Stock Modeling for Locomotives and Marine Vessels

[formal cover page to be added]

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 2004	Final Report September 2003 – September 2004

4. TITLE AND SUBTITLE
Stock Modeling for Railroad Locomotives and Marine Vessels

5. FUNDING NUMBERS
RS91/AM119

6. AUTHOR(S)
Kevin Green,* David Chien,** and Judy Yahoodik*

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
*U.S. Department of Transportation
John A. Volpe National Transportation Systems Center
55 Broadway
Cambridge, MA 02142

8. PERFORMING ORGANIZATION REPORT NUMBER
DOT-VNTSC-RSPA-04-10

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Department of Transportation
Research and Special Programs Administration
400 7th Street, S.W.
Washington, D.C. 20590

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
**Bureau of Transportation Statistics

12a. DISTRIBUTION/AVAILABILITY STATEMENT
This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Stock modeling is the process of estimating the number of pieces of equipment in service in a given year manufactured in each of all relevant prior years. This type of modeling is important for, among other things, estimating the rate at which new technologies might penetrate the in-service fleet and thereby achieve results such as reductions in national emissions and energy consumption. Such estimates of future results are often the basis for policy decisions regarding technology-related incentives or requirements, as well as broader policy decisions regarding energy and greenhouse gas emissions. This paper furthers the development of stock models for railroad locomotives and marine vessels by reviewing basic theoretical concepts and their application to light-duty vehicles, and attempts to use available data to develop tentative locomotive and marine vessel survival curves.

14. SUBJECT TERMS
Emissions, energy consumption, locomotives, marine vessels, scrappage, stock models, survival curves

15. NUMBER OF PAGES
32

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

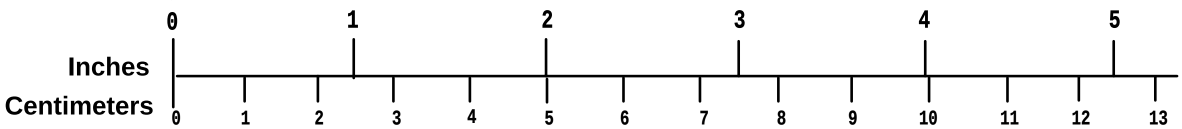
NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANS Std. 239-18
298-102

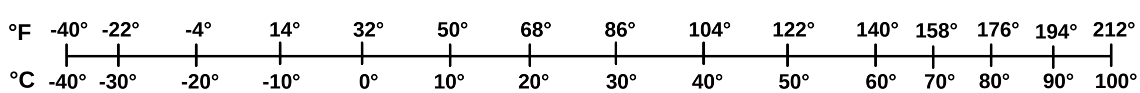
METRIC/ENGLISH CONVERSION FACTORS

ENGLISH TO METRIC		METRIC TO ENGLISH	
LENGTH (APPROXIMATE)		**LENGTH (APPROXIMATE)**	
1 inch (in)	= 2.5 centimeters (cm)	1 millimeter (mm)	= 0.04 inch (in)
1 foot (ft)	= 30 centimeters (cm)	1 centimeter (cm)	= 0.4 inch (in)
1 yard (yd)	= 0.9 meter (m)	1 meter (m)	= 3.3 feet (ft)
1 mile (mi)	= 1.6 kilometers (km)	1 meter (m)	= 1.1 yards (yd)
		1 kilometer (km)	= 0.6 mile (mi)
AREA (APPROXIMATE)		**AREA (APPROXIMATE)**	
1 square inch (sq in, in^2)	= 6.5 square centimeters (cm^2)	1 square centimeter (cm^2)	= 0.16 square inch (sq in, in^2)
1 square foot (sq ft, ft^2)	= 0.09 square meter (m^2)	1 square meter (m^2)	= 1.2 square yards (sq yd, yd^2)
1 square yard (sq yd, yd^2)	= 0.8 square meter (m^2)	1 square kilometer (km^2)	= 0.4 square mile (sq mi, mi^2)
1 square mile (sq mi, mi^2)	= 2.6 square kilometers (km^2)	10,000 square meters (m^2)	= 1 hectare (ha) = 2.5 acres
1 acre = 0.4 hectare (he)	= 4,000 square meters (m^2)		
MASS - WEIGHT (APPROXIMATE)		**MASS - WEIGHT (APPROXIMATE)**	
1 ounce (oz)	= 28 grams (gm)	1 gram (gm)	= 0.036 ounce (oz)
1 pound (lb)	= 0.45 kilogram (kg)	1 kilogram (kg)	= 2.2 pounds (lb)
1 short ton = 2,000 pounds (lb)	= 0.9 tonne (t)	1 tonne (t)	= 1,000 kilograms (kg)
			= 1.1 short tons
VOLUME (APPROXIMATE)		**VOLUME (APPROXIMATE)**	
1 teaspoon (tsp)	= 5 milliliters (ml)	1 milliliter (ml)	= 0.03 fluid ounce (fl oz)
1 tablespoon (tbsp)	= 15 milliliters (ml)	1 liter (l)	= 2.1 pints (pt)
1 fluid ounce (fl oz)	= 30 milliliters (ml)	1 liter (l)	= 1.06 quarts (qt)
1 cup (c)	= 0.24 liter (l)	1 liter (l)	= 0.26 gallon (gal)
1 pint (pt)	= 0.47 liter (l)		
1 quart (qt)	= 0.96 liter (l)		
1 gallon (gal)	= 3.8 liters (l)		
1 cubic foot (cu ft, ft^3)	= 0.03 cubic meter (m^3)	1 cubic meter (m^3)	= 36 cubic feet (cu ft, ft^3)
1 cubic yard (cu yd, yd^3)	= 0.76 cubic meter (m^3)	1 cubic meter (m^3)	= 1.3 cubic yards (cu yd, yd^3)
TEMPERATURE (EXACT)		**TEMPERATURE (EXACT)**	
[(x-32)(5/9)] °F	= y °C	[(9/5) y + 32] °C	= x °F

QUICK INCH - CENTIMETER LENGTH CONVERSION



QUICK FAHRENHEIT - CELSIUS TEMPERATURE CONVERSION



For more exact and or other conversion factors, see NIST Miscellaneous Publication 286, Units of Weights and Measures. Price $2.50 SD Catalog No. C13 10286

Updated 6/17/98

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AAR Association of American Railroads

EIA Energy Information Administration
EPA Environmental Protection Agency

mpg miles per gallon
NEMS National Energy Modeling System

RSD Regulatory Support Document

TRAN [NEMS] Transportation Model

Executive Summary

Policy development related to climate change often involves the consideration of the potential to improve the efficiency and greenhouse gas emission rates of new equipment, in particular by introducing new technologies. On a unit basis, comparison of alternatives is often straightforward. For example, a hybrid electric car achieving 0.02 gallons per mile (i.e., 50 miles per gallon) might be compared to an otherwise similar conventional car achieving 0.033 gallons per mile, indicating a 40 percent reduction in energy consumption and emissions. However, by themselves, unit comparisons do not provide a meaningful basis for judging the potential overall benefits—typically measured in annually avoided billions of gallons of petroleum product imports and millions of metric tons of avoided carbon dioxide emissions—of more efficient equipment.

Effectively estimating the impact new equipment will have on future energy and emissions generally begins by attempting to answer the following question: how quickly will more efficient new equipment replace less efficient older equipment? Answering this question requires the application of stock models—computational methods for representing equipment turnover. The equipment survival functions critical to such models require historical data regarding the sale of new equipment as a function of time and the number of units remaining in service as a function of both time and vintage.

The scope and quality of data for light duty vehicles has supported the development of comparatively robust stock models. For railroad locomotives and marine vessels—the focus of this analysis—data are much more limited, such that the survival curves derived here are more tentative. Two functional forms are fitted to available data in this analysis, yielding qualitatively similar survival curves that begin at 100 percent (i.e., all new units are initially in service) and gradually decline toward 0 percent (i.e., all units are eventually taken out of service). For locomotives, this approach suggests a typical useful life (in other words, the vintage at which a locomotive is as likely to be scrapped as to remain in service) of 26 to 28 years. Similar analysis of marine vessel data suggests a typical useful life of 24 to 31 years. A preliminary analysis of data regarding one specific category of marine vessels, ferries, suggests a typical useful life of 39 years.

These curves are consistent with anecdotal evidence that locomotives and marine vessels remain in service considerably longer than do automobiles, for which a typical useful life of 15 years has recently been estimated by the Energy Information Administration. These survival curves probably perform well enough to support the initial development of corresponding stock models for use in broad-based energy and emissions models (i.e., national-scale models used for mid- to long-range forecasting). Future efforts could build on this analysis—compiling data regarding older locomotives, examining other sources of marine vessel data, possibly focusing on economically meaningful specific categories of locomotives and marine vessels, and by identifying data regarding trends in the annual utilization of those locomotives and marine vessels that do remain in service.

1. Background

Stock modeling is the process of estimating the number of pieces of equipment in service in a given year manufactured in each of all relevant prior years. This type of modeling is important for, among other things, estimating the rate at which new technologies might penetrate the in-service fleet and thereby achieve results such as reductions in national emissions and energy consumption. Such estimates of future results are often the basis for policy decisions regarding technology-related incentives or requirements, as well as broader policy decisions regarding energy and greenhouse gas emissions.

A wealth of underlying data has enabled relatively well-established stock modeling practices for light-duty highway vehicles—practices that are used in current energy and emissions modeling systems. Information regarding other types of transportation equipment is more limited, and corresponding stock models are underdeveloped and, in some cases, nonexistent.

2. Purpose and Scope

The purpose of this paper is to further the development of stock models for railroad locomotives and marine vessels by reviewing basic theoretical concepts and their application to light-duty vehicles, and then attempting to use available data to develop tentative locomotive and marine vessel survival curves.

3. Theory

Within the context of energy and emissions forecasting, stock modeling typically involves estimating the number of units (i.e., vehicles) sold in each of the relevant past years that remain in service during the year in question.[1] If the importation of used equipment is negligible, the domestic in-service population of a given type of transportation equipment is expressed as the sum of the products of the numbers originally placed in service and the equipment survival rates:

$$N(y) = \sum_{MY<y} N_{MY}(y) = \sum_{MY<y} SALES_{MY} SURVIVAL(y - MY) \tag{1.1}$$

Here $N(y)$ is the total number surviving in year y, $SALES_{MY}$ is the number that were placed in service in year MY (i.e., model year), $N_{MY}(y)$ is the number from that MY that are still in service year y, and $SURVIVAL(v)$ is the fraction of equipment of vintage v that remain in service.

If this approach is used to model stock turnover, the prevalence and effects of a technology introduced on a known schedule can be determined through straightforward accounting. For example, considering a hypothetical scenario in which historical and forecasted sales volumes are available for vehicles with and without a new technology—referred to here as "Tech 1"—and survival rates are known, the population of all vehicles and those with the new technology can

[1] Sales estimates are also required, and typically involve a combination of historical information and sales forecasts. Normally, sales forecasts are developed exogenously (e.g., as an element of economic forecasting), although a theoretically more robust (though more data-intensive) approach might be to explicitly represent economic influences on both sales levels and survival rates.

easily be forecasted, as can the average fuel economy of new and in-service vehicles. The results of this type of simulation are illustrated in Figure 1 and Figure 2. This example uses the hypothetical survival curve discussed below and shown in Figure 3, and assumes that vehicles with and without “Tech 1” achieve 50 and 27.5 miles per gallon, respectively.

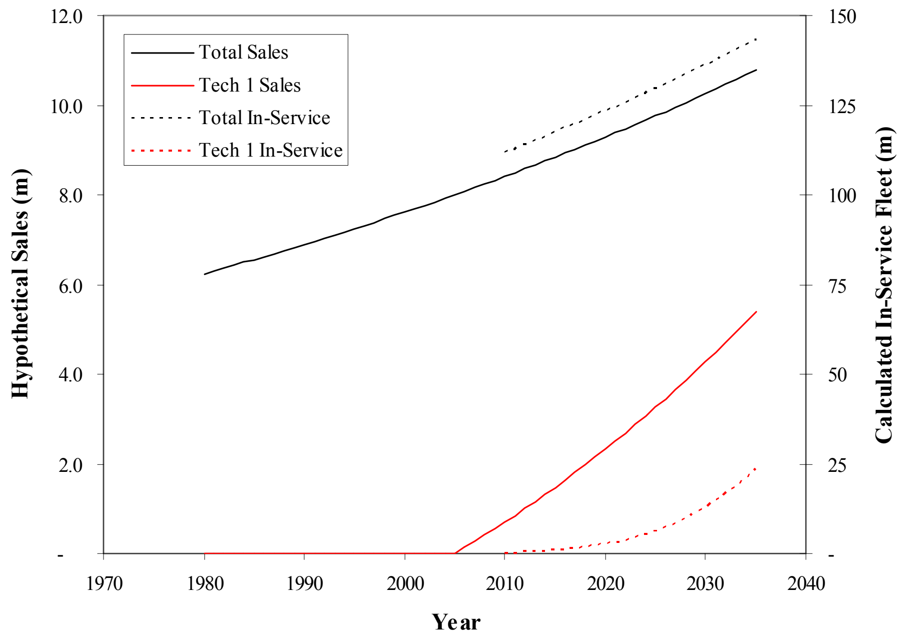


Figure 1. Hypothetical Technology Penetration Rate Simulation

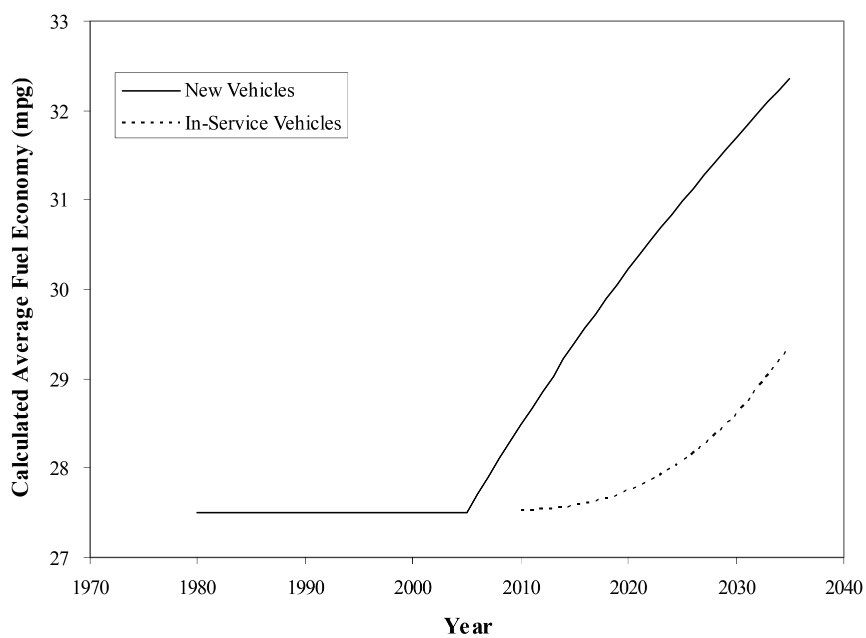


Figure 2. Hypothetical Fuel Economy Calculation

While this simple example uses population weighting to calculate average in-service fuel economy, vintage-specific estimates of vehicle utilization (i.e., miles driven per year) would enable usage weighting, which would yield average fuel economy estimates more directly related to total energy consumption and carbon dioxide emissions. However, with or without vintage-specific utilization estimates, an estimated survival function is required. Without a survival function, there is no way to clearly relate the sales and characteristics of new equipment to the population and characteristics of in-service equipment. Because a technology's ultimate effects—such as changes in total national energy consumption and carbon dioxide emissions—depend on its presence in the in-service fleet, estimation of the survival function is a key step in estimating such effects.

The historical survival function is estimated by solving (1.1). Doing so requires information regarding both the in-service population and historical sales volumes. If both sales volumes and in-service populations are known for enough model years, the survival function can be determined directly by solving the model year-specific terms of (1.1).[2] For example, working backwards from the 2002 model year, the survival function is estimated by solving the following system of equations:

$$\begin{Bmatrix} N_{2002}(2002) = SALES_{2002}SURVIVAL(0) \\ N_{2001}(2002) = SALES_{2001}SURVIVAL(1) \\ N_{2000}(2002) = SALES_{2000}SURVIVAL(2) \\ \vdots \end{Bmatrix} \tag{1.2}$$

A variety of functional forms can used to represent equipment survival rates. Prominent examples include the "Iowa Curves," developed in the early 20th century at Iowa State University and the "h-System," developed at the New York State Department of Public Service in 1947. As documented in a 1996 publication by the National Association of Regulatory Utility Commissioners, such survival curves (or equivalently, retirement frequency curves) provide a foundation for depreciation analysis used in the public utility sector.[3] Many of these curves exhibit a sigmoid shape (i.e., that of a stretched and reversed "S") shared by logistic curves, with the following form:[4]

$$SURVIVAL(v) = \frac{1+e^{v_{50}/k}}{e^{v_{50}/k}} \frac{e^{(v_{50}-v)/k}}{1+e^{(v_{50}-v)/k}} \tag{1.3}$$

---

[2] If only the overall population *N(y)* is known, it may still be possible to fit a survival rate function if model year-specific sales volumes are known. However, this would, at a minimum, require an a priori assumption regarding the functional form of the survival function. Of course, even if *N(y)* and $SALES_{MY}$ are known, substitution of (1.3) into (1.1) leaves one equation with two unknowns (*k* and v_{50}). However, if enough historical sales information is available and *N(y)* is known for more than one calendar year, it should be possible either to develop an additional equation, or even to statistically fit these constants (e.g., using nonlinear least-square techniques).

[3] National Association of Regulatory Utility Commissioners. "Public Utility Depreciation Practices." Washington, D.C. 1996.

[4] The underlying form omits the first ratio, a normalization constant applied to ensure that the survival rate is 100 percent when the vintage is equal to zero.

In equation (1.3), v_{50} and k are constants. Figure 3 shows a typical example of this functional form. The first constant, v_{50}, determines the vintage at which the survival rate is 50 percent (in this example, 15 years).[5] The second constant, k, determines the rate at which the function transitions from high to low survival rates.

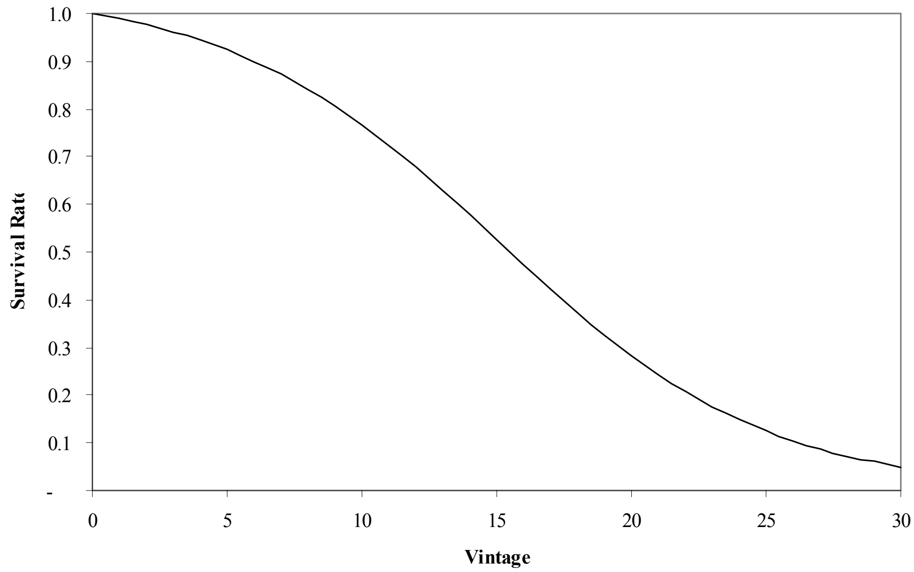


Figure 3. Sample Survival Function

The absence of national-scale data regarding either in-service populations or historical sales volumes could pose a significant challenge to this type of analysis. If companies are relatively uniform in their equipment procurement and use practices, it might be possible to collect and analyze the above-mentioned information from a limited number of representative companies. However, this could introduce problems associated with equipment transfers between companies. For example, if both $N_{MY}(y)$ and $SALES_{MY}$ are known for one particular company, but some used equipment was sold or leased to other domestic companies rather than being scrapped, biases in the derivation of a survival function could be introduced unless the analysis is extended to cover the other companies involved in such transactions. Such extension of coverage could quickly lead to national coverage at a company-by-company level of detail.

In addition, the possibility of one or more rebuilds prior to equipment scrappage introduces uncertainty regarding the definition of "new" equipment. However, as long as rebuilds are properly recorded, adjustments can be made to avoid double counting.

[5] Due to normalization in (1.3), this relationship is approximate. However, for k/v_{50} less than about 0.2, the survival rate falls to half at a vintage close to v_{50}.

4. Light-Duty Vehicles

Modeling of the domestic stock of automobiles and light trucks is highly evolved, and illustrates the application of the basic theory discussed above. This state of evolution is possible largely because the stock is relatively well bounded and there exists a comparative wealth of data. Because the level of importation of used vehicles from foreign countries is small, little error or bias is introduced by assuming that vehicles of a given vintage were represented in domestic sales data for the corresponding model year. Extensive historical data regarding the number of vehicles sold and registered domestically are available from a combination of public and commercial sources, and one major commercial source of vehicle registration data disaggregates this information by vintage.

Despite a few specific limitations, the scope and quality of information regarding light vehicles is sufficient to support fairly sophisticated analysis, a prominent example of which was released in 1996 by Greenspan and Cohen.[6] Central to this analysis was registration data available from R. L. Polk & Company. This data provides the equivalent of both sales and survival rates as defined in (1.1). Because it does so for multiple years of initial vehicle registration, it also provides a basis for inference regarding changes in vehicle durability, and for differentiating between what the authors refer to as "engineering" and "cyclical" scrappage.[7]

For all model years considered by Greenspan and Cohen, the Polk data exhibit a sigmoid shape similar to that shown above in Figure 3. Greenspan and Cohen fit the following basic form, in which k_1 and k_2 are constants, and v and SURVIVAL(v) are as in (1.3):

$$SURVIVAL(v) \propto e^{k_1 + k_2 v^2} \tag{1.4}$$

An earlier analysis by Miaou also utilizes the Polk data, but applies a form that integrates implicit engineering factors with explicit socioeconomic factors:[8]

$$SURVIVAL(v) \propto 1 - \frac{1}{a + e^{b + (c + k \bullet x)v}} \tag{1.5}$$

Here, a, b, and c are constants, $\mathbf{x}$ is a vector of socioeconomic variables, and $\mathbf{k}$ is a vector of corresponding coefficients.

As basic functional forms, (1.3), (1.4), and (1.5) perform in a qualitatively similar manner with respect to vintage. Data limitations for locomotives and marine vessels clearly limit the

---

[6] Greenspan, Alan and Cohen, Darrel. "Motor Vehicle Stocks, Scrappage, and Sales." Federal Reserve Board. Washington, DC. 1996. Available at http://www.federalreserve.gov/pubs/feds/1996/199640/199640pap.pdf.

[7] The ability to forecast vehicle sales volumes appears to provide the underlying motivation for this approach. Greenspan and Cohen apply "cyclical" scrappage on an aggregated rather than vintage-specific basis. While such aggregation may be appropriate for sales forecasting, stock modeling for purposes of, for example, energy forecasting, requires vintage-specific accounting.

[8] Miaou, S.-P. 1995. "Factors Associated with Aggregated Car Scrappage Rate in the United States: 1966-1992," Trans. Res. Rec., Vol.1475, 1995 pp.3-9. Report No. J95-75695.

feasibility of a multivariate form such as (1.5). Although this analysis focuses on the logistic form in (1.3), the exponential form in (1.4) would warrant consideration prior to actual implementation of stock models for locomotives and marine vessels. These two forms are sufficiently similar, though, that either can provide a basis for exploring general feasibility given the scope and quality of the data.

As an example of survival curves currently used in stock modeling, Figure 4 shows car and light truck survival rate estimates used by the Energy Information Administration (EIA) in preparing the 2004 Annual Energy Outlook (AEO 2004).[9]

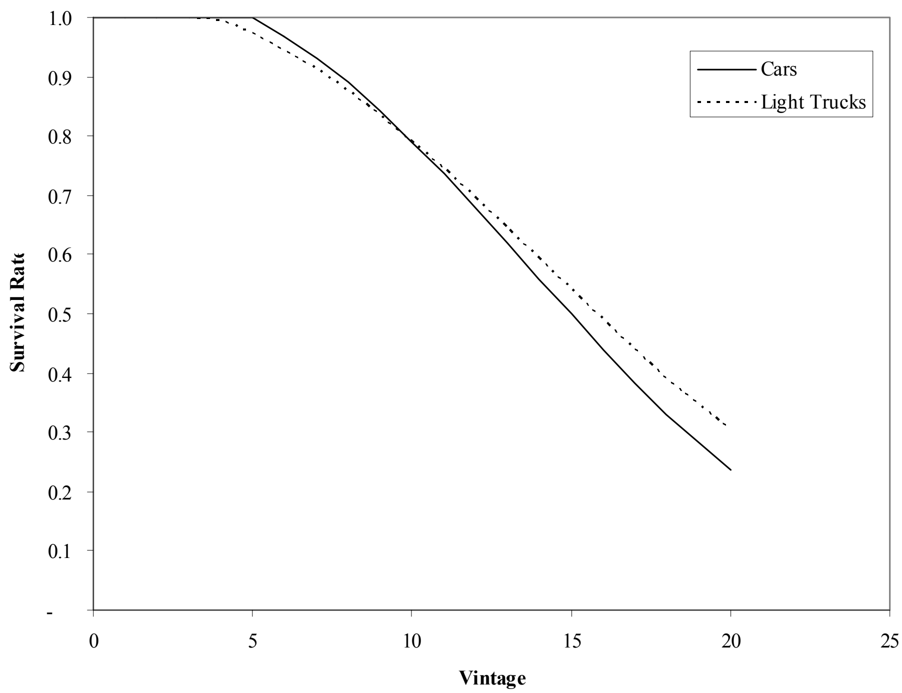


Figure 4. AEO 2004 Survival Rates for Cars and Light Trucks

These survival rate curves are defined as discrete vintage-specific input values rather than as explicit functional forms. However, both curves clearly exhibit at least the initially sigmoid shape illustrated by Figure 3. Although EIA's assumptions do not extend past vintages of 20 years, adding "flattened tails" to the curves in Figure 4 would clearly be much more reasonable than linearly extrapolating the sloped portions of the curves toward (much less below) an intersection with the vintage axis. Consequently, EIA's assumptions are consistent with sigmoid forms such as (1.3) and (1.4).

EIA's stock modeling is performed under the National Energy Modeling System (NEMS), which integrates the simulation of energy supply and energy demand with a broader macroeconomic simulation. Within NEMS, light vehicle stock modeling is performed by the Transportation

[9] Documentation is available from EIA at http://www.eia.doe.gov/bookshelf/docs.html.

Model (TRAN), which predicts transportation sector energy demand. By coupling sales and technology forecasting with the survival rates shown in Figure 4, TRAN produces forecasts of the prevalence of different technologies (e.g., advanced transmissions), as well as the resultant changes in energy consumption and greenhouse gas emissions.

At least three aspects of EIA's treatment warrant emphasis. First, although actual survival rates likely vary considerably for different groups of cars and light trucks, TRAN represents each general vehicle type in the aggregate rather than attempting to differentiate between brands or market segments based on expected quality. Second, although vehicle maintenance practices and service intensity likely vary widely between different groups of owners (e.g., police departments, taxi companies, households), TRAN effectively represents the average owner rather than attempting to account for such differences. Third, TRAN treats survival rates as fixed with respect to vehicle prices and macroeconomic measures like personal income, even though survival rates might reasonably be assumed to depend on such factors. TRAN does attempt to account for other dependencies on macroeconomic measures (for example, personal income as one determinant of travel demand). In all three of these areas, TRAN's design is consistent with the limits of its purpose. TRAN is intended to support broad-based national energy forecasting rather than, for example, vehicle acquisition planning by individuals or firms.

Although TRAN also has algorithms to forecast energy demand by locomotives and marine vessels, these algorithms use simple assumptions regarding modal energy intensity, measured in the amount of energy consumed per unit of underlying economic activity. Any technology-based efficiency improvements are purely implicit. Because TRAN does not contain stock models for locomotives and marine vessels, it cannot be used without modification to forecast the effects of new technologies for such equipment.

On the other hand, the factors underlying the turnover of locomotives and marine vessels differ from those underlying the turnover of cars and light trucks. For cars and light trucks, scrappage is driven primarily by maintenance costs and vintage is a reasonable proxy for these costs. It is probably also safe to assume that locomotive and marine vessel scrappage is at least partially driven by vintage-dependent maintenance costs. However, while most cars and light trucks are used for personal transportation, locomotives and non-recreational marine vessels are used for commercial purposes. Therefore, locomotive and marine vessel scrappage could also be influenced by changes in the basic nature of the relevant commercial activities, such as changes in freight flows or operational requirements. Scrappage related to such changes might have little to do with vintage, suggesting that the "engineering scrappage" approach considered here might well be enhanced by attempts to also represent these other factors.

5. Locomotives

For Class I Railroads, the Association of American Railroads (AAR) maintains a database of new locomotive purchases and leases as well as vintage-specific counts of in-service units. Unfortunately, this accounting uses model year cohorts (i.e., sets of locomotives grouped based on time of production) with temporal widths of 5 years.[10] Nonetheless, with two simplifying

[10] The most recent 5-year period is covered at a resolution of 1 year, and all locomotives older than 25 years are aggregated into a single cohort.

assumptions—that (1) each cohort is represented by its midpoint vintage in any given calendar year, and (2) survival rates have not changed significantly during the past 15-20 years—this data can be put in a form to which survival curves may reasonably be fitted. Figure 5 shows this data with cohorts indicated, along with fitted curves using forms (1.3) and (1.4). Appendix A lists estimated survival rates for each cohort and provides a summary of the statistical properties of the fitted curve.

To conduct this analysis, it was necessary to ignore locomotives older than 25 years, as there was no way to reasonably estimate the average age of such locomotives. Therefore, although the fitted function appears to perform relatively well for vintages of up to 15 years, whether it does so or not at high vintages (e.g., 40 years) is unclear.

This analysis could also be affected by the fact that locomotives are often leased and rebuilt during their service lives.[11,12] The AAR data appear structured in a manner that limits errors that might otherwise be associated with both leasing and rebuilds (i.e., leases and rebuilds are reported separately from purchases and builds of new units). However, because the AAR data only cover Class I railroads, the data do not account for the fact that some units removed from Class I service may still be used elsewhere, such as for smaller regional services. This suggests that the fitted curve shown in Figure 5 may understate survival rates at relatively high vintages.

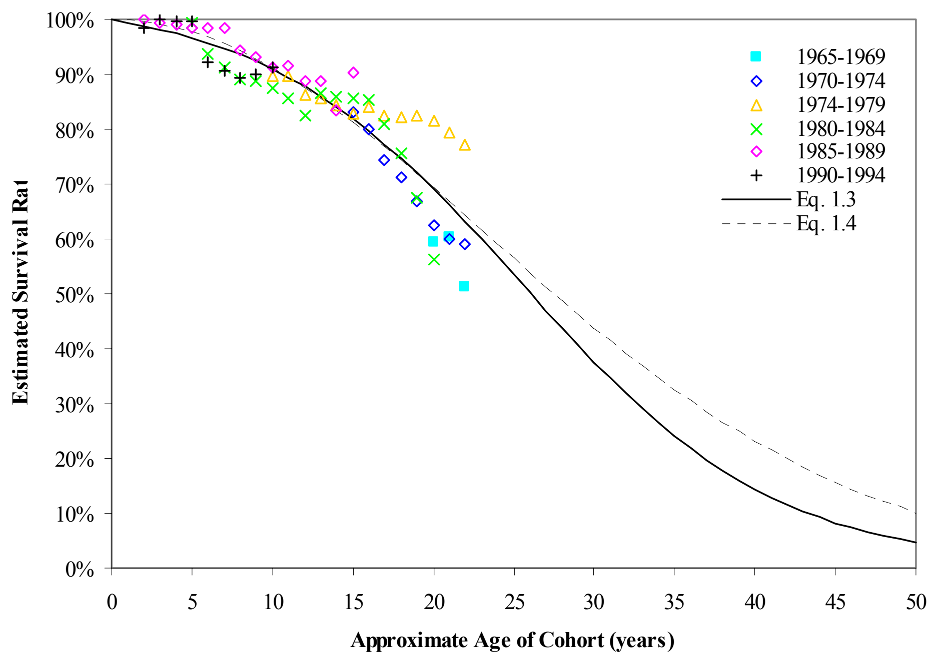


Figure 5. Locomotive Survival Rate vs. Vintage

[11] The rebuilding or upgrading of locomotives can significantly extend operation, and a locomotive may be rebuilt and upgraded three or four times. As locomotives age, the cost of overhauls and upgrades to maintain performance typically increases due to engineering limitations, and eventually forces investment in new equipment.

[12] For example, information available from Railspot's *All Time Roster Listing* (Available at http://www.railspot.com/gif/railspot/msql/ALLTIME.HTM) shows cases of individual locomotives having as many as five or six separate lessors before being removed from service.

At least one previous analysis, U.S. Environmental Protection Agency's (EPA's) Regulatory Support Document (RSD) for its 1998 locomotive emissions rulemaking, uses an estimated survival curve for locomotives.[13] Although EPA's analysis uses a slightly different formulation—the percent of expected useful life remaining as a function of vintage, the fitted curves shown in Figure 5 may be compared to that used by EPA by first calculating the expected useful life remaining implied by the former. As shown by Figure 6, the three curves predict virtually identical survival rates for locomotives less than 12 years old. For older locomotives, the curves diverge somewhat, EPA's estimates being more consistent with (1.4).

While either of the survival curves shown in Figure 5 could provide a reasonable basis for the initial development of a locomotive stock model used, for example, within the context of national-scale energy or emissions modeling, they would almost certainly be of limited relevance within a narrower context. The survival rate of any particular locomotive could depend on a wide range of factors, such as source (i.e., original manufacturer), engine type (2- or 4-cycle), service profile, and maintenance practices. When planning equipment purchases and retirements, railway managers might well have little use for a model that does not explicitly account for such factors. However, at the national scale and over relatively long time scales, variation in survival rates based on such factors should be "smoothed out," such that it would only be meaningful to attempt to explicitly account for these factors if their effects were well-understood and corresponding policies were under consideration.[14]

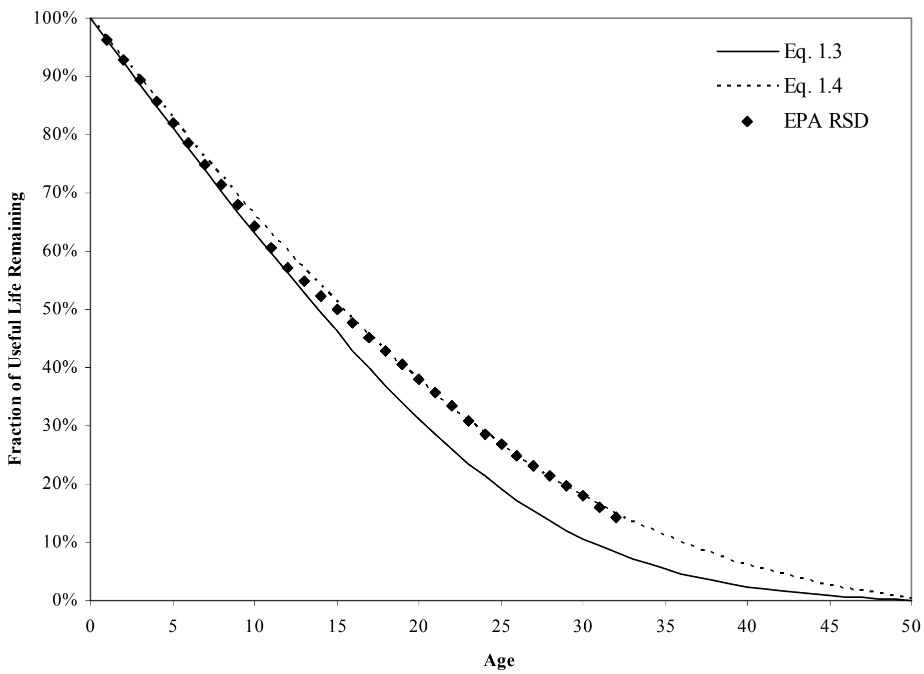


Figure 6. Comparison of Locomotive Survival Functions

[13] U.S. Environmental Protection Agency, Locomotive Emissions Standards—Regulatory Support Document, 1998.

[14] As mentioned above, despite a comparative wealth of data, automotive stock modeling does not explicitly account for factors such as brand name and maintenance practices. For example, even though it might be reasonable to attempt to account for relatively recent federal requirements regarding vehicle inspection and on-board diagnostics because those requirements might influence automobile survival rates, NEMS does not attempt to do so.

6. Marine Vessels

For marine vessels, consistent historical data regarding both sales (i.e., construction) and the population of in-service equipment are much more limited. The source closest to meeting the requirements discussed above is a database maintained by the U.S. Army Corps of Engineers.[15,16] This database provides profiles, including vintage, of the in-service fleet for several calendar years. However, because it provides very limited data regarding the introduction of new vessels into service, it does not provide a direct basis for estimating survival rates.[17]

To attempt to get around this limitation in the data, (1.1), (1.3), and (1.4) were manipulated and a major simplifying assumption was made in order to develop equations that did not include sales and could, therefore, be fitted. This is not a robust technique.

First, the share of the in-service fleet *SHARE (v)* was defined as a function of vintage and in terms of (1.1), after recasting vintage and (1.1) on a continuous basis (i.e., by making *v* a continuous variable):

$$SHARE(v) = \frac{SALES(v) \cdot SURVIVAL(v)}{\int_0^{\infty} SALES(v) \cdot SURVIVAL(v) dv} \quad (1.6)$$

To proceed without information regarding sales, *SALES(v)* was removed from (1.6) by assuming that *SALES(v)* has been constant. This is clearly not a sound assumption, but it was necessary in order to proceed. As a result, (1.6) reduces to the following:

$$SHARE(v) = \frac{SURVIVAL(v)}{\int_0^{\infty} SURVIVAL(v) dv} \quad (1.7)$$

---

[15] U.S. Army Corps of Engineers, *Waterborne Transportation Lines of the United States*, available at http://www.iwr.usace.army.mil/ndc/veslchar/veslchar.htm.

[16] Future efforts could consider cross-referencing the Army Corps of Engineers data with vessel documentation from the U.S. Coast Guard's *Merchant Vessels of the United States* series. The latter's historical coverage and detail would enable definitive estimation of the year in which vessels are removed from service, as well as the observation over many years of a sufficient number of older cohorts (e.g., vessels built in the 1960s) to provide a basis for directly (rather than by manipulation to remove sales) fitting forms such as (1.3) or (1.4). However, except for a few very recent years, data from this series appear to be available only in printed form, such that this type of analysis would require very extensive re-encoding of data.

[17] While the *Waterborne Transportation Lines of the United States* series does report (in Table 3) vessel construction and rebuilding, this data extends back only as far as 1989. Assuming marine vessel survival rates are very high through vintages of about fifteen years, this data covers too limited a period to provide a meaningful indication of longer-term survival rates. In other words, if a form similar in shape to that illustrated in Figure 1 is assumed, new vessel construction data from *Waterborne Transportation Lines of the United States* covers only the leftmost near-flat region of the curve. While the data regarding older locomotives is also limited, as illustrated by Figure 5, it extends far enough into the more steeply sloped region to enable at least a tentative fit of the forms considered here.

If *SURVIVAL(v)* is of form (1.3), (1.7) yields

$$SHARE(v) = \frac{\left[\dfrac{e^{-(v-v_{50})/k}}{1-e^{-(v-v_{50})/k}}\right]}{\int_0^\infty \left[\dfrac{e^{-(v-v_{50})/k}}{1-e^{-(v-v_{50})/k}}\right] dv} = \frac{\left[\dfrac{e^{-(v-v_{50})/k}}{1-e^{-(v-v_{50})/k}}\right]}{k \ln\left[1+e^{v_{50}/k}\right]} \tag{1.8}$$

For (1.4), integration of (1.7) yields

$$SHARE(v) = \frac{e^{k_1+k_2v^2}}{\int_0^\infty e^{k_1+k_2v^2}\, dv} = \frac{e^{k_2v^2}}{\int_0^\infty e^{k_2v^2}\, dv} = \frac{e^{k_2v^2}}{\dfrac{1}{2}\sqrt{\dfrac{\pi}{-k_2}}} = 2\sqrt{\frac{-k_2}{\pi}}e^{k_2v^2} \tag{1.9}$$

A nonlinear least-square algorithm was used to fit (1.8) and (1.9) to data regarding *SHARE(v)* for both 1997 and 2002.

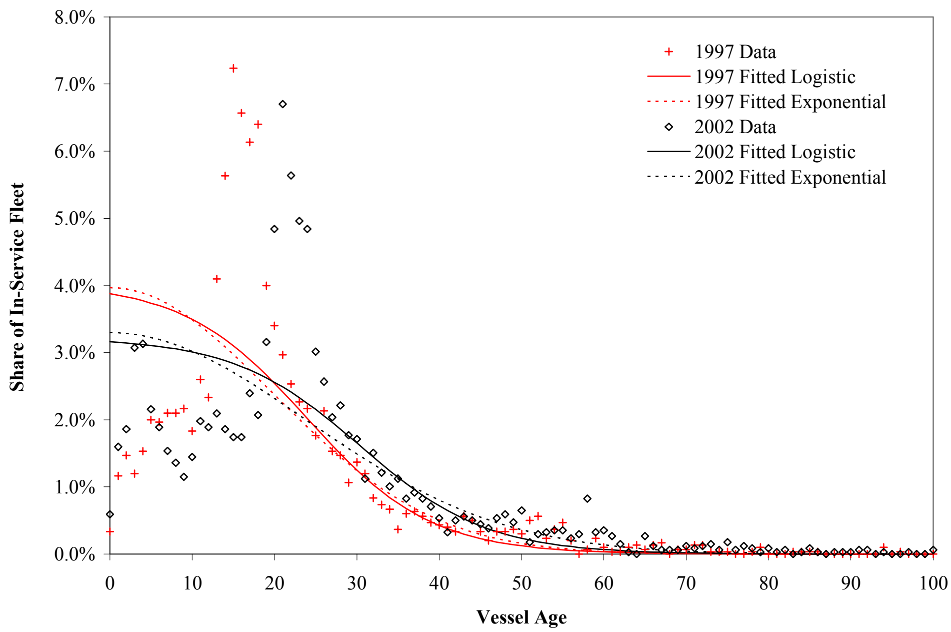


Figure 7. Age Distribution of Vessels in Service in 1997 and 2002

As indicated by Figure 7, (1.7) has a shape that does not fit the data available for ships built after the late 1970s. However, it appears safe to assume not that this indicates a fundamental problem with the assumption that the survival curve has a sigmoid shape like that of (1.3), but rather that new ships were being introduced at a historically unusual rate during the late 1970s and early

1980s, and that the constant-sales assumption is not at all representative of this period.[18] Therefore, the curves shown in Figure 7 were fitted only to the data for vessels built prior to the indicated peak in construction. Appendix B shows the summarized data regarding the 2002 and 1997 fleet of marine vessels and summarizes the statistical properties of the corresponding fitted curves.

Figure 7 includes all recorded marine vessels in a single data set. Given the diversity of marine vessel types, it might be appropriate to fit type-specific survival functions. Although the accompanying data manipulation would be a nontrivial undertaking, the analytical techniques should be at least as applicable to specific vessel types as to all marine vessels as a single group. For example, focusing on the 2002 ferry fleet and fitting equation (1.9) to the entire set of vintages (because of the lack of an obviously distorting peak) yields the curve shown in Figure 8. The statistical properties of this function are also summarized in Appendix B.

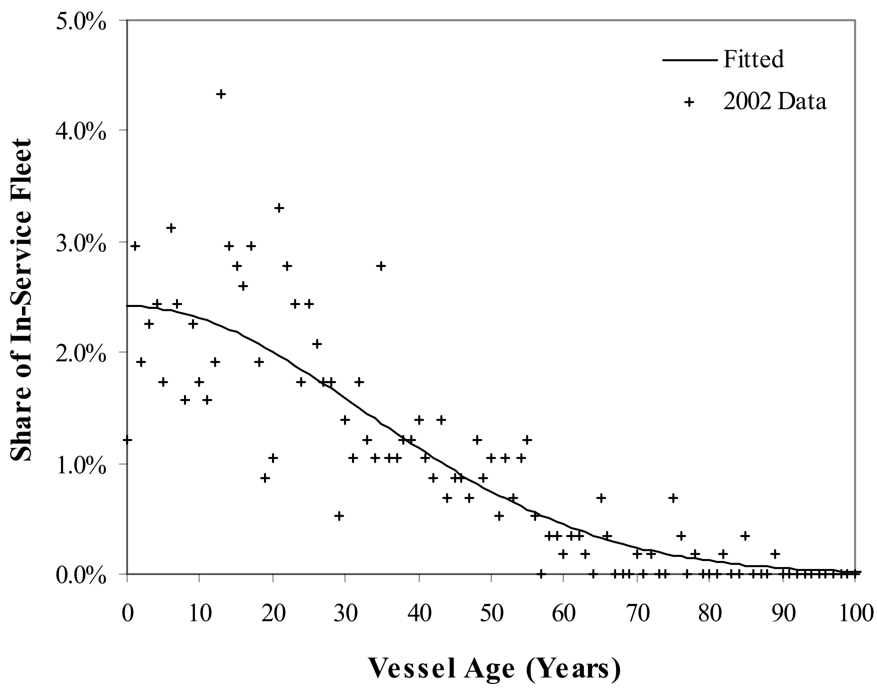


Figure 8. Age Distribution of Ferries in Service in 2002

Figure 9 shows the survival curves corresponding to the fitted share-of-fleet curves shown above. The significantly more gradual curve fitted to the ferry data appears to support the possibility of type-specific marine vessel stock modeling. In general, the necessity of modal disaggregation

[18] Data available from the Bureau of Transportation Statistics data at http://www.bts.gov/products/maritime_trade_and_transportation/2002/excel/figure_02_01.xls indicates that domestic orders for new marine vessels fell dramatically between 1975 and 1986, recovering slightly after 1994. Notwithstanding domestic orders for new foreign-built vessels (and the converse), this appears to support the assumption that the "young" side of the peaks in Figure 7 is dominated by sales trends—not survival rates. Data regarding trends in pre-1975 orders would help to understand the extent to which this assumption may reasonably be applied to the "old" side of the same peaks.

will depend on the level of disaggregation at which activity can be estimated, policies might be considered, and technologies might be utilized. For light-duty highway vehicles, it is appropriate to distinguish between automobiles and light trucks because they are subject to significantly different fuel economy standards. For marine vessels, activity level estimation would probably be a more important near-term consideration than vessel-related policies. Insofar as different types of vessels could be quantitatively related to distinct economic activities, it could be appropriate to develop correspondingly specific models of both activity levels and vessel survival.

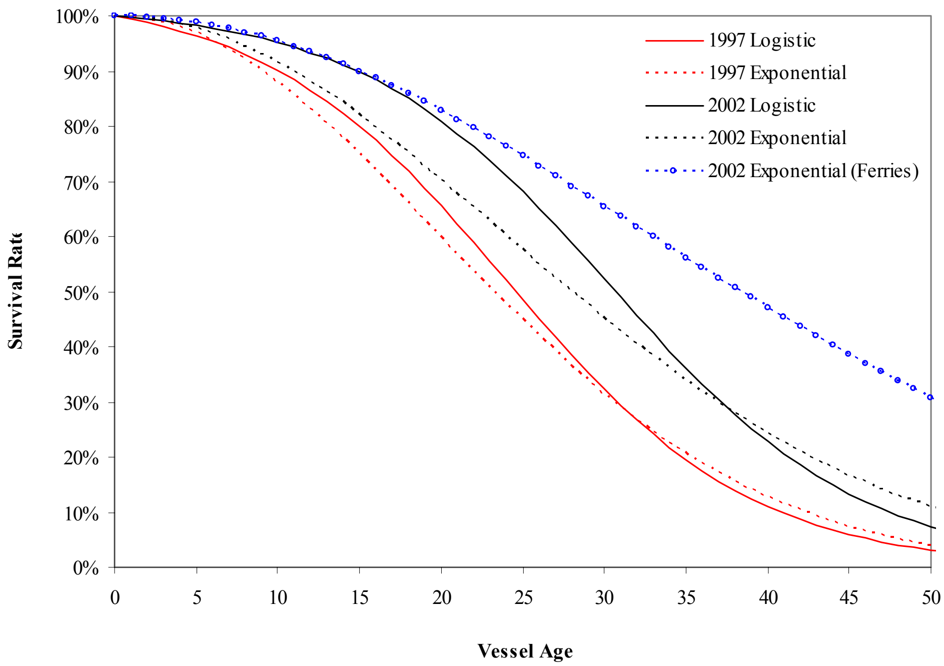


Figure 9. Fitted Survival Curves for Marine Vessels

7. Conclusions

The development of stock models that provide a means of forecasting the effects of new equipment depends most critically on the availability of data that can be used to estimate the probability of survival (i.e., the likelihood that a locomotive remains in service) as a function of vintage. Robust analysis would require historical data regarding units of sale as a function of time, as well as number of units remaining in service as a function of both time and vintage. A suggested format for such data is provided in Appendix C. For automobiles and light trucks, ample data are available in this structure. Data regarding locomotives and marine vessels appear to be considerably more limited.

A sigmoid (i.e., "S"-shaped) function can be fitted to available data for locomotives operated by Class I railroads. However, because of aggregated reporting for locomotives with vintages over 25 years, it is not clear how well this function represents the survival of older locomotives. Because the data do not include information regarding any transfer of older locomotives to other railroads for further service, the fitted survival function may understate survival probabilities at relatively high vintages. If possible, disaggregating data regarding older locomotives and accounting for use by non-Class 1 railroads would help to better understand locomotive survival.

Similar functions can also be fitted to data regarding in-service marine vessels. However, due to the scarcity of information regarding sales (or the equivalent), the analysis requires significant and possibly weak simplifying assumptions. Future studies could more closely examine other sources that might provide more sales data extensive enough to make such assumptions unnecessary. Also, because the one single-type function developed here (for ferries) corresponds to a significantly longer useful life than the functions developed for all marine vessels, future studies could focus on a variety of specific types of marine vessels. On the other hand, the eventual utility of type-specific marine vessel models would depend on the context of their use. The level of specificity would likely best be consistent with that of any accompanying models (e.g., economic models) used to forecast activity levels.

Despite these uncertainties and limitations, the survival functions developed here are consistent with the anecdotal indications that locomotives and marine vessels typically remain in service considerably longer than automobiles, as indicated by Figure 10.[19] The curves fitted to available data also suggest similar survival profiles for locomotives and marine vessels when the latter are considered in the aggregate.

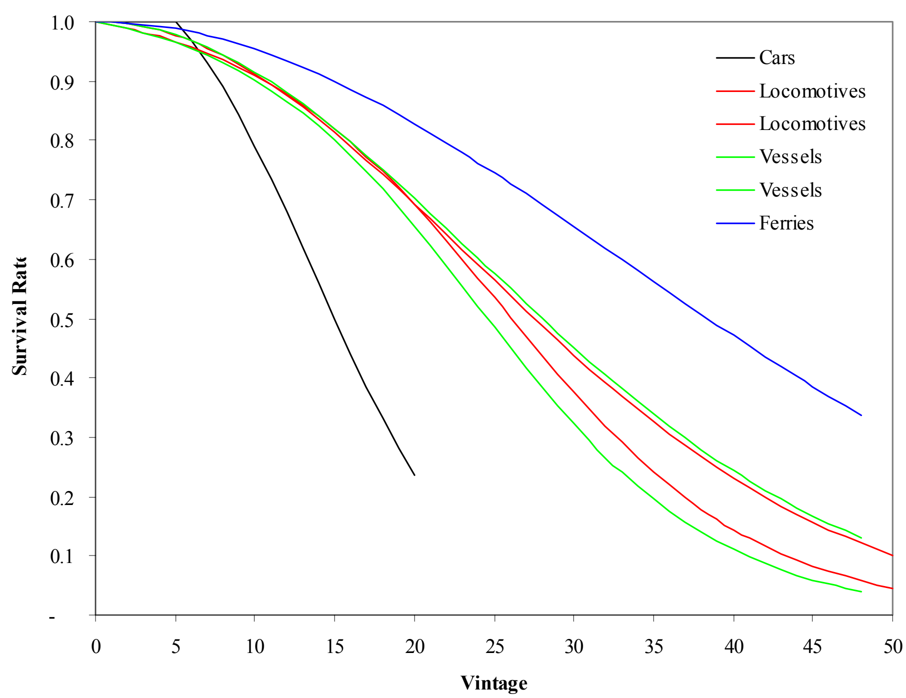


Figure 10. Intermodal Comparison of Survival Curves

[19] For marine vessels, only the two bounding curves shown in Figure 9 are repeated in Figure 10.

Given projections regarding the introduction of new equipment, using these functions to project the future composition of the in-use population of locomotives and marine vessels would be computationally straightforward. Within the context of broad-based energy and emissions forecasting, these survival functions may perform well enough to justify the initial development of stock models for this equipment, such that the potential effects of corresponding new technologies could begin to be quantified. However, like automobile survival forecasting tools, these models would likely be of limited use for very specific purposes, such as corporate fleet planning. Specific information regarding factors such as operating profile, maintenance practices, and choice of supplier could be very important within such contexts.

Estimates of the generally recognized decline in service intensity (e.g., annual hours of operation) over the useful life of most equipment would also be important when using stock models to forecast energy consumption and emissions. However, for locomotives and marine vessels, no corresponding data support the development of such estimates.

Appendix A. Locomotive Data and Fitted Curves

Table A-1. Estimated Locomotive Survival Rates

	Estimated Survival Rates						
Age	1965-1969	1970-1974	1974-1979	1980-1984	1985-1989	1990-1994	1995-1999
0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
1	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
2	#N/A	#N/A	#N/A	#N/A	100%	98%	88%
3	#N/A	#N/A	#N/A	#N/A	99%	100%	88%
4	#N/A	#N/A	#N/A	#N/A	99%	100%	88%
5	#N/A	#N/A	#N/A	99%	98%	100%	100%
6	#N/A	#N/A	#N/A	94%	98%	92%	
7	#N/A	#N/A	#N/A	91%	98%	91%	
8	#N/A	#N/A	#N/A	89%	94%	89%	
9	#N/A	#N/A	#N/A	89%	93%	90%	
10	#N/A	#N/A	90%	87%	91%	91%	
11	#N/A	#N/A	90%	86%	91%		
12	#N/A	#N/A	86%	82%	89%		
13	#N/A	#N/A	86%	86%	89%		
14	#N/A	#N/A	84%	86%	84%		
15	#N/A	83%	83%	86%	90%		
16	#N/A	80%	84%	85%			
17	#N/A	74%	83%	81%			
18	#N/A	71%	82%	76%			
19	#N/A	67%	82%	68%			
20	59%	63%	81%	56%			
21	60%	60%	79%				
22	51%	59%	77%				

Table A-2. Statistical Properties of Logistic Curve Fitted to Locomotive Data

Dependent Variable: SURVIVAL
Method: Least Squares
Date: 05/10/04 Time: 13:41
Sample: 1 67
Included observations: 67
Convergence achieved after 7 iterations
SURVIVAL=(EXP((C(2)-AGE)/C(1))+EXP((2*C(2)-AGE)/C(1)))/(EXP(C(2)/C(1))+EXP((2*C(2)-AGE)/C(1)))

	Coefficient	Std. Error	t-Statistic	Prob.
C(2)	25.45011	0.812147	31.33682	0.0000
C(1)	7.971970	1.145194	6.961243	0.0000

R-squared	0.758747	Mean dependent var	0.847732
Adjusted R-squared	0.755036	S.D. dependent var	0.119135
S.E. of regression	0.058965	Akaike info criterion	-2.794365
Sum squared resid	0.225993	Schwarz criterion	-2.728553
Log likelihood	95.61122	Durbin-Watson stat	1.253580

Table A-3. Statistical Properties of Exponential Curve Fitted to Locomotive Data

Dependent Variable: SURVIVAL
Method: Least Squares
Date: 04/16/04 Time: 07:16
Sample: 1 67
Included observations: 67
Convergence achieved after 3 iterations
SURVIVAL=EXP(C(2)*AGE^2) NOTE: see footnote[20]

	Coefficient	Std. Error	t-Statistic	Prob.
C(2)	-0.000915	4.33E-05	-21.10854	0.0000

R-squared	0.743312	Mean dependent var	0.847732
Adjusted R-squared	0.743312	S.D. dependent var	0.119135
S.E. of regression	0.060359	Akaike info criterion	-2.762200
Sum squared resid	0.240452	Schwarz criterion	-2.729294
Log likelihood	93.53369	Durbin-Watson stat	1.270725

[20] Regression results using SURVIVAL = EXP(C(1)+C(2)*AGE^2) indicated that the null hypothesis for C(1) could not be rejected at the 10 percent significance level.

Appendix B. Marine Vessel Data and Fitted Curves

Table B-1. 1997 Marine Vessel Data

Age	Number	Share	Age	Number	Share	Age	Number	Share	Age	Number	Share
0	10	0.3%	25	53	1.8%	50	9	0.3%	75	1	0.0%
1	35	1.2%	26	64	2.1%	51	15	0.5%	76	0	0.0%
2	44	1.5%	27	46	1.5%	52	17	0.6%	77	0	0.0%
3	36	1.2%	28	44	1.5%	53	7	0.2%	78	1	0.0%
4	46	1.5%	29	32	1.1%	54	11	0.4%	79	3	0.1%
5	60	2.0%	30	41	1.4%	55	14	0.5%	80	0	0.0%
6	59	2.0%	31	36	1.2%	56	6	0.2%	81	0	0.0%
7	63	2.1%	32	25	0.8%	57	0	0.0%	82	0	0.0%
8	63	2.1%	33	22	0.7%	58	2	0.1%	83	1	0.0%
9	65	2.2%	34	20	0.7%	59	7	0.2%	84	1	0.0%
10	55	1.8%	35	11	0.4%	60	3	0.1%	85	1	0.0%
11	78	2.6%	36	18	0.6%	61	1	0.0%	86	1	0.0%
12	70	2.3%	37	19	0.6%	62	1	0.0%	87	0	0.0%
13	123	4.1%	38	17	0.6%	63	3	0.1%	88	0	0.0%
14	169	5.6%	39	14	0.5%	64	4	0.1%	89	0	0.0%
15	217	7.2%	40	13	0.4%	65	2	0.1%	90	0	0.0%
16	197	6.6%	41	12	0.4%	66	4	0.1%	91	0	0.0%
17	184	6.1%	42	10	0.3%	67	5	0.2%	92	0	0.0%
18	192	6.4%	43	17	0.6%	68	0	0.0%	93	0	0.0%
19	120	4.0%	44	15	0.5%	69	2	0.1%	94	3	0.1%
20	102	3.4%	45	10	0.3%	70	2	0.1%	95	0	0.0%
21	89	3.0%	46	6	0.2%	71	4	0.1%	96	1	0.0%
22	76	2.5%	47	10	0.3%	72	4	0.1%	97	0	0.0%
23	68	2.3%	48	10	0.3%	73	1	0.0%	98	0	0.0%
24	65	2.2%	49	11	0.4%	74	1	0.0%	99	0	0.0%

Table B-2. Statistical Properties of Logistic Curve Fitted to 1997 Marine Vessel Data

Dependent Variable: SHARE
Method: Least Squares
Date: 10/21/03 Time: 08:26
Sample: 1 126
Included observations: 126
Convergence achieved after 9 iterations
SHARE=(EXP(-(AGE-C(1))/C(2))/(1+EXP(-(AGE-C(1))/C(2)))/(C(2)
*LOG(1+EXP(C(1)/C(2)))))

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	30.40182	0.839124	36.23042	0.0000
C(2)	7.706826	0.499986	15.41409	0.0000

R-squared	0.917015	Mean dependent var	0.002721
Adjusted R-squared	0.916346	S.D. dependent var	0.005378
S.E. of regression	0.001555	Akaike info criterion	-10.07830
Sum squared resid	0.000300	Schwarz criterion	-10.03328
Log likelihood	636.9327	Durbin-Watson stat	0.778786

Table B-3. Statistical Properties of Exponential Curve Fitted to 1997 Marine Vessel Data

Dependent Variable: SHARE
Method: Least Squares
Date: 04/06/04 Time: 09:42
Sample: 1 130
Included observations: 130
Convergence achieved after 9 iterations
SHARE=2*((-C(1)/3.141592654)^0.5)*EXP(C(1)*AGE^2)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.001288	4.37E-05	-29.47491	0.0000

R-squared	0.939415	Mean dependent var	0.002595
Adjusted R-squared	0.939415	S.D. dependent var	0.005491
S.E. of regression	0.001352	Akaike info criterion	-10.36739
Sum squared resid	0.000236	Schwarz criterion	-10.34533
Log likelihood	674.8805	Durbin-Watson stat	0.739627

Table B-4. 2002 Marine Vessel Data

Age	Number	Share	Age	Number	Share	Age	Number	Share	Age	Number	Share
0	20	0.6%	25	102	3.0%	50	22	0.6%	75	6	0.2%
1	54	1.6%	26	87	2.6%	51	6	0.2%	76	2	0.1%
2	63	1.9%	27	69	2.0%	52	10	0.3%	77	4	0.1%
3	104	3.1%	28	75	2.2%	53	11	0.3%	78	3	0.1%
4	106	3.1%	29	60	1.8%	54	12	0.4%	79	1	0.0%
5	73	2.2%	30	58	1.7%	55	12	0.4%	80	3	0.1%
6	64	1.9%	31	38	1.1%	56	8	0.2%	81	1	0.0%
7	52	1.5%	32	51	1.5%	57	10	0.3%	82	2	0.1%
8	46	1.4%	33	41	1.2%	58	28	0.8%	83	0	0.0%
9	39	1.2%	34	34	1.0%	59	11	0.3%	84	1	0.0%
10	49	1.4%	35	38	1.1%	60	12	0.4%	85	3	0.1%
11	67	2.0%	36	28	0.8%	61	9	0.3%	86	1	0.0%
12	64	1.9%	37	31	0.9%	62	5	0.1%	87	0	0.0%
13	71	2.1%	38	28	0.8%	63	1	0.0%	88	1	0.0%
14	63	1.9%	39	24	0.7%	64	0	0.0%	89	1	0.0%
15	59	1.7%	40	18	0.5%	65	9	0.3%	90	1	0.0%
16	59	1.7%	41	11	0.3%	66	4	0.1%	91	2	0.1%
17	81	2.4%	42	17	0.5%	67	2	0.1%	92	2	0.1%
18	70	2.1%	43	19	0.6%	68	2	0.1%	93	0	0.0%
19	107	3.2%	44	17	0.5%	69	2	0.1%	94	1	0.0%
20	164	4.8%	45	15	0.4%	70	4	0.1%	95	0	0.0%
21	227	6.7%	46	13	0.4%	71	3	0.1%	96	0	0.0%
22	191	5.6%	47	18	0.5%	72	4	0.1%	97	1	0.0%
23	168	5.0%	48	20	0.6%	73	5	0.1%	98	0	0.0%
24	164	4.8%	49	16	0.5%	74	2	0.1%	99	0	0.0%

Table B-5. Statistical Properties of Logistic Curve Fitted to 2002 Marine Vessel Data

Dependent Variable: SHARE
Method: Least Squares
Date: 10/22/03 Time: 07:24
Sample: 1 130
Included observations: 130
Convergence achieved after 10 iterations
SHARE=(EXP(-(AGE-C(1))/C(2))/(1+EXP(-(AGE-C(1))/C(2)))/(C(2)
*LOG(1+EXP(C(1)/C(2)))))

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	23.98774	0.691485	34.69016	0.0000
C(2)	7.507098	0.370120	20.28288	0.0000

R-squared	0.948593	Mean dependent var	0.002595
Adjusted R-squared	0.948191	S.D. dependent var	0.005491
S.E. of regression	0.001250	Akaike info criterion	-10.51627
Sum squared resid	0.000200	Schwarz criterion	-10.47215
Log likelihood	685.5573	Durbin-Watson stat	0.859887

Table B-6. Statistical Properties of Exponential Curve Fitted to 2002 Marine Vessel Data

Dependent Variable: SHARE
Method: Least Squares
Date: 04/06/04 Time: 09:44
Sample: 1 126
Included observations: 126
Convergence achieved after 10 iterations
SHARE=2*((-C(1)/3.141592654)^0.5)*EXP(C(1)*AGE^2)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.000883	4.14E-05	-21.32021	0.0000

R-squared	0.900014	Mean dependent var	0.002721
Adjusted R-squared	0.900014	S.D. dependent var	0.005378
S.E. of regression	0.001701	Akaike info criterion	-9.907796
Sum squared resid	0.000361	Schwarz criterion	-9.885286
Log likelihood	625.1911	Durbin-Watson stat	0.656496

Table B-7. 2002 Marine Vessel Data (Ferries Only)

Age	Number	Share	Age	Number	Share	Age	Number	Share	Age	Number	Share
0	7	1.2%	25	14	2.4%	50	6	1.0%	75	4	0.7%
1	17	3.0%	26	12	2.1%	51	3	0.5%	76	2	0.3%
2	11	1.9%	27	10	1.7%	52	6	1.0%	77	0	0.0%
3	13	2.3%	28	10	1.7%	53	4	0.7%	78	1	0.2%
4	14	2.4%	29	3	0.5%	54	6	1.0%	79	0	0.0%
5	10	1.7%	30	8	1.4%	55	7	1.2%	80	0	0.0%
6	18	3.1%	31	6	1.0%	56	3	0.5%	81	0	0.0%
7	14	2.4%	32	10	1.7%	57	0	0.0%	82	1	0.2%
8	9	1.6%	33	7	1.2%	58	2	0.3%	83	0	0.0%
9	13	2.3%	34	6	1.0%	59	2	0.3%	84	0	0.0%
10	10	1.7%	35	16	2.8%	60	1	0.2%	85	2	0.3%
11	9	1.6%	36	6	1.0%	61	2	0.3%	86	0	0.0%
12	11	1.9%	37	6	1.0%	62	2	0.3%	87	0	0.0%
13	25	4.3%	38	7	1.2%	63	1	0.2%	88	0	0.0%
14	17	3.0%	39	7	1.2%	64	0	0.0%	89	1	0.2%
15	16	2.8%	40	8	1.4%	65	4	0.7%	90	0	0.0%
16	15	2.6%	41	6	1.0%	66	2	0.3%	91	0	0.0%
17	17	3.0%	42	5	0.9%	67	0	0.0%	92	0	0.0%
18	11	1.9%	43	8	1.4%	68	0	0.0%	93	0	0.0%
19	5	0.9%	44	4	0.7%	69	0	0.0%	94	0	0.0%
20	6	1.0%	45	5	0.9%	70	1	0.2%	95	0	0.0%
21	19	3.3%	46	5	0.9%	71	0	0.0%	96	0	0.0%
22	16	2.8%	47	4	0.7%	72	1	0.2%	97	0	0.0%
23	14	2.4%	48	7	1.2%	73	0	0.0%	98	0	0.0%
24	10	1.7%	49	5	0.9%	74	0	0.0%	99	0	0.0%

Table B-8. Statistical Properties of Exponential Curve Fitted to 2002 Ferry Data

Dependent Variable: SHARE
Method: Least Squares
Date: 04/14/04 Time: 11:42
Sample: 1 151
Included observations: 151
Convergence achieved after 10 iterations
SHARE=2*((-C(1)/3.141592654)^0.5)*EXP(C(1)*AGE^2)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.000471	3.22E-05	-14.62286	0.0000

R-squared	0.825987	Mean dependent var	0.006623
Adjusted R-squared	0.825987	S.D. dependent var	0.009446
S.E. of regression	0.003941	Akaike info criterion	-8.228405
Sum squared resid	0.002329	Schwarz criterion	-8.208423
Log likelihood	622.2446	Durbin-Watson stat	1.718737

Appendix C. Suggested Data Structure

The development of survival curves requires historical data regarding both the number of units initially built and the number of units still in service. Consistency regarding geographic coverage is required. For example, annual historical data are available regarding the number of domestically manufactured and imported cars and light trucks sold for domestic use, as are data regarding the number of units (by year of initial sale) remaining in domestic service.

If 2005 is the most recent year for which such information is available, such data would have the following structure:

Table C-1. Suggested Data Structure to Support Future Analyses

		Number Built	Number of Units Still in Domestic Service in Calendar Year									
			2005	2004	2003	2002	2001	2000	1999	1998	…	Y_{min}
Year Built	2005											
	2004											
	2003											
	2002											
	2001											
	2000											
	1999											
	1998											
	…											
	Y_{min}*											

Here, Y_{min} is the earliest year of coverage. If anecdotal evidence suggests a typical useful life (i.e., vintage beyond which failure is more expected than not) of τ, having data covering vintages through 1.5τ to 2τ would probably be ideal. In other words, Y_{min} would ideally be at most Y_{max} - 1.5 τ. In terms of Table C-1, this means coverage of equipment with an expected typical life of 25 years should extend from 2005 back through roughly 1955-1967.

As indicated by Table C-2, currently available data regarding the introduction of new locomotives extends from 1965 through 1999, and vintage-specific counts of in-service units cover roughly half the potential extent of such counts. As indicated by Figure 5, these data provide a good basis for estimating particularly the "early" portion (i.e., left half) of locomotive survival curves. More extensive data regarding older in-service locomotives would facilitate estimation of the "later" portion (i.e., right half) of such curves.

Table C-2. Extent of Currently Available Data Regarding Locomotives

		Num. Built	Number of Units Still in Domestic Service in Calendar Year							
			'95-'99	'90-'94	'85-'89	'80-'84	'75-'79	'70-74	'65-'69	'60-'64
Year Built	'95-'99	✓	✓							
	'90-'94	✓	✓	✓						
	'85-'89	✓	✓	✓	✓					
	'80-'84	✓	✓	✓	✓					
	'75-'79	✓	✓	✓	✓					
	'70-'74	✓		✓	✓					
	'65-'69	✓			✓					

Table C-3 demonstrates the relative paucity of data regarding the introduction and survival of marine vessels. Information regarding the placement into service of new vessels extends back only to 1987, and although annual “snapshots” of the in-service fleet appear to cover all vintages, they are only available between 1997 and 2002.[21] Because of these data limitations, survival curves for marine vessels can be estimated only approximately and with significant simplifying assumptions, as suggested by Figure 7.

Table C-3. Extent of Currently Available Data Regarding Marine Vessels

		Num.	Number of Units Still in Domestic Service in Calendar Year							
		Built	’97-‘02	’92-‘96	’87-‘91	’82-‘86	’77-‘81	’72-‘76	’67-‘71	’62-‘66
Year Built	’97-‘02	✓	✓							
	’92-‘96	✓	✓							
	’87-‘91	✓	✓							
	’82-‘86		✓							
	’77-‘81		✓							
	’72-‘76		✓							
	’67-‘71		✓							
	’62-‘66		✓							

[21] Although not shown in Table C-3, these profiles of the in-service fleet include very small numbers of very old (more than fifty year-old) vessels.