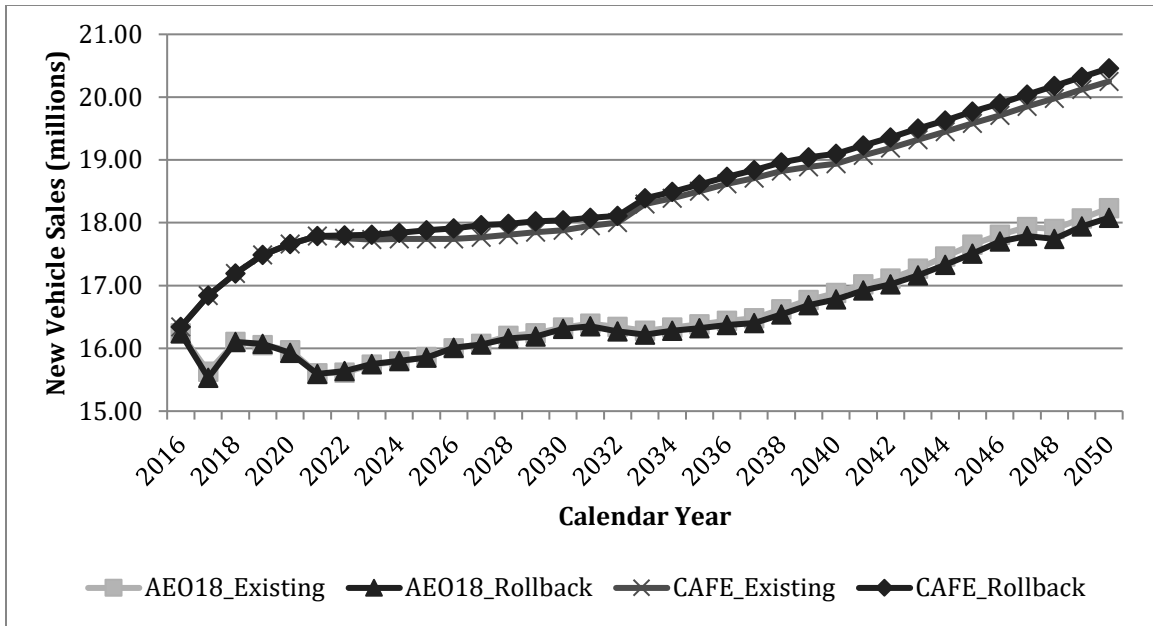


Figure 5-4. New Vehicle Sales (AEO 2018 and CAFE Model)

To summarize: For both AEO18 and CAFE, fleet sizes are larger for Existing than for Rollback. For new vehicle sales, Rollback sales are higher than Existing for CAFE, but the reverse is true for AEO18. These are potentially important qualitative differences. However, here is another potentially important observation: In all four scenarios, new vehicle sales are either growing or flat in almost all years after 2021. (CAFE_Existing shows slight declines only in 2022 and 2023, but no other years. Both AEO18 scenarios show slight declines in 2032 and 2033.)

First, consider AEO18. New vehicle sales generally are growing in both scenarios, so economic theory suggests that fleet sizes should also be growing (they are). Specifically, although the Gruenspecht effect logic suggests that increasing new vehicle sales should lead to increased used vehicle scrap rates, the total “value” of the fleet is increasing, so this would suggest an increase in the fleet size. Moreover, new vehicle sales are higher under Existing, so the fleet size should be also. Based on these observations, *AEO18 results are consistent with economic theory.*

Now, consider CAFE results. The first part of the AEO18 argument is exactly the same: New sales, and fleet sizes, are increasing under both scenarios. However, new sales are higher under the Rollback, so therefore fleet sizes should be larger. But, the opposite is true. *The CAFE results are not consistent with economic theory.*

Recall that, in NEMS, our understanding is that the model produces estimates of total market size and new vehicle sales in a rather direct fashion, so that scrappage is likely to be an inferred/derived quantity. However, in NHTSA’s approach their scrappage model is playing an active role. This (and the results in section 4) suggest that a closer look at scrappage rates is warranted.

See Figure 5-6. The AEO18 results are unremarkable, because we were required to use the total fleet size and new vehicle sales results to estimate scrap rates. So, these will be consistent with the previous results by definition.

However, consider the scrap rates for the CAFE results. They display a pattern consistent with the results in section 4: Scrap rates are lower for Existing than for Rollback, and the differences are *large*. Based on the discussion above, one would expect the Rollback scappage rates to be lower than Existing (because new vehicle sales are larger). However, the opposite is true.

What is the reason for this problem with the CAFE results? Why are they inconsistent with economic theory?

The reasons are explained in the modeling review of section 3. In section 3.2 we established the importance of developing models consistent with theory, and that capture the behavioral and structural features of the market. In particular, the new and used markets are related, and the behavior of both is driven by a combination of consumer preferences and other economic factors.

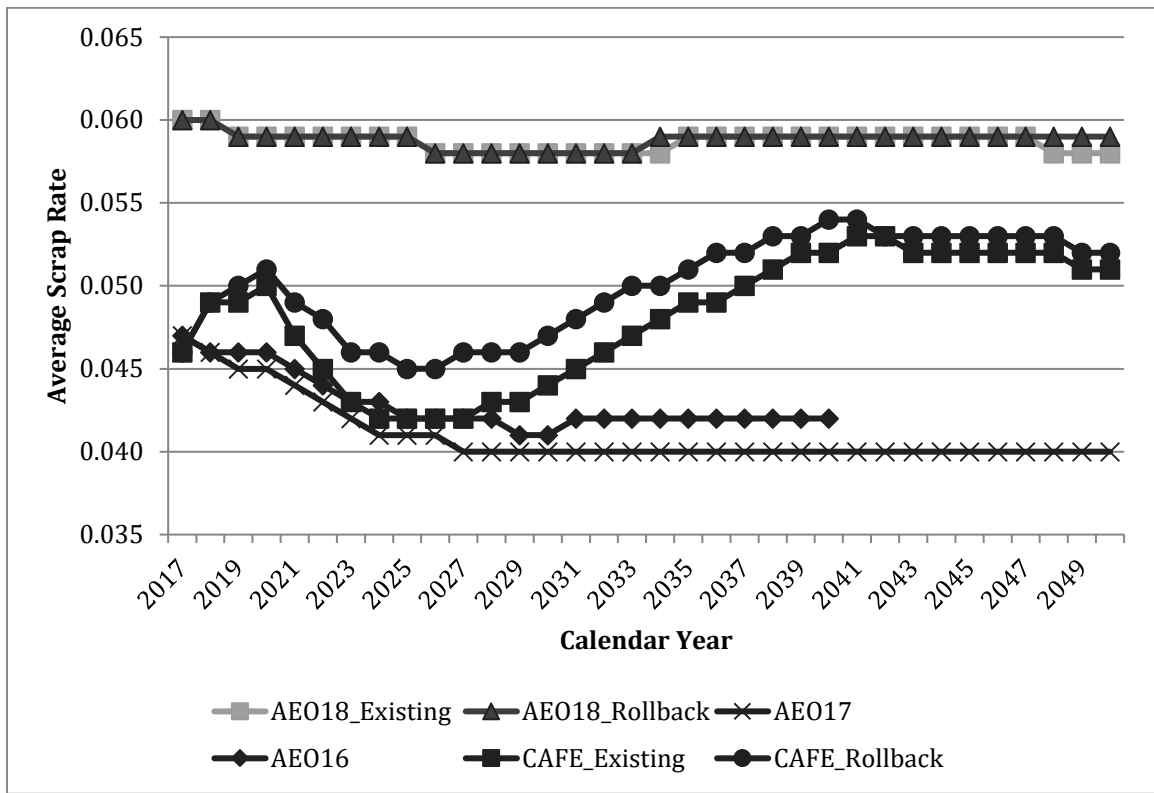


Figure 5-6. Average Scrap Rates (NEMS and CAFE Model)

In section 3.1.2 we pointed out that there was no structural relationship between the new and used vehicle markets in NHTSA’s modeling approach, only an attempt to create a “correlation” based on new vehicle “prices.” However, we also emphasized in section 3.5 that the full *value* of new vehicles need to be taken into

account, and not just their *prices*. (This is also the point of section 3.1.1, which focuses on the common miss-statement of the Gruenspecht effect.)

The failure of the CAFE economic modeling demonstrated above is due to the fact that there are multiple factors driving new vehicle demand (as well as scrappage) other than price and GDP growth, and these were inadequately captured by the reduced form aggregate-level time series approach adopted by NHTSA.

More colloquially: The attempt to create a desired “Gruenspecht effect” by basing both models almost exclusively on new vehicle “prices” was doomed to failure

Put more simply: New vehicle prices and GDP growth rates yielded effects in both the dynamic scrappage and new vehicle sales models

To explore the issue of new vehicle sales and scrappage behavior from another perspective, see Figures 5-7 and 5-8. Figure 5-7 uses AEO results to compute *percentage change from the prior year* for new vehicle sales levels, and scrappage rates, respectively. They are plotted on the same graph for comparison purposes. Figure 5-8 contains the corresponding plots for CAFE model results.

First consider the AEO 2018 results in Figure 5-7. As an overall matter, scrap rates appear to experience very small declines over time, with little variation. They creep into positive territory for the last two years. In contrast, new vehicle sales first decline and then increase with sales growth being positive until the very last year (2032).

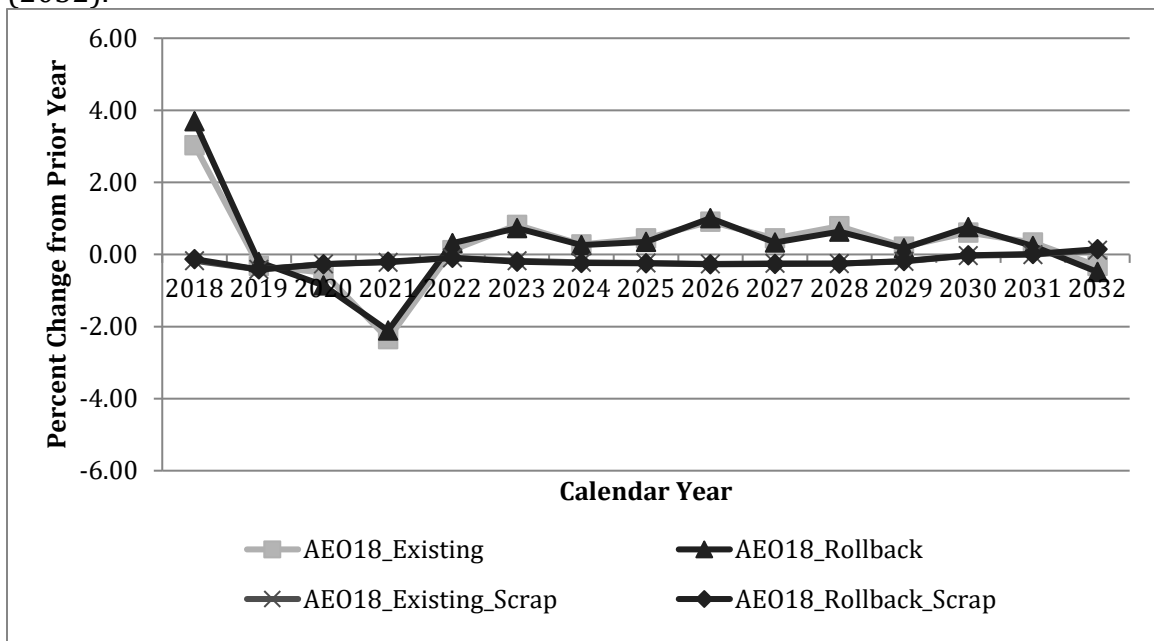


Figure 5-7. Percentage Changes from Previous Year for New Vehicle Sales and Scrap Rates (AEO 2018).

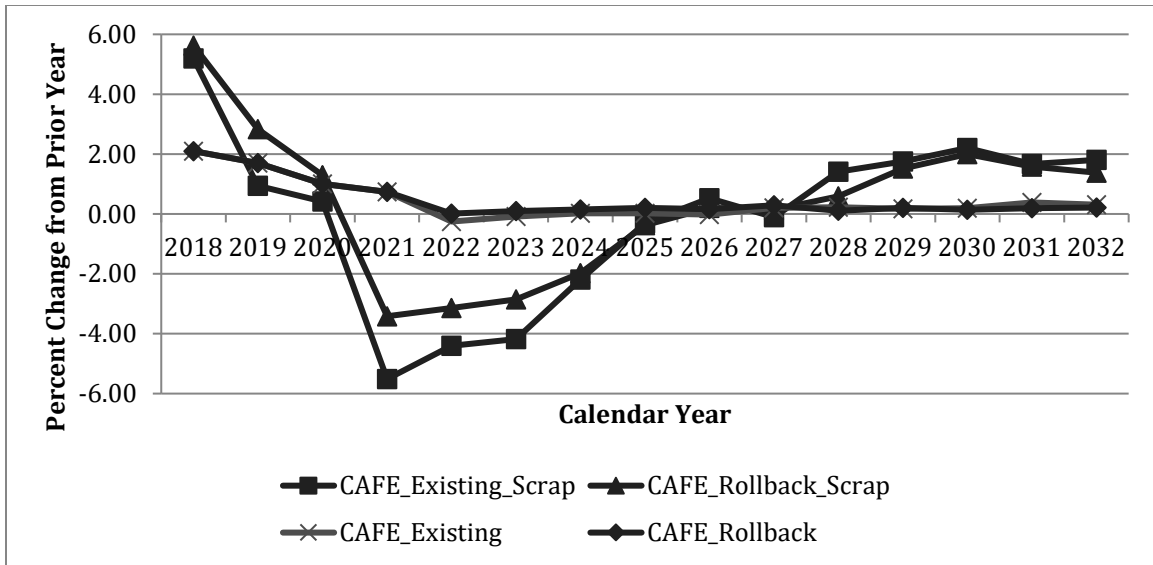


Figure 5-8. Percentage Changes from Previous Year for New Vehicle Sales and Scrap Rates (CAFE Model).

As in previous figures, there are only very small differences between the Existing and Rollback AEO results. As a practical matter, there seems to be very little interaction between new vehicle sales and scrap rates, with both being relatively stable after 2021. It might be tempting to attribute sales declines from 2018 to 2021 to increasingly stringent CAFE standards; however, examination of other NEMS macroeconomic factors suggests that these are due to a short-term projected slowdown in economic growth

Next, consider the CAFE model results in Figure 5-8. In many ways the patterns in this figure are the exact “opposite” of the AEO results. Sales changes are largely flat starting in 2022, slightly positive, with very little change (similar to scrap rate changes in the AEO results, except they are slightly negative). The CAFE results show much larger amounts of variation in scrap rates (analogous to sales in AEO). Both sets of results have their largest sales rate increases in 2018.

With regard to the Gruenspecht effect, these CAFE model results demonstrate that there is very little relationship between changes in new vehicle sales and changes in scrap rates. Sales changes are very small, and stay flat over most of the period. At the same time, changes in scrap rates make large swings from negative to positive. This behavior is largely consistent with the more detailed analysis in section 4.

The “disconnect” between sales and scrappage rates is clearly demonstrated in this figure, providing an empirical demonstration of the previous observations based on theoretical considerations.

Another frequently considered aggregate measure is “turnover rate,” which (as discussed in section 3) is usually replaced by the ratio of new vehicle sales to fleet

size. (We will use the term turnover rate for this measure.) For AEO and CAFE model results, see Figure 5-9. In this figure we have expanded the range of years (2017-2015).

Based on experience with the previous results, the patterns here are not surprising. The turnover rate is “flatter” for the AEO results when compared to the CAFE model results (although one must consider the scale being used). In previous figures we looked at results through 2032, and the behavior in this figure is consistent with previous results, with everything going flat afterwards (this is consistent with the way the CAFE model works in years after 2032). As with previous results, there could be a question about differences between AEO and CAFE model results for the initial year. But, again, scale is potentially an issue.

In looking at the CAFE results, we see patterns that are consistent with what has already been learned about the scrappage model. On this scale, turnover rate drops rather quickly from 0.072 (a figure we have seen cited in online articles) to 0.060 and below. So, all CAFE results yield a drop in fleet turnover, compromising the efficacy of any CAFE policy. Also, as in previous results, there is a clear gap between the curves for the Rollback and Existing scenarios. By 2032 this is on the order of 0.02, which is about 3%.

On one hand, this may not be considered large. On the other hand, the effect associated with this gap is what gives rise to NHTSA’s claim that the Rollback has substantial net benefits relative to the Existing standards (as discussed in section 4).

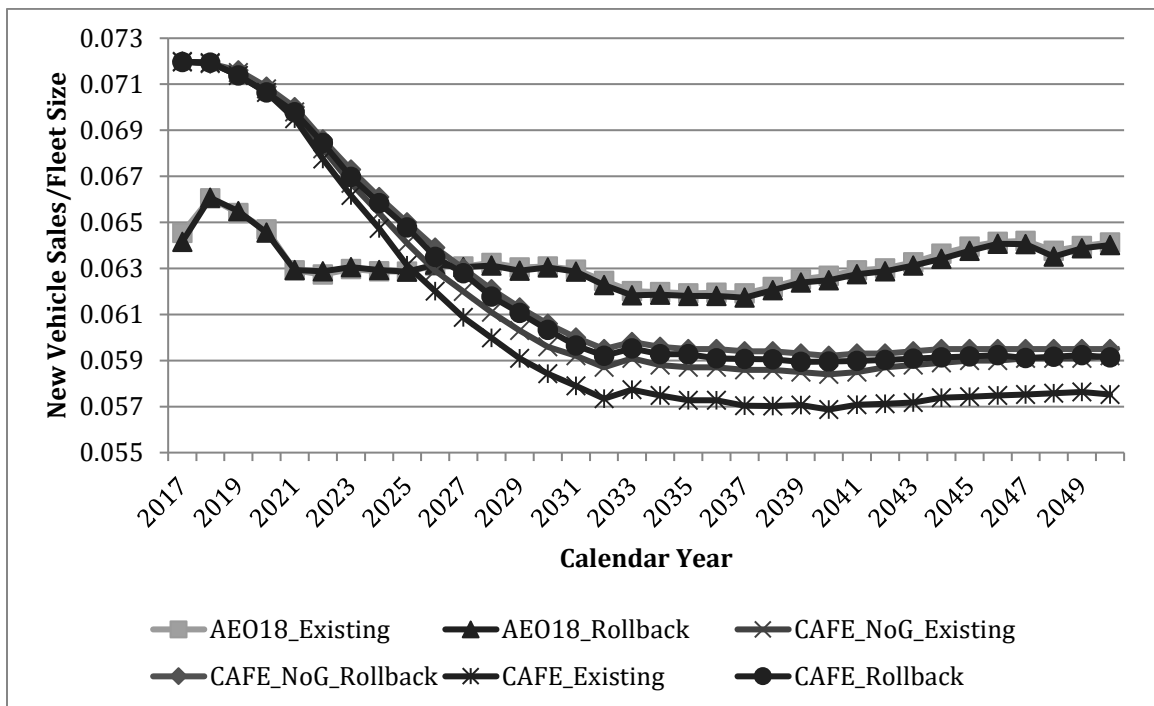


Figure 5-9. Turnover (New Vehicle Sales/Fleet Size) for AEO and CAFE Results

There has been much discussion about the dropping turnover rate in the US vehicle fleet. To put all of this in perspective, we obtained data from the Transportation Energy Data Book to show the much larger trend. See Figure 5-10.

First, over this time scale we can see that: (1) actual observed turnover rates can vary over a wide range, and (2) as has been recently reported, turnover has been experiencing a downward trend, from 0.10 in 1970 to roughly 0.07 in 2015 (with a large dip during the great recession). On this scale, the differences between AEO and CAFE look rather small, and suggest a continuation of a downward trend to a level of 0.06 on average.

This perspective is entirely consistent with the results reported in section 4 (see, e.g. Figure 4-9) that demonstrate how, *due to the statistical error in the dynamic scrappage model, the difference between the Rollback and Existing vehicle fleet-related results* are not meaningful in a practical sense.

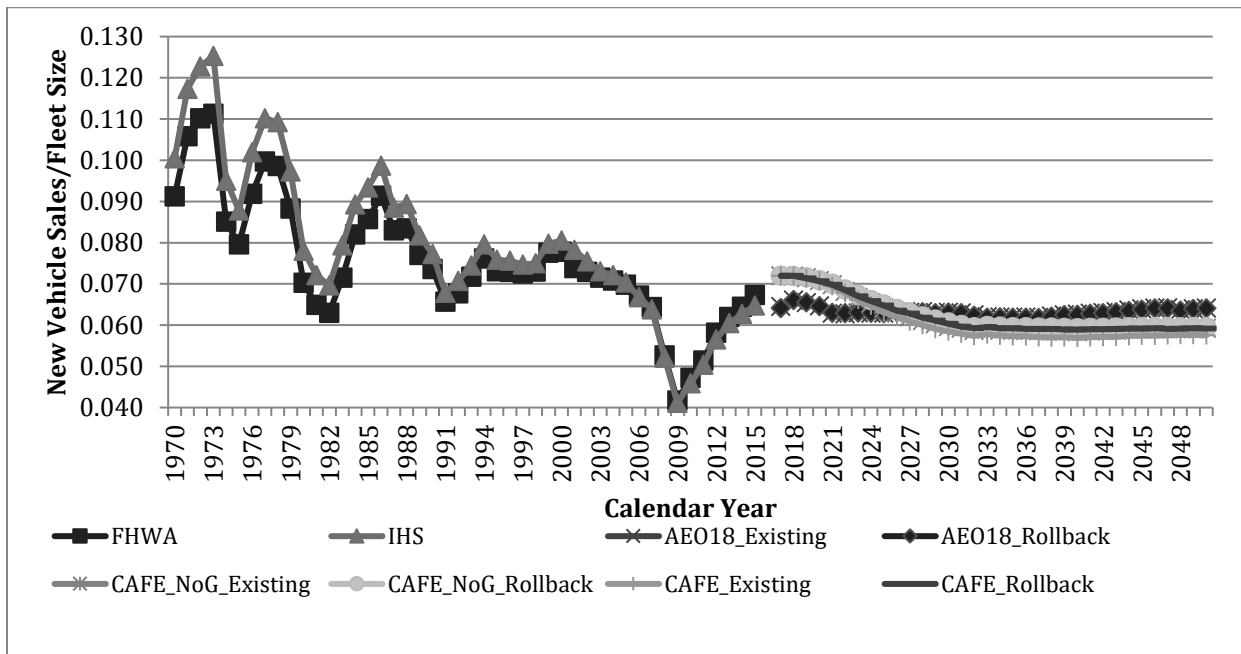


Figure 5-10. Turnover Comparison: FHWA Data and AEO/CAFÉ Model Results

5.2 Additional Exploration of CAFE Model and NEMS Behavior

One thing we have not yet directly is the potential role of new vehicle *prices* (more specifically, new vehicle price *increases* attributable to differences in compliance costs under different CAFE policies). Both the CAFE model and NEMS execute procedures to ensure that manufacturers make decisions so as to comply with CAFE, which will generally require adoption of new technology. Because this can place upward pressure on prices, it is instructive to compare average prices from the two models.

See Figure 5-11. A number of features are clear. AEO18 prices are systematically *lower* than CAFE model prices. Although there could be any number of reasons for this, this is at least consistent with the higher sales levels in AEO18. The price *differences* between the Existing and Rollback scenarios are *smaller* for AEO18 than for the CAFE model, which is a reminder of potential concerns regarding the Tech Cost results in the 2018 NPRM, and much they have changed since the 2016 TAR. Finally the pattern of AEO prices changes appears to track to the compliance schedule much more closely than the CAFE model results, which would be an indication of differences in the algorithms used the manufacturer decision models for two systems.

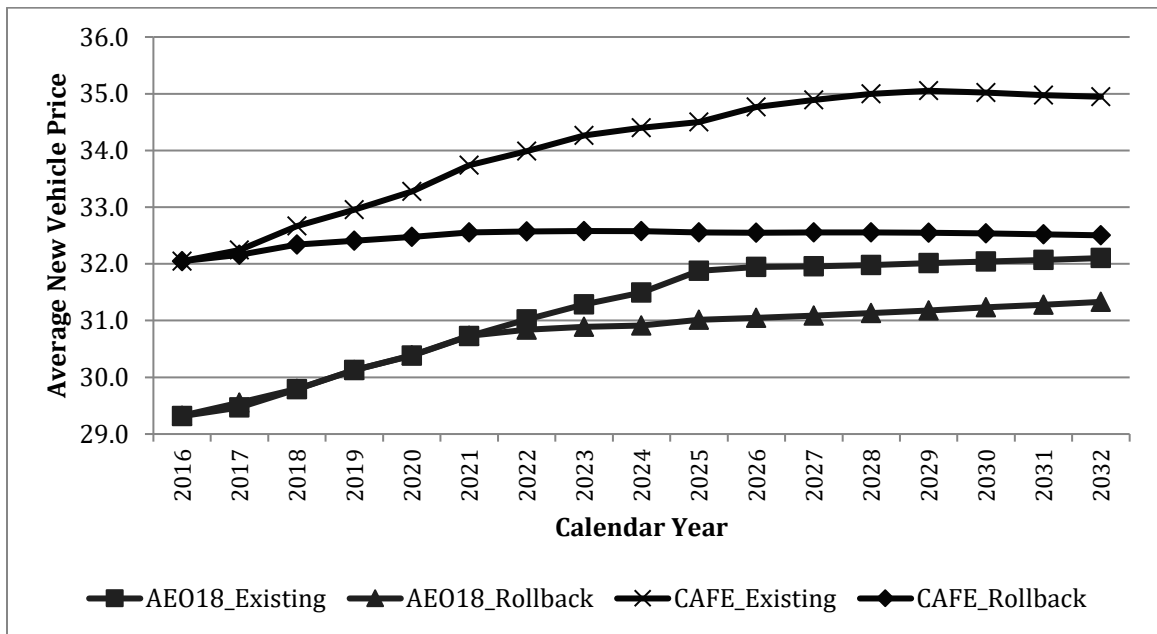


Figure 5-11. Average New Vehicle Prices from AEO18 and CAFÉ Model.

Returning to economic considerations, it is worth remembering that *neither of these two models has any type of behavioral model for the used vehicle market, nor do they have a representation of used vehicle prices.* On the other hand, NEMS *does* have a discrete choice model (nested logit) that yields new sales shares for a relatively large number of vehicle classes and fuel technology types. Scrapage is an implied behavior determined by projecting total fleet size and new vehicle sales. Through this mechanism, all else equal, an increase in new vehicle sales would yield an increase in scrapage. In this way, consumer responses to price changes in the new vehicle market would influence scrapage²⁵.

²⁵ Note that this represents yet a third choice when compared to GC99 and the current CAFÉ model. In GC99, total vehicle stock and scrapage are modeled, which determines new vehicle sales. The CAFÉ model forecasts new vehicle sales and scrapage, which determines vehicle stock. Based on earlier discussions, our view is that it is always preferable to model future levels of vehicle stock.

In contrast, in the CAFE model’s overall fleet size is an implied “behavior” determined by whatever results are produced from the new vehicle sales model and the scrappage model. As discussed previously, the outputs of these two models are theoretically “correlated” because they both use new vehicle price as an explanatory variable, but there is no cause-and-effect mechanism that directly links new sales demand to scrappage rates. New vehicle price changes affect scrappage rates directly by being an explanatory variable in the model.

We have no independent source for an estimate of what the likely impact on scrappage would be from *new* vehicle price increases (i.e., the scrappage elasticity for changes in new vehicle prices). Recall that we do have such estimates for *used* vehicle prices (see section 3). Recall that Bento et al. (2018) report an estimate of -0.4 based on models using aggregate data. Jacobsen and van Benthem (2015) report a somewhat higher value using models estimated on highly disaggregated data. They estimate multiple models yielding a range of estimates that depend on various vehicle characteristics, reporting an overall “central estimate” of -0.7.

Now, note that, with the numerical results we have compiled in this section, we can compute estimates of scrappage elasticities for new vehicle price changes. Specifically, we can compute the percent changes in new vehicle prices using the available data. Then, for each calendar year, we can divide the percentage change in scrap rate by the corresponding percentage change in new vehicle price. This provides a measure of elasticity. See the results in Table 5-1.

Year	AE018_Exist	CAFE_Exist	AE018_Rollback	CAFE_Rollback
2018	-0.16	3.98	-0.16	10.30
2019	-0.37	1.07	-0.38	12.63
2020	-0.31	0.42	-0.31	6.63
2021	-0.19	-3.97	-0.18	-13.47
2022	-0.10	-5.92	-0.28	-60.47
2023	-0.21	-5.16	-1.22	-142.79
2024	-0.33	-5.68	-3.12	163.88
2025	-0.19	-1.01	-0.73	6.62
2026	-1.27	0.68	-2.59	-8.92
2027	-5.65	-0.32	-2.17	11.83
2028	-3.37	4.45	-1.81	54.81
2029	-2.11	11.53	-1.13	-69.56
2030	-0.37	-23.33	-0.14	-46.43
2031	-0.04	-13.72	0.06	-32.59
2032	1.12	-21.46	0.89	-26.41
average	-0.90	-3.90	-0.88	-8.93

Table 5-1. Estimates of Scrap Rate Elasticity With Respect To Changes In New Vehicle Price

In comparing the two sets of results, the AEO18 values are much more reasonable than the CAFÉ model values. In the above table we have identified in **bold** the entries that are reasonable based on economic theory and the results in the literature. In this case, elasticities should be negative, and generally be smaller than 3 (in absolute value.) The vast majority of AEO18 values satisfy this condition. Even the larger AEO18 values are “reasonable.” Finally, the average values are -0.90 and -0.88 for the Existing and Rollback scenarios, respectively. On one hand, these are reasonably close to the Jacobsen and van Benthem (2015) estimate for scrap elasticity with respect to *used* vehicle prices. On the other hand, the Bento et al. (2018) estimate was -0.4, and one might expect the elasticity with respect to new vehicle price to be smaller. In any case, these results are not unreasonable.

However, the results from the CAFE model do not fare nearly so well. Although the negative values outnumber the positive ones (17 to 13), this is a relatively even mix of negative and positive values (and therefore clearly inconsistent with economic theory). Most of the values are extremely large in absolute value. (Given the mix of positive and negative values, the average values are not even valid for consideration.)

Considering that the dynamic scrappage model was specifically intended as a direct implementation of the Gruenspecht effect by using new vehicle “prices” as an explanatory variable, this further demonstrates of the failure of the approach. One possible reason for the relative success of NEMS results is the cause-and-effect structure describe above, combined with a discrete choice model of consumer behavior for new vehicle sales.

Is it possible to explain in more detail while this approach did not work? Recall that GC96 and Parks (1977) do use a variable related to new vehicle price in their scrappage models (specifically, the new vehicle price index). However, recall that NHTSA’s approach used unadjusted new vehicle prices (as discussed in section 3.6). But, this completely ignores the well-known phenomenon occurring in new vehicles: they are constantly being improved using rapid technological advances in multiple areas, so that quality-adjusted new vehicle prices have been dropping steadily.

However, the bigger problem is that these models are just very limited when compared to approaches that use, e.g., discrete choice models of consumer behavior. In this regard, recent research reinforces how important it is to use a relatively high level of detail when estimating these models in order to get unbiased parameter estimates. See, e.g., Wong, Brownstone, and Bunch (2018), who show that, even discrete choice models based on a relatively large number of vehicle classes are subject to problems with bias due to the use of attribute averages. However, at the same time, we recognize that these approaches could be very challenging to use by many analysts due to their technical requirements.

6. Conclusions

The purpose of this report was to provide a review and analysis of NHTSA's economic modeling approach, and determine any implications for the benefit-cost analysis results in the NPRM. The reasons for specifying the purpose in this way were: (1) a preliminary review of the NPRM revealed there were major differences between the results and conclusions in the NPRM versus the 2016 TAR, (2) it was unlikely that such a major reversal was due to "new information," (3) further review revealed that the reversal hinged largely on the benefit-cost analysis, and (4) one of the major changes made since 2016 was the introduction of economic models that seemed motivated by a desire to implement the Gruenspecht effect.

As noted elsewhere, there are also a number of other problems with NHTSA's analysis related to issues such as technology costs, the rebound effect, etc., but these are outside the scope of this report.

This report began in section 2 by reviewing benefit-cost measurement issues, and showing that the dynamic scrappage model was a major driver of the results in the NPRM. Specifically, if the dynamic scrappage model is turned off and replaced with the most recently developed scrappage rates made available by the Agencies, *the results and conclusions of the NPRM are reversed in favor of the Existing standards.*

The remainder of the report shows that the difference in results is not due to an improvement in the "quality" of the modeling employed by the Agencies. Rather, the evidence is that the models were inherently limited based on theoretical considerations, were implemented using questionable approaches, and were not subjected to rigorous testing, validation, or peer review.

In particular, we provide a thorough analysis using numerical results informed by comparison to other models in the literature, and also with NEMS. These results *empirically* demonstrate that the dynamic scrappage model has many serious flaws. It behaves in a manner that is inconsistent with theory, and, in particular, produces results for vehicle market behavior that are inconsistent with economic theory (violating the Agencies' own requirements for regulatory analysis). Finally, the differences between the Existing and Rollback net benefit results in the NPRM can be directly attributed to specific differences in scrappage rates that are not statistically meaningful due to the prediction error properties of the dynamic scrappage model.

Bibliography

- Bento, A. et al. (2018), "Vehicle Lifetime Trends and Scrappage Behavior in the U.S. Used Car Market." *The Energy Journal*, vol. 39, no. 1, Jan. 2018, doi:10.5547/01956574.39.1.aben
- Bates, J. (2008), "History of Demand Modelling," Handbook of Transport I: Transport Modeling (Second Edition), in David A. Hensher and Kenneth J. Button, editors, Pergamon, (2008) pp. 11-34.
- Berkovec, J. (1985), "New Car Sales and Used Car Stocks: A Model of the Automobile Market," *Rand Journal of Economics*, Vol. 16, no. 2 (Summer), pp. 195-214.
- Bunch, David S. (2009), "Follow-on Development of CARBITS: A Response Model for the California Passenger Vehicle Market," Final Report (Contract 05-303) prepared for State of California Air Resources Board, April 30, 2009.
- Bunch, D. S. and B. Chen (2008), "Automobile Demand and Type Choice," Handbook of Transport I: Transport Modeling (Second Edition), in David A. Hensher and Kenneth J. Button, editors, Pergamon, (2008) pp. 463-479.
- Energy Information Agency (2009), "The National Energy Modeling System: An Overview 2009," (October), Energy Information Administration, Office of Integrated Analysis and Forecasting U.S. Department of Energy Washington, DC 20585 (www.eia.doe.gov/oiaf/aeo/overview/).
- Environmental Protection Agency, National Highway Traffic Safety Administration, and California Air Resources Board (2016), "Draft Technical Assessment Report: Midterm Evaluation of Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards for Model Years 2022-2025," EPA-420-D-16-900 July 2016.
- National Highway Traffic Safety Administration (NHTSA) and Environmental Protection Agency (EPA) (2018a), "Preliminary Regulatory Impact Analysis: The Safer Affordable Fuel-Efficient (SAFE) Vehicles Rule for Model Year 2021-2026 Passenger Cars and Light Trucks," July 2018.
- National Highway Traffic Safety Administration (NHTSA) and Environmental Protection Agency (EPA) (2018b), Notice of Proposed Rulemaking (NPRM), *Federal Register*, Vol. 83, No. 165, Friday, August 24, 2018, Proposed Rules, pp. 42986- 43500.
- Gilbert A. Q. and B. K. Sovacool (2016), "Looking the Wrong Way: Bias, Renewable Electricity, and Energy Modelling in the United States," *Energy*, 94, pp. 533-541.

- Gillingham, K. (2018), "The Rebound Effect of Fuel Economy Standards: Comment on the Safer Affordable Fuel-Efficient (SAFE) Vehicles Proposed Rule for Model Years 2021-2026 Passenger Cars and Light Trucks," October 19.
- Greenspan, A. and Cohen, D. (1999), "Motor Vehicle Stocks, Scrapage, and Sales," Review of Economics and Statistics, vol. 81, no. 3, 1999, pp. 369-383, doi:10.1162/003465399558300.
- Gruenspecht, H. K. (1982), "Differentiated Regulation - The Case of Auto Emissions Standards," American Economic Review, American Economic Association, 1 Jan. 1982, econpapers.repec.org/RePEc:aea:aecrev:v:72:y:1982:i:2:p:328-31.
- Jacobsen, M. R. and van Benthem. A. A. (2015), "Vehicle Scrapage and Gasoline Policy," American Economic Review, vol. 105, no. 3, 2015, pp. 1312-1338, doi:10.1257/aer.20130935.
- Liu, C., D. L. Greene, and D. S. Bunch (2014), "Fuel Economy and CO2 Emissions: Standards, Manufacturer Pricing Strategies, and Feebates," with C. Liu and D. L. Greene, The Energy Journal, Vol. 35, No. 3 (2014), pp. 71-89.
- Parks, R. W. (1977), "Determinants of Scrapping Rates for Postwar Vintage Automobiles," Econometrica, vol. 45, no. 5, 1977, p. 1099, doi:10.2307/1914061.
- Train K. (1986), Qualitative Choice Analysis. MIT Press, Cambridge, MA.
- United States Department of Transportation-National Highway Traffic Safety Administration (2018), CAFE Model Peer Review, DOT HS 812 590, July [https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/812590-cafe-peer-review.pdf, downloaded September 11, 2018.]
- Walker, F. V. (1968), "Determinants of Auto Scrapage," The Review of Economics and Statistics, vol. 50, no. 4, Nov. 1968, pp. 503-506.
- Wong, T., D. Brownstone, and D. S. Bunch (2018), "Aggregation Biases in Discrete Choice Models," Journal of Choice Modelling, In Press, (available online February 2018), <http://dx.doi.org/10.1016/j.jocm.2018.02.001>.

Appendix A: Potential Impact of Dynamic Scrappage Model Statistical Error

The results in Tables 2-1 to 2-3 were all presented in terms of *net benefits*. However, the outputs from the , model are cost estimates. Recall that the original net benefit estimate corresponding to the NPRM result was \$196.6B (column 1 of Table 2-3). This estimate was obtained from *two* different runs of the CAFE model. The total cost estimate of the Rollback is \$16,338.34B (from one run), whereas the total cost estimate for the Existing standards is \$16,534.91B (from another run). The net benefit is obtained by subtraction: \$16,534.91B - \$16,338.34B = \$196.57B.

However, because of the presence of a statistical model (dynamic scrappage) inside the CAFE model, the prediction errors in the scrappage model imply that there must be some amount of uncertainty in the final results.

In section 4 we discuss the construction of a 95% confidence interval for the error in predicted scrappage rates from the scrappage model. Using the reported results in the PRIA () made it possible to construct an upper and lower bound defining a 95% confidence interval. We obtained CAFE model runs to generate curves in section 4 that show the scrappage model prediction error.

However, these same runs can be used to complete the calculations used to produce total cost estimates.

Existing:	Central estimate =	\$16,534.9B
	Bounds = (\$15,351.8B,	\$17,545.0B).
Rollback:	Central estimate =	\$16,338.3B
	Bounds = (\$15,160.8B,	\$17,376.6B).

The purpose of showing these results is to demonstrate that the potential impact of prediction error from embedding a model like the scrappage model within the CAFE model is *an issue that should have been investigated*. We caution that we are not claiming that these final results are necessarily “95% confidence intervals”, nor do the conclusions and findings of this report rely on the results shown in this appendix. Rather, this speaks to the need to rigorously test and validate models before using them for the purpose of making important policy decisions.

Appendix B. Brief Description of National Energy Modeling System (NEMS)

EIA's NEMS model frequently plays an important role in policy analysis (including at the state level): Its projections are widely used by researchers performing such analyses, representing a type of "standard" used to define reference scenarios. NEMS is used by EIA to produce its Annual Energy Outlook, which includes projections of important macro-level statistics. For example, fuel price projections (including reference, low price, and high price scenarios), are widely used in policy analysis (as they are in the CAFE model). Moreover, in our own work we have also used AEO projections of future vehicle fleet sizes and new vehicle sales for purposes of model calibration.

Although a number of researchers have concerns about possible biases in specific sub-sectors (e.g., renewable energy), even those researchers characterize NEMS/AEO as follows (Gilbert and Sovacool 2016):

"Released annually, AEO contains long-term projections of energy supply, demand, and prices in the U.S [10]. AEO projections are relied upon by industry, government, academia, and the public sector for regulatory proceedings, rulemakings, environmental projections, financial decisions, creating other energy models, and more.

... One colleague of ours even refers to it colloquially as 'The Bible of energy information.'

Indeed, many high-profile regulatory proceedings in the U.S. rely on AEO or NEMS to assess the costs and benefits of regulatory policies. ... One of the major challenges with energy economic models is a lack of transparency: it is usually difficult or impossible for third parties to be able to "independently verify published results" [14]. Unlike other energy models, AEO projections have been published for many years and are well documented, making them a prime candidate to test the effectiveness of energy model projections."

Based on the preliminary description of NEMS capabilities in the area of vehicle markets in section 3.4, it is evident that includes a much more advanced version of the same type functionality that NHTSA has attempted to add to the CAFE model (under a very short time frame), incorporating many of the desirable features described in section 3.2 (discrete choice models for vehicle classes based on attribute preferences, equilibration, and manufacturer decision making at a high level of detail). The remainder of this appendix includes additional detail on NEMS, which draws heavily from the overview provided in EIA (2009).

NEMS is a large-scale modeling system that incorporates many components, managed by an "integrating module" that performs iterations of the entire system until convergence to a general equilibrium (analogous to the Berkovec framework,

although with a much larger scope). It addresses essentially all sectors of the economy, since all sectors have some implications for future energy use.

At the “top level,” NEMS addresses “domestic spending (I), income (II) and tax policy (III) sectors” and simulates the “the central circular flow of behavior as measured by the national income and product accounts.” Domestic spending is decomposed into according to a hierarchy of categories, including consumer spending on durable goods. Durable goods spending “is divided into nine categories: light vehicles; used automobiles; motor-vehicle parts; other vehicles; computers; software; other household equipment and furnishings; ophthalmic and orthopedic products and ‘other’.” Consumer spending on non-durable goods and services is similarly subdivided.

In terms of factors that affect these projections: “In nearly all cases, real consumption expenditures are motivated by real income and the consumer price of a particular category relative to the prices of other consumer goods. Durable and semi-durable goods are also especially sensitive to current financing costs, and consumer speculation on whether it is a ‘good time to buy’. Clearly, overall macroeconomic effects that affect future vehicle usage are taken into account.

One of the many modules used by NEMS is the Transportation Demand Module, which is described as follows: “The transportation demand module (TRAN) projects the consumption of transportation sector fuels by transportation mode, including the use of renewables and alternative fuels, subject to delivered prices of energy and macroeconomic variables, including disposable personal income, gross domestic product, level of imports and exports, industrial output, new car and light truck sales, and population.”

The module makes extensive use of data on vehicle technology capabilities and costs, and *models the decisions by manufacturers to add new technologies*. The demand side incorporates a *vehicle choice model* that includes detailed vehicle classes, including the capability to address future alternative fuel technologies. It captures the effect of tradeoffs among different types of vehicle attributes (e.g., price and fuel operating cost), as well as the degree of substitution and competition among similar vehicle types. Moreover, projections of future vehicle counts are provided at this level of detail.

An important feature is that NEMS is specifically designed to take into account the effect of CAFE standards:

Proposed changes in CAFE standards: This class of simulations is based on changing (increasing) the combined average fuel economy of new light vehicles relative to the baseline CAFE standards. Increases in the CAFE standards are associated with an increase in the cost of production of new light vehicles, which are calculated by the Transportation Module of the NEMS. This increased cost is passed to the MAM. The additional cost per new light vehicle is added to the reference average price of new light duty vehicles (PLVAVG).

Once the MAM solves its series of models using the new assumption, it writes its new projection to the global data structure. The other modules of the NEMS read the new MAM and CAFE assumptions and recalculate their projections. The resulting new energy prices and quantities along with the incremental cost for new light vehicles are returned to the MAM. The MAM uses the newly estimated energy market assumptions to re-solve. This process continues until the NEMS forecast converges.

In the 2018 AEO, NEMS was used to produce results for two different regulation scenarios (Existing and Rollback), facilitating the comparison study in section 5.

Appendix C. Prediction Uncertainty in the Dynamic Scrappage Models

NHTSA's scrappage model development is documented in section 8.10 of the PRIA. As discussed in section 3.5, three other references in the literature on this subject are Bento et al. (2018), Jacobsen and van Benthem (2018), and Greenspan and Cohen (1999). The models in all of these references are developed by first adopting a particular form of equation (or, in the case of a two-stage model, equations) and then "calibrating" or "estimating" the model parameters by fitting it to observed data.

The most familiar version of this exercise in introductory statistics is linear regression, where a straight-line formula is fitted to observed data. In this case, the researcher assumes that the following is the "true model" (with unknown parameters α , β , and σ):

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

where x_i is an explanatory variable and y_i is the dependent variable that the researcher is interested in understanding and/or predicting. Pairs (y_i, x_i) , $i = 1, \dots, n$, are observed data that can be used to obtain estimates of the parameters. In this model, y_i is subject to random variation due to unobservable effects on its value. Specifically, the true model explains the *average* value of y_i (\hat{y}_i) for a particular value of x_i (given by $\hat{y}_i = \alpha + \beta x_i$), and the observed value of y_i is subject to random variation (ε_i). The mean and variance of ε_i are 0 and σ^2 , respectively. The variance is a measure of how much "noise" there is in the observed values (the larger the y_i , the more noise).

Using the observed data, the researcher finds estimates for α and β (called a and b , respectively) that provide the best "fit" to the data, as well as s (an estimate of the noise). Figure B-1 shows the output of this process, with a scatterplot of observations and a fitted line. The fitted line gives the predicted value from the model at various values of X . By plotting this together with the observations, it is possible to see the amount of random variation in the data. A measure of this error

is the root-mean-square-error: $\sqrt{\left(\sum_{i=1}^n (y_i - \hat{y}_i)^2\right) / n}$.

It is possible to create an upper and lower bound to define a prediction interval with a 95% confidence level. The interpretation of this is: "If this process of collecting data and performing this analysis were conducted over and over again (with a correct model), then the true value of Y would lie within this interval 95% of the time." For an example of a similar plot that also includes upper and lower bounds for a 95% prediction interval, see Figure B-2.

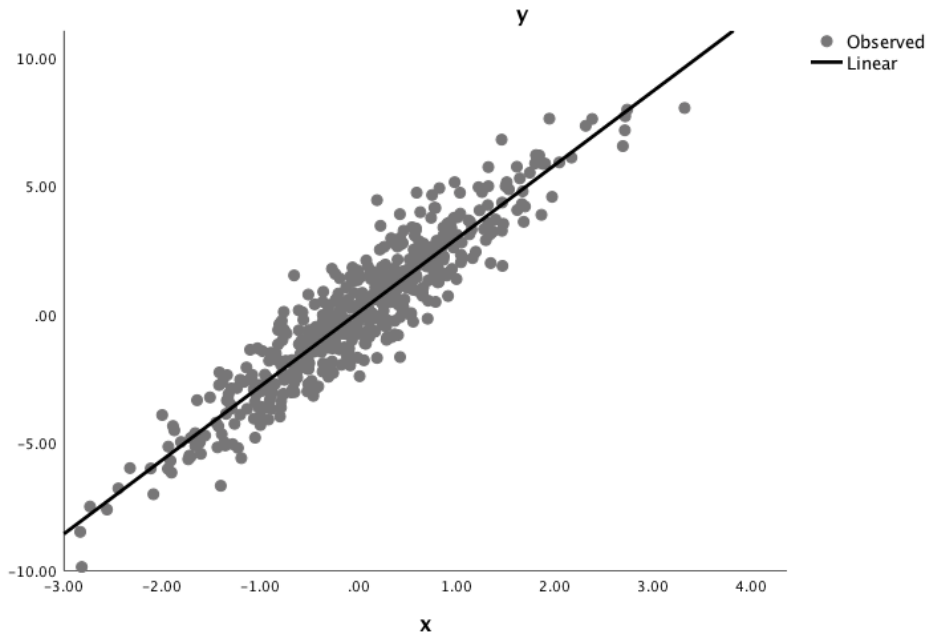


Figure B-1. Straight Line Fitted to Observations

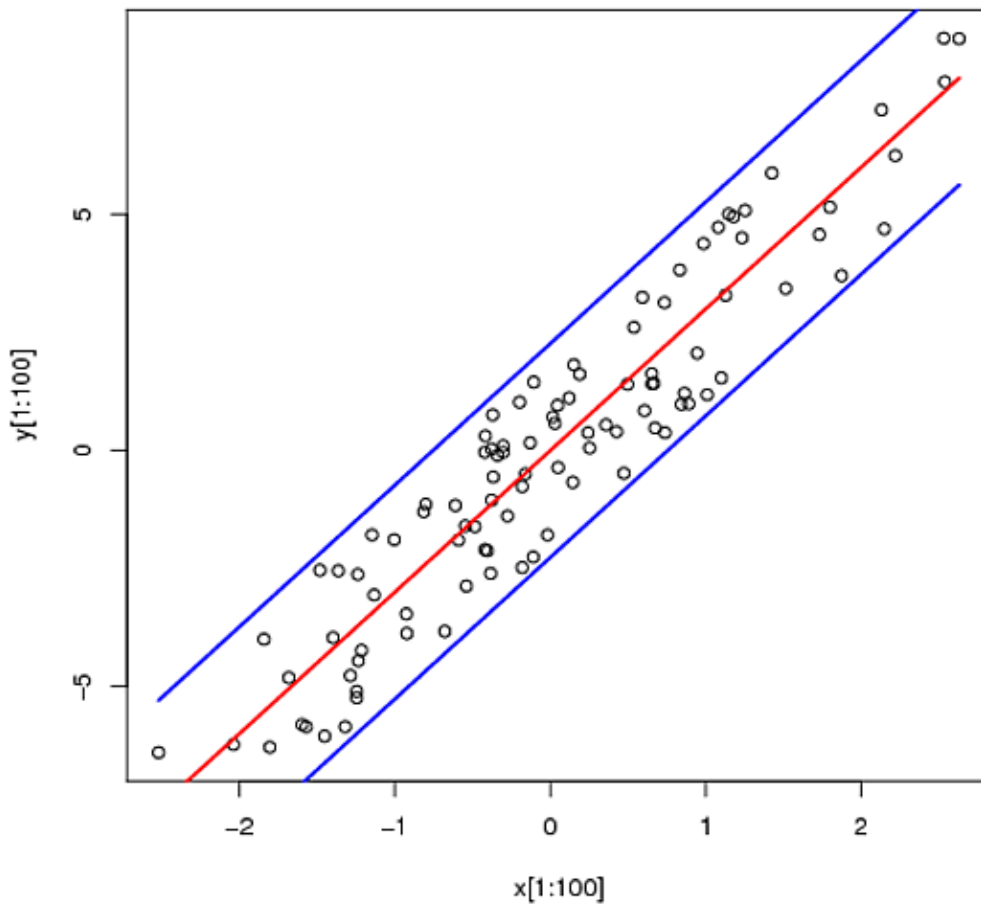


Figure B-2. Fitted Straight Line, Observations, and 95% Prediction Bounds

Based on information available in Chapter 8 of the PRIA, we were able to use the CAFE model as a “black box” to make runs that determine the upper and lower lines corresponding to those in Figure B-2. NHTSA estimated a linear regression like the ones depicted in the figures above. However, in this case the fitted line was for $y = \ln(s/(1-s))$, where s is the scrap rate (see page 1040 of the PRIA), and \ln is the natural logarithm. The scrap rates themselves are obtained by transforming this expression for y using equation 8-4 in the PRIA. (This is all done automatically inside the CAFE model).

We obtained the root-mean-square error (RMSE) for the regression for Cars from Table 8-10 in the PRIA (0.15). In order to produce the required results, we shifted the intercept in Table 8-10 and re-ran the CAFE model. This required two runs: One for the lower bound, and one for the upper bound. The reported intercept for the Cars model is in the “Scrapage Model Values” worksheet contained in the input Excel worksheet containing “parameters”. The reported value is -0.985368. The RMSE is 0.15. To get the intercepts for the two runs, the intercept was shifted by plus-or-minus $\text{RMSE} \times 1.96$ (1.96 is the value that produces the 95% interval), and entered into the parameter worksheet for two runs.