INSTITUTE FOR DEFENSE ANALYSES

Hedonic Price Indices for Ground Vehicles

David M. Tate

May 2015

IDA Document NS D-5467
H 15-000310

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication
This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-14-D-0001, Project Number BA-7-3054, "Cost Indices Assessment" for the Director, Cost Assessment and Program Evaluation. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

INSTITUTE FOR DEFENSE ANALYSES

Hedonic Price Indices for Ground Vehicles

David M. Tate
Institute for Defense Analyses (CARD)
4850 Mark Center Dr., Alexandria, VA 22311
Ph: 703-575-1409 (W)
Fax: 703-824-9103
E-mail: dtate@ida.org

May 2015

Contents

A. Introduction

As part of ongoing support to the Office of the Secretary of Defense, Cost Assessment and Program Evaluation (OSD(CAPE)), the Institute for Defense Analyses (IDA) has been developing hedonic price indices for various types of military systems. Hedonic indices attempt to correct overall price changes over time for changes in the quality of the good being purchased, in order to distinguish price growth (for a given product) from demand shift (to a product with different characteristics). This paper reports preliminary results of investigations into hedonic indices for military ground vehicles, including both tactical vehicles and combat vehicles.

B. Data

For this effort, we collected data on as many ground vehicle systems as possible. Only new builds (as opposed to upgrades or modifications of existing units) were considered. This restricted the number of vehicle types available to the analysis, particularly for the years of the "procurement holiday" in the mid-1990s. In addition, hedonic models require both price information and detailed technical specifications, which further limited the range of vehicles included. Jane's, manufacturer product sheets, the Federation of American Scientists website, Gary's Combat Vehicles, Forecast International, and other miscellaneous sources provided specification data.

The price data used were taken from a variety of sources at different levels of detail. Prior to 1996, only top-level Selected Acquisition Report (SAR) procurement costs (minus spares) and quantities, unadjusted for advance procurement, were available. Named block upgrades (e.g., M2A0 Bradley versus M2A1 Bradley) were treated as separate vehicles. The systems included in this period were as follows:

- M1 Abrams tank (2 blocks)
- M2 Bradley Fighting Vehicle (3 blocks)
- M9 Armored Combat Earthmover
- M998 HMMWV (2 blocks)
- M992 Field Artillery Ammunition and Support Vehicle (FAASV) (2 blocks)

Beginning in 1996, more detailed quantity and price information became available in the form of Army budget justification forms (Exhibit P-5) itemizing different variants and separating base vehicle costs from other expenditures. In general, the "price" of each vehicle was defined as any costs allocated by the Army to unit cost in the P-5. Other costs, such as Engineering Change Orders, documentation, quality assurance, government-furnished equipment, upgrade kits, government testing, and field support, were not included in the price. For the period 1996–2012, data were compiled for the following:

- M4 Command and Control Vehicle (C2V)
- Stryker family (8 variants)
- M992A2 FAASV
- Armored Security Vehicle (ASV)
- Family of Heavy Tactical Vehicles (10 variants, generally 2 blocks each)
- HMMWV (3 variants)

In addition to these data, CAPE-CA provided a 2011 spreadsheet documenting Army and Marine Corps Mine-Resistant Ambush-Protected (MRAP) vehicle purchases by manufacturer, vehicle category (i.e., Category I, Category II, or Category III), and contract date. The individual MRAP models reflected in those buys for eight MRAP types were identified and included in the data set. We augmented the MRAP data from CAPE-CA with contract award announcements in those cases where the announcement specified quantity, price, and the specific vehicle variant to be provided. The eight MRAP types included were as follows:

- Buffalo
- Cougar H
- Cougar HE
- MaxxPro
- MaxxPro Dash
- MaxxPro Dash DXM
- MATV
- MATV-UIK

C. Data Characteristics

In all, data was compiled on 319 purchases of 53 distinct vehicle types between 1981 and 2012. Each vehicle variant/block was assigned to one of the families in Table 1.

Table 1. Families of Purchased Vehicles

Family	Data Points
Light tactical vehicle (LTV, e.g., HMMWV)	31
Medium tactical vehicle (MTV, e.g., 2.5-ton truck)	12
Heavy tactical vehicle (HTV, e.g., HEMTT)	99
Force protection vehicle (FPV, e.g., MRAP or ASV)	56
Tracked combat vehicle (TCV, Abrams or Bradley)	22
Tracked support vehicle (TSV, e.g., M9 ACE)	22
Wheeled combat vehicle (WCV, e.g., Stryker Mobile Gun System)	36
Wheeled support vehicle (WSV, e.g. Stryker Medevac vehicle)	41
TOTAL	**319**

Note: HMMWV - High Mobility Multipurpose Wheeled Vehicle; HEMTT – Heavy Expanded Mobility Tactical Truck; ASV – Armored Security Vehicle; ACE – Armored Combat Earthmover.

Note the uneven distribution, with heavy tactical vehicles and force protection vehicles contributing nearly half of the data points. Some family assignments, such as those for the Stryker reconnaissance vehicles (which were classified as support vehicles), may be open to dispute.

The coverage of years was also uneven, as noted above. Figure 1 shows a histogram of the number of data points by year. The very few points in 1994 and 1995 make it impossible to directly estimate credible price indices for those years. In the end, they were left out, and that period was made the omitted base period to which other years were compared.

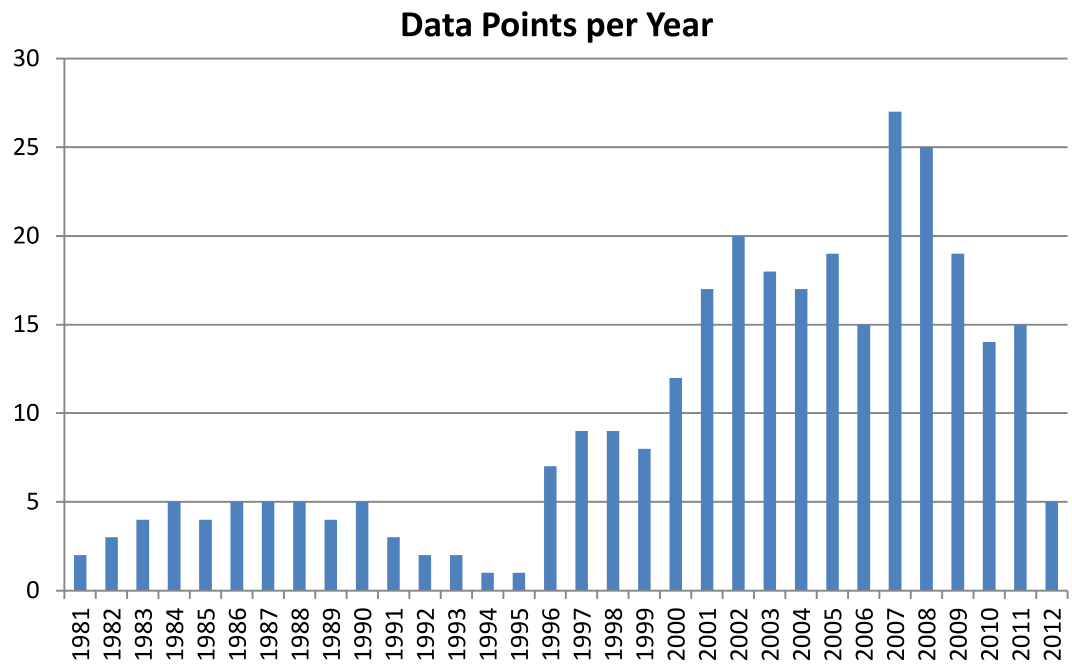

Figure 1. Time Distribution of Data

More subtly, there is also a time bias in the data that could make it difficult to estimate price indices. All of the expensive tracked combat vehicles appear prior to 1995, using fully-loaded SAR costs. All of the wheeled combat vehicles appear after 2000, using unloaded P-5 line item costs. MRAP purchases, for which the Services are rumored to have paid an expediting premium, occur only in the last few years of the time span. The correlation between the year and the family mix of vehicles, combined with the correlation between year and data source, could distort the underlying price growth.

D. Model Specification

1. Basic Hedonic Regression

The typical specification for deriving a hedonic index has three basic parts: a set of fixed effects for individual years (i.e., the price index to be estimated), a set of non-quality factors that affect the unit price of individual vehicle buys, and quality factors that constitute a cost-estimating relationship (CER) for vehicles. There is also the potential for "latent" quality changes that are not captured by the vehicle characteristic data available. Each of these is discussed in turn.

To estimate the overall price growth due to all factors, we fit a naïve regression using a fixed effect ("constant dollar price") for each vehicle type and a single common constant inflation rate. That model estimates the annual price growth at roughly 14 percent annually, which is unreasonably high if quality is truly constant over time for a given vehicle variant and block. In addition, this naïve model gives a very good fit, with an adjusted R^2 of 0.97 and few obvious outliers. This suggests the presence of latent quality growth within vehicle types; this is discussed further below.

a. Price Index

In principle, the price index variables would consist of one fixed-effect variable per year in the data set, minus one omitted reference year. The coefficients on these variables could then be interpreted (after undoing any data transformations used in the regression) as the relative price of vehicles in each year—a multiple of the price in the reference year.

In practice, the singleton data points in the years 1994 and 1995 make the model unstable. The regression is free to assign arbitrary values to the index in order to fit those two points. As a result, a choice was made to omit both 1994 and 1995 from the model, and use the aggregated 1994–1995 price level as the reference level. There is insufficient information in the dataset to fit individual year indices in those years.

This formulation ignores our prior knowledge that the price index is a time series, and that there should be significant serial correlation among the yearly values. Follow-on research might explore formulations in which the parameter to be estimated is the annual year-over-year increase, or overlapping multi-year average price increases. There are also more sophisticated time series methods that could be invoked.

b. Non-Quality Factors

In addition to overall change in prices, other factors besides vehicle quality affect the price in individual buys. It is useful to distinguish these factors from pure quality factors, in order to better isolate that portion of price change that is unexplainable by other means.

In previous work with aircraft, Harmon et al. (2014) found that cost progress curves (i.e., learning curves) explained a significant fraction of the observed lot-by-lot variation in price. For ground vehicles, we found that learning is not a significant driver of unit cost. Of the 53 vehicle types examined, only two or three showed even modest evidence of a learning curve effect. This is almost certainly due to the very high volume of production, where nearly all systems are produced in lots of hundreds or even thousands. Any learning effect would no longer be detectible after the first lot, at those rates.

Lot size, however, does appear to be an important factor, with volume discounts (perhaps explained by fixed cost dilution) for many types of vehicles. This factor was significant across several distinct families of quality model specification.

c. Quality-Based CERs

In a typical hedonic regression, price is explained by a combination of the price index (to be estimated) and set of product characteristics that capture, to the extent possible, what consumers are looking for in the product. Thus, for laptop computers, price might be modeled as a function of characteristics such as dimensions, weight, memory, processor speed, and screen size.

For military ground vehicles, the desirable features vary, depending on the purpose of the vehicle. For combat vehicles, the important characteristics are armor, weaponry, capacity, and mobility. Mobility can be further subdivided into speed on-road, speed off-road, and "trafficability," which measures what fraction of the terrain the vehicle can traverse. For tactical vehicles, the important characteristics are mobility, payload, reliability, and (in recent years) force protection.

It is difficult to find explicit data on many of these characteristics, and exact levels of force protection tend to be classified. The choice was made to work with proxy measures that are related to the characteristics of interest. From previous IDA work by David Gillingham (2009), we know that several aspects of mobility increase with the vehicle's horsepower-to-weight ratio (HP/ton), and that trafficability is closely related to the vehicle's ground pressure (i.e., pounds per square inch of ground contact). Weight, horsepower, and ground pressure could be found or estimated for each the vehicles in the data set described in Section B.[1]

For a given armor material, the degree of force protection on a vehicle is often measured by the areal density of the armor—the weight of armor per square foot of surface. As a proxy for this, we computed a notional surface area *A* for the vehicle based on length *L*, width *W*, and height *H* using the equation

$$A \approx 2(LH + LW + HW).$$

Gross vehicle weight (GVW) was then divided by *A* to get a rough areal density metric. This quantity was not only highly predictive of price: it was also less correlated with other quality variables than GVW.

2. Latent Quality Growth

We were not able to find data on every quality measure of interest for every vehicle type. In particular, detailed force protection data were only available for a small subset of vehicle types. Perhaps more importantly, we found evidence that programs made significant investments over time in quality aspects for which we had no useful metric,

[1] For wheeled vehicles, on-road contact area can be estimated accurately from tire size, by assuming an optimal inflation pressure for on-road travel. Off-road tire pressures are usually lower, increasing surface area to trade speed for traction. We did not attempt to model off-road performance of vehicles.

such as vehicle suspension, passenger seating, spall liners, reliability, internal climate control, and so forth. At the same time, we noticed that year-over-year price growth for most vehicle systems seemed higher than could be accounted for by simple inflation. Figure 2 shows unit price by year for tactical and support vehicles in the database. There is a clear trend of price growth, particularly in the early years of production, which simple regression estimates at roughly 14 percent per year within vehicle types over the entire data set.

To help explain this phenomenon, we considered models that include annual price growth due to latent quality growth. These models used the logarithm of lot number as a predictor. The basic model included a single term for all vehicles, which is equivalent to assuming that price growth over time is proportional for all vehicle types. Other models used instead the interaction of log(LotNum) with the vehicle type or vehicle family, allowing for the possibility that different families or types show different rates of price growth due to latent quality improvements.

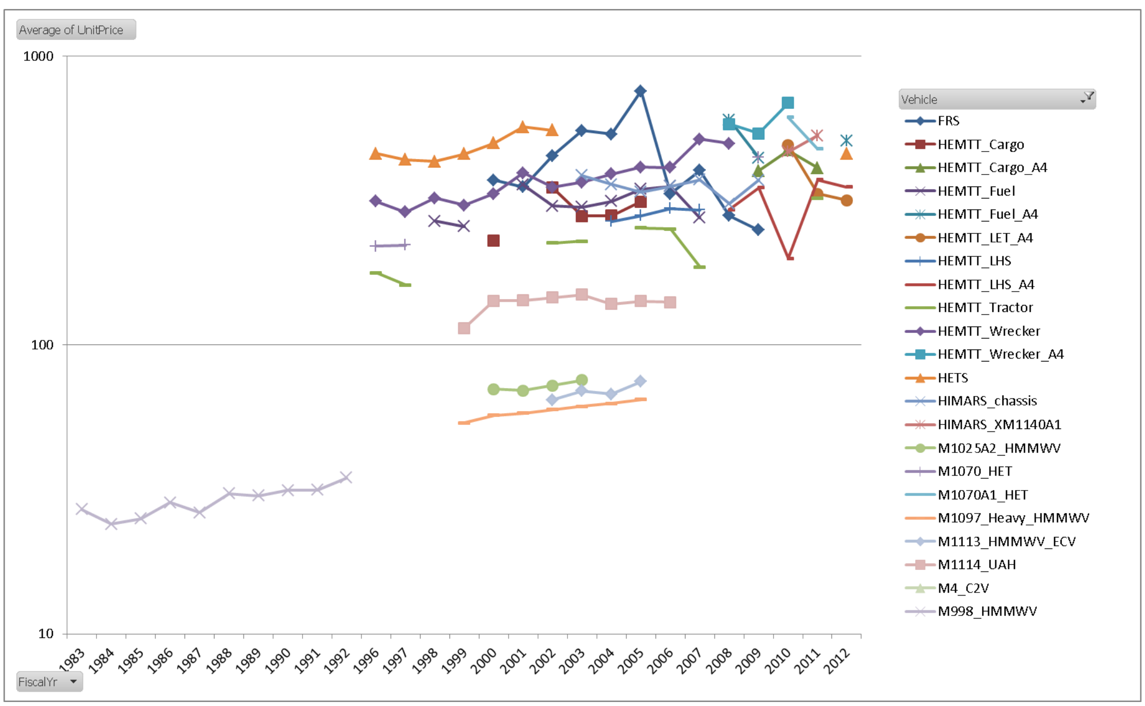

Figure 2. Unit Price by Year

3. An Alternative Formulation

The behavior of the hedonic models described above suggests that it might be reasonable to assume that vehicle type completely summarizes the initial quality of a vehicle, leading to a model in which price is explained solely by a fixed effect for vehicle type (i.e., “initial quality”), a rate effect capturing volume discounts, and possibly a "quality growth" factor for each vehicle family (e.g., LTV or TCV) or even vehicle type. This formulation essentially eliminates all explicit quality metrics from the model, and assumes that lot-by-lot price changes are entirely due to a combination of rate effects, steady latent quality growth, and inflation. Preliminary results and issues associated with this model are described in Section E.

E. Issues and Preliminary Results

1. Paucity of Data

As noted above, the fact that only one vehicle type provided data for the years 1994 and 1995 made it impossible to estimate price indices for those years. To make matters worse, there are indications that this was a period of particularly rapid price growth in the vehicle sector. As it stands, it is difficult to find stable estimates for the price index prior to 1996.

In addition, we have very little data on TCV prices in the current database. The M1 Abrams and M2 Bradley are the only TCVs in the data set. There are no TCV data points after 1993, and only one SAR-level data point per year prior to that for each of those programs. We do not have separate data for other Bradley variants (such as the M3 Cavalry vehicle or the Bradley Fire Support Team (BFIST) vehicle). Later variants of the Abrams tank were built by modifying existing hulls;, their prices cannot be compared to new production. Neither the Army nor the Marines have achieved full-rate production of a new TCV in the last two decades. Between the scarcity of pre-1994 data and the lack of post-1995 tracked vehicle production, our data may only support good estimates of a price index for tactical vehicles, rather than one for all ground vehicles.

2. Confounding Time Trends

As noted above, there have been significant changes in the mix of vehicle types being purchased over time. The 1980s were dominated by heavy tracked vehicles and HMMWVs. The 1990s were dominated by HMMWVs and tracked support vehicles. The 2000s were dominated by the Stryker family of vehicles, and later by force protection vehicles, as well as heavy tactical vehicles. This systematic change in mix could introduce systematic errors into estimates of the price index.

There have been other trends as well. The new interest in force protection vehicles to counter improvised explosive device (IED) and sniper attacks in Iraq and Afghanistan was paralleled by efforts to up-armor tactical and combat vehicles of all kinds. In some cases, we have separate model specifications for the up-armored variants, but in others (e.g., Stryker vehicles) we do not. Even where we have complete data, though, the new emphasis on force protection reflects a revaluation over time of that dimension of quality, which makes it difficult to apply the usual hedonic modeling framework.

3. Index Sensitivity to Sample

A potential issue in estimating the hedonic price indices is that the regression model can explain almost all of the variation in price without resorting to annual fixed effects. It is not difficult to find hedonic regression models that given an overall adjusted R^2 approaching 0.9, in then-year dollars, with no reference to time at all. This may be due in part to the time trends discussed above. In consequence, the model is to some extent treating the annual fixed effects—the index of interest—as a post-processing adjustment to minimize the magnitude of the residuals for an already good fit. There is a risk of overfitting here, with the index estimates being overly influenced by random variation, especially in the pre-1994 years. To test this, we ran jackknife regressions using random subsamples of the data, and bootstrap regressions using oversampled data.[2] In both cases, for a given regression model, there was considerable variation in the estimates of the specific yearly index values. Figure 3 shows the results of 50 bootstrap replications of a model whose adjusted R^2 on the complete data set is ~0.95.

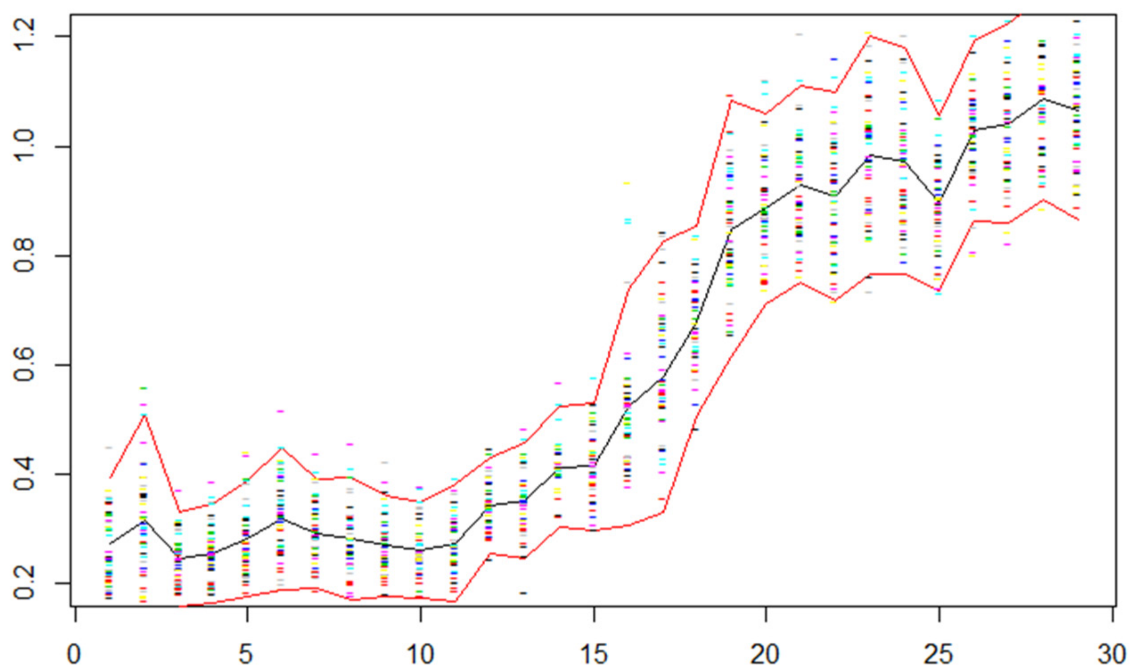

Figure 3. Results of 50 Bootstrapped Regressions

While the trend of the index is clear, it is not clear that precise values of the index can be stated with high confidence; however, bootstrapping the estimates does reduce the standard error of the estimates significantly. The red 90 percent confidence bands in the figure are roughly one-third as wide as the corresponding confidence bands derived from

[2] For discussions on these regression techniques, see Sections F.3.a. and F.3.b. of this paper on page 12.

the standard errors of the parameter estimates in the best single regression. We recommend bootstrapping to improve the precision of the index estimate, regardless of which model specification(s) is chosen.

4. Index Sensitivity to Specification

Perhaps more worrisome than the sensitivity to sample, which can (to some extent) be compensated for by bootstrapping, is the sensitivity of the price index estimates to the form of the regression model. We found a number of different model specifications that feature highly significant predictor variables, high adjusted R^2, and intuitively plausible coefficient values. Each of these models results in a somewhat different estimate of the price index, and some specifications result in a substantially different shape of the index.

There are at least two different mechanisms at work here. The first is sensitivity within a particular family of models that differ only slightly in their set of predictor variables. This is a significant source of instability in the price index estimates, and may be exacerbated by both multicollinearity within the predictors and correlation between the predictors and time. This instability could be mitigated using model averaging, if we were confident of the preferred basic form of the specification.

The second sensitivity is to the basic form of the specification. For example, specifications that use "vehicle family" as a treatment variable with multiple levels give similar fits (but somewhat different index estimates) to specifications that use a series of indicator variables for specific vehicle features, such as being armored, having a turret, having tracks, or being intended for a combat environment. Given comparable fits, significance levels, and coefficient plausibility, there are no strong grounds for preferring one specification type over the other—but they do lead to different estimates not only of the exact index values, but also of the overall shape of the index trend.

Finally, there is the question of latent quality growth. As might be expected, models that include latent quality growth terms attribute some of the observed price growth to that, leaving less unexplained price growth to be accounted for by the price index. This holds consistently for all specifications, and leads to yet a third proposed overall shape for the price index over time. Here, there is a philosophical issue that should be resolved before any further attempt is made to refine the estimate of the index: if latent quality growth is real and important, then only models including latent quality variables should be averaged (and/or bootstrapped) to estimate the index. Conversely, if we decide that we do not wish to hypothesize unobservable quality growth within vehicle types, we should not average over models that include latent quality variables.

The one piece of good news in all this is that there seems to be strong consensus among all specifications and samples concerning whether prices moved up or down in a given year. For example, nearly every model/sample combination tested produces an index in which there was a one-year drop in prices in 2006. This consistency provides some reassurance that we are modeling actual yearly changes in prices across an industry, and not just random noise.

F. Way Forward

1. Additional Data

One obvious way to improve the quality of the estimate of the price index is to find more data. This would help both to reduce the sensitivity of the estimate to sample and specification effects, and to allow extension of the index to additional years. Several potential data sources are available that we could pursue.

a. Family of Medium Tactical Vehicles

Currently, the data set in use includes many LTVs in the HMMWV family, and many HTVs in the HEMTT family. It does not include many MTVs and, in particular, does not include the many trucks bought by the Family of Medium Tactical Vehicles (FMTV) program during the 1990s and 2000s. The addition of FMTV data would improve the model's ability to distinguish family effects for MTVs, especially if we choose to include latent quality growth in the model.

Electronic budget justification forms are available back to about 1998, describing purchases from 1996 to date. This would not help to fill the gap in 1994 and 1995, and would not help flesh out the pre-1994 data set.

b. Army Tracked and Wheeled Combat Vehicle Database

Technomics Corporation maintains the Army Tracked and Wheeled Combat Vehicle Database, containing vehicle costs and specifications, as part of the Automated Cost Database (ACDB). Arrangements to receive this database are currently underway; however, at this time, we do not know exactly which vehicle systems are described in this database, or what price and quality data about those systems are included.

c. Contractor Cost Data Reports (CCDRs)

For the original tactical aircraft hedonic index work, Harmon et al. (2014) were able to draw extensively on contractor cost data reports (CCDRs), which provide a more detailed description of what exactly is being purchased—and how it differs from lot to lot—than is available through other sources. However, the data are proprietary, and availability is unknown.

2. Normalized Data

As noted previously, one concern regarding the current data set is that it uses primarily line-item data from the annual budget justification exhibits for units obligated after 1995, but SAR-level data for pre-1995 purchases. This makes it difficult to do an apples-to-apples comparison of prices, since the pre-1995 prices include things (such as documentation and engineering change orders) that are not included in post-1995 prices. This is one of several barriers to estimating the pre-1995 index. Where possible, we prefer to substitute line item data for SAR-level data currently being used.

3. Addressing Sensitivity to Sample

As noted previously, regression models that estimate the price index using annual fixed effects show quite a bit of sensitivity to the data sample used. Several traditional methods exist for trying to improve the precision of the estimates in the face of this sensitivity. Three of the most widely used techniques are the jackknife, the bootstrap, and cross-validation.

a. Jackknife

Jackknife techniques work by fitting the regression model to repeated random subsamples of the full data set, then averaging the parameter estimates over those repetitions. There are obvious limits to the utility of the method. Smaller subsamples allow for more repetitions (with associated Central Limit Theorem benefits in the averaging), but if the subsamples are too small, the loss in predictive power can offset the gains from averaging. If the subsamples are too large, you are essentially fitting the same data set repeatedly, gaining no new information. A theoretical rationale for choosing specific subsample sizes and repetition counts is beyond the scope of this paper.

b. Bootstrap

Bootstrap techniques are similar to jackknife techniques, except that sample size is held constant by sampling the full data set *with replacement*. Thus, in each repetition, some points occur multiple times, while others are omitted. This eliminates the problem of deciding on a subsample size and iteration count. There is a surprising body of theoretical justification for this seemingly ad hoc technique, including asymptotic results for the standard error of the averaged estimator.

c. Cross-Validation

Cross-validation can be thought of as a systematic (as opposed to randomly-sampled) jackknife. The dataset S is randomly partitioned into k (approximately) equal-sized subsets $S_1,\ldots,S_k$. The regression model is fit successively to $S \setminus S_j$, the data set with S_j omitted, for $j = 1 \ldots k$. The parameter estimates for the k repetitions provide an estimate of the sensitivity of the estimate to the data, and the averaged estimates over the k repetitions have lower standard error than the single-regression estimate over the full data set S.

The original form of cross-validation, used for very small samples, defined the subset S_j to simply be the j^{th} observation in S. This has the advantage both of preserving the largest possible sample for each regression, and of mitigating the influence of high-leverage data points. For large samples, k is typically chosen to be at least 25 so that the averaged estimators are approximately normally distributed.

4. Addressing Sensitivity to Specification

It is bad enough that the estimate of the price index is sensitive to the exact data used in the regression. In the absence of a sound theoretical basis for the specification of the hedonic regression, it is even more worrisome that the estimate of the price index is highly sensitive to the choice of specification. Sensitivity to data can be mitigated by the techniques described above. Sensitivity to the specification may not be as easy to get around.

There are two basic approaches to mitigating specification sensitivity in the index. The first has to do with increasing our confidence that we are using the "correct" sort of specification. The second uses averaging techniques to improve the precision of the estimate within a given family of specifications. Either or both may be appropriate to our situation.

a. CER Validation

Whether or not there is one "true" specification that best describes the relationship between the data and the index to be estimated, it is clear that some specifications are better than others. Given two models with identical adjusted R^2 and standard errors, we will always prefer the model whose parameters and coefficients do not conflict with our experience and intuition about what drives cost. For example, a model in which the predictive variable "vehicle weight" is assigned a negative coefficient is far less plausible than one that assigns the same variable a positive coefficient.

The first step in dealing with sensitivity to specification is thus to eliminate specifications that do not meet our threshold standards of plausibility or coherence. This sounds straightforward, but can be more difficult in practice than in principle. For example, for the vehicle data collected to date, it sometimes happens that specifications that include either predictor A or predictor B give plausible results with high adjusted R^2, but the specification that includes both A and B gives very different and counterintuitive results, with even higher adjusted R^2—and A and B are apparently uncorrelated.

Qualitative treatment variables contribute to this difficulty. Specifications that use the treatment "vehicle family" (e.g., LTV or FPV) seem to work well, but not all of the treatment levels are significant. Is it necessary to eliminate the insignificant treatment levels from the model before trying to estimate the price index? Or should we rely on our a priori certainty that there really are differences between vehicle families, and leave the less significant categories (and their nonzero coefficients) in the model?

Absent an engineering model of how cost arises from vehicle performance characteristics (with accompanying detailed performance data on all of the vehicles in the data set), it will thus be hard to say with any confidence which specification (or family of specifications) is preferred.

b. Model Averaging

We saw above that when a model is sensitive to the input data, we can use techniques like jackknifing or bootstrapping to average over a set of similar samples to reduce the variance of the estimate. Model averaging applies the same intuition to the specification, averaging over a set of alternative specifications to arrive at a consensus estimate of the unknown parameter. While this has mechanical similarities to bootstrapping, the underlying theory and implementation details are considerably more complex.

There are both Bayesian and Frequentist versions of model averaging. Both are potentially computationally demanding, but there are existing Stata and R packages available on the web that would allow us to implement either technique. There is some risk that the method would become a "black box" to us as analysts if we were to use those canned routines.

c. Latent Quality Growth

Before implementing model averaging, it would be important to decide whether we believe that there is unobserved quality growth over time within vehicle programs. If we do, we should average only over models that attempt to identify that growth. If we do not, we should average only over models that assume we have data on all of the relevant quality measures for each lot.

5. Pure Price Formulation

It is worth taking a moment to think about the role of the quality variables in a hedonic regression. If the available quality data perfectly characterize "what is being bought," those regression coefficients should capture the buyer's value tradeoffs among performance dimensions. At that point, it no longer matters which particular type of vehicle a given point represents; the type of vehicle adds no information beyond that contained in the quality variables, and should not be statistically significant.

If, on the other hand, the available quality data do not perfectly characterize "what is being bought"—that is, if there are aspects of vehicle performance that matter to the buyer and are not captured in our quality data—we can no longer be confident of our ability to compare quality across vehicles. We can, however, assume that a given vehicle represents the same quality bundle over time. Given a data set with enough different vehicle types, and multiple observations in all years, we can use this information to reconstruct a price index without attempting to model quality at all. In this "pure price" formulation, the "base price" (in real dollars) for each vehicle type would be modeled as a constant, modified only by production rate effects and a common price index.

This model gives good fits, but the average price growth rate over all vehicle types is roughly 14 percent annually, which seems extreme. This suggests (again) that there is latent quality growth within each vehicle program. We can modify the pure price formulation by explicitly accounting for this latent quality growth. The most straightforward model assigns each vehicle an initial base price, which is modified by production rate, common price index, and a growth function that is monotonically increasing over time. This growth function can be applied at the individual vehicle level (though this may lead to overfitting), or at the vehicle family level, or as a common growth rate for all systems.

The most straightforward version of this model is given by

$$\ln(U_j) = \beta_0 + \sum_{k=1}^{V} \beta_k I_{jk} + \sum_{k=1}^{V} \beta_{V+k} I_{jk} \ln(N_j) + \beta_{2V+1} \ln(Q_j) + \sum_{t=1}^{T} \alpha_t Y_{jt} ,$$

where V is the number of vehicle types in the dataset, U_j is the unit price of purchase j, Q_j is the lot size of purchase j, N_j is the lot number of purchase j, I_{jk} is an indicator that purchase j is of vehicle type k, T is the number of years covered by the data, and Y_{jt} is an indicator that purchase j occurred in year t. The parameters of the model are

- an intercept term;
- V fixed-effect terms, giving the base price of each vehicle type;
- V terms, giving the average annual quality growth rate for each vehicle type;
- a rate effect parameter quantifying returns to scale for a single annual lot; and
- T annual price index variables.

Applying this model to the post-1995 dataset yields an adjusted R^2 of 0.981. Using latent growth rates by vehicle family (rather than by individual vehicle growth rates) gives an adjusted R^2 of 0.979. The corresponding price indices are shown in Figure 4. The sudden drop in 2012 should be taken with a grain of salt, given that it is based on only five data points, all of which are new-model HTVs. Ignoring the anomalous sharp decline in 2012, both models show roughly 5 percent annual price growth unattributable to latent quality growth within vehicle types. Both models also confirm the real price drop in 2006 that was common to the various hedonic models. Table 2 shows these two estimates normalized to 2011, to avoid the questionable final year.

We have not yet checked the sensitivity of these models to data, but it seems likely that the variability would be lower than for the more complex quality-based specifications, and that bootstrapping would still be appropriate as a variance-reduction technique. We could then think of the resulting index estimate as a lower bound on the true index. We could also fix the base price parameters and re-fit the index without rate effects, if we wish to consider changes in lot sizes over time as part of the price change. This would parallel the "preferred model" (vice the "full CER model") of the work in 2014 on aircraft by Harmon et al.

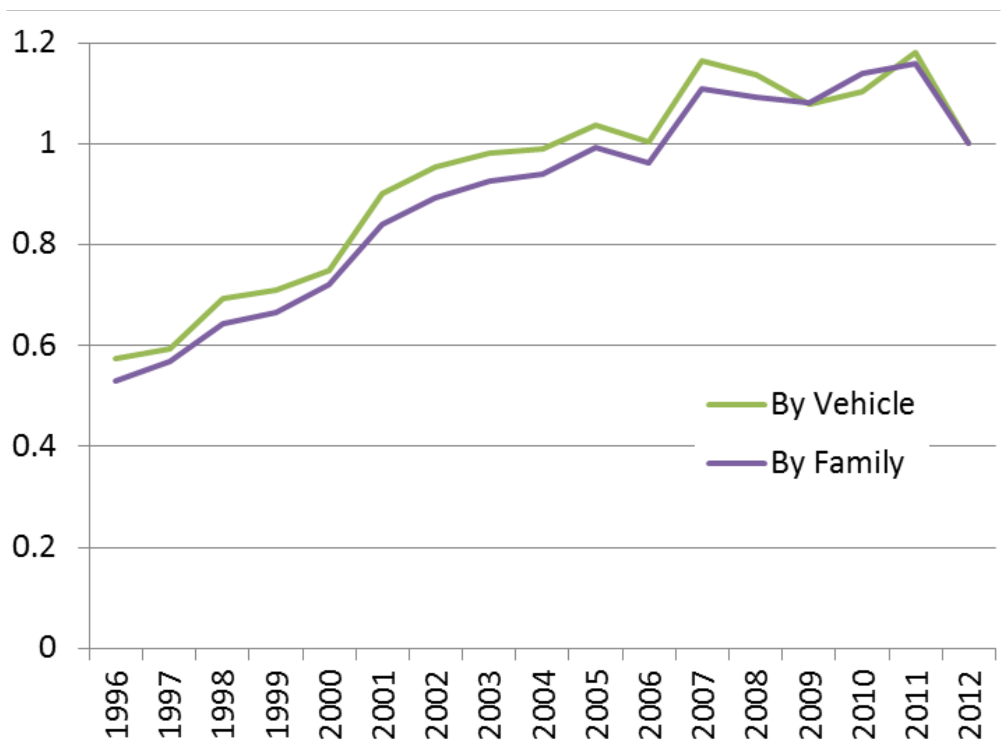

Figure 4. Pure Price indices

Table 2. Pure Price Index Estimates

Year	Index (by vehicle)	Index (by family)
1996	48.6%	45.7%
1997	50.3%	49.1%
1998	58.9%	55.5%
1999	60.2%	57.4%
2000	63.4%	62.2%
2001	76.4%	72.6%
2002	80.8%	77.0%
2003	83.2%	80.0%
2004	84.0%	81.2%
2005	88.0%	85.6%
2006	85.0%	83.1%
2007	98.7%	95.8%
2008	96.3%	94.4%
2009	91.3%	93.3%
2010	93.5%	98.3%
2011	100.0%	100.0%

G. Summary

We collected data on the price and specifications of a wide variety of tactical vehicles and a few combat vehicles. We then constructed various regression models attempting to predict unit prices in historical purchases as a function of the quality characteristics of the vehicle and the year of purchase. It was not difficult to find regression models with high adjusted R^2, but it was difficult to draw firm conclusions about how quality-adjusted prices have changed over time.

The current data set is not sufficient to estimate stable indices for years prior to 1996. Hedonic regression models using quality variables as predictors show considerable instability in their estimates of the price index; they are sensitive to both data variation and model variation. Sensitivity to data variation can be mitigated using bootstrap techniques. Sensitivity to model specification is somewhat more problematic, but there are a couple of avenues available to us. These include various forms of model averaging, or (in a completely different direction) moving to "pure price" models that eliminate the explicit quality variables and treat each vehicle type as a distinct product, with or without latent quality growth.

There is evidence of latent quality growth within vehicle programs, with quality improvements that are recognized by the buyer but not captured in the vehicle specification data available to us at this time. The main evidence for this is that naïve specifications that assume no latent quality growth lead to very high estimates of price growth—on the order of 15 percent annually since 1996.

Additional data would be extremely useful, especially if we want our index to apply to combat vehicles as well as tactical vehicles, and to years prior to 1996. There is some hope that the Army Wheeled and Tracked Combat Vehicle database will provide useful additional data.

References

Gillingham, David R. "Mobility of Tactical Wheeled Vehicles and Design Rules of Thumb." IDA Document NS D-3747. Alexandria, VA: Institute for Defense Analyses, June 2009.

Harmon, Bruce R., Daniel B. Levine, and Stanley A Horowitz. "Inflation Adjustments for Defense Acquisition." IDA Document D-5112. Alexandria, VA: Institute for Defense Analyses, October 2014.

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
xx-05-2015	Final	

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Hedonic Price Indices for Ground Vehicles	HQ0034-14-D-0001
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
Tate, David, M.	5e. TASK NUMBER
	BA-7-3054
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Institute for Defense Analyses 4850 Mark Center Drive Alexandria, VA 22311-1882	IDA Document NS D-5467 H 15-000310
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
Office of the Secretary of Defense Cost Assessment and Program Evaluation Economics and Manpower Analysis Division Room BE 798, The Pentagon Washington, DC 20301-1800	OSD(CAPE)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

As part of ongoing support to OSD(CAPE), IDA has been developing hedonic price indices for various types of military system. This paper reports preliminary results of investigations into hedonic indices for military ground vehicles, including both tactical and combat vehicles. We discuss issues associated with traditional hedonic methods, and propose a novel method designed to account for unobservable quality growth in military systems.

15. SUBJECT TERMS

Hedonic price index, military vehicle, quality attributes, unobservable factors, bootstrap, model averaging

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Summers, Garrett
Unclassified	Unclassified	Unclassified	Same as Report	24	19b. TELEPHONE NUMBER *(Include area code)* (703) 695-0722

Reset