An Analysis of Automobile Fuel Efficiency and the Time Dependence of Gasoline Prices: A Single-Vehicle Approach

Steve Sekula

September 12, 2008

Abstract

Data from a single automobile spanning the period from 2003-2008 are analyzed. Fuel economy and fuel cost are determined from the data. The raw data contained several challenges and inconsistencies which needed a specific data-handling procedure in order to proceed. The fuel economy data is compared favorably to the manufacturer's reported fuel economy, and seasonal variations in the data are studied. The fuel cost data are analyzed to look for microscopic and macroscopic trends.

1 Introduction

The price of gasoline has risen significantly over the past decade [2]. While market data is certainly one approach to studying this trend, another approach is to use extensive data from a single vehicle. Refueling records for a single car, regardless of its fuel efficiency, are an interesting probe of the larger economic trend of gas prices. In addition, long-term data from a single vehicle can allow a study of the fuel efficiency trends in that vehicle, both long-term and seasonal.

I report on the data from a single vehicle, taken during the period of January, 2003 to August, 2008. The data contains the total vehicle mileage, the total fuel put into the car at the time of each refueling, and the total cost of the fuel. The analysis of this data is shown to be a useful means to test data-handling skills. I show that the trend in fuel prices over time toward higher values should not have come as an unexpected effect in the current year, when gas prices are at their historical highest in the United States. In addition, I report on observations of the fuel economy of the vehicle over the period of the data.

2 Automobile Information

The automobile used in this study is a 2002 Dodge Stratus, purchased new from a Connecticut dealership. The sticker fuel economy is approximately 22 MPG [3]. The vehicle has received regular maintainence, which should insure stable fuel economy over time. This assumption will be tested. In addition, the vehicle has remained primarily in the New England area, making it subject to the varying weather conditions of the Northeastern United States. Possible trends in fuel economy, the result of varying climate conditions, will be tested. The data contains enough information to determine the approximate fuel cost (per gallon) as a function of time. Trends in that cost will be studied.

3 The Data Sample

The data were collected by the author's parents over a period of five years. The data are extensive, with only a few kinds of anomalies discovered during the analysis. The data were kept in a small notebook and arranged in columns. The columns contained the date of refueling, the mileage, the total cost of the fuel, and the total number of gallons pumped.

3.1 Possible Anomalies in the Data

Possible anomalies in a hand-recorded dataset are as follows:

- Errors in transcription of the numbers: one expects errors in the reported mileage, fuel cost, and fuel put into the vehicle. Such errors can be transcription errors (errors in recording), rounding (errors in the treatment of the raw data by the recorder), and data swapping (putting the wrong number in the wrong columns)
- Errors of omission: this analysis assumes that some fueling information is not recorded, leaving "gaps in the fossil fuel record". For instance, the author of this paper is guilty on several occasions of not recording the mileage when fueling the vehicle[1].

We can detect these errors as follows. One a priori expects several things to be true of the data. The price of gas varies significantly over long periods of time (months to years), but from refueling to refueling - even at different stations - one expects the price to vary fairly smoothly from one fueling to the next. Errors in transcription can easily be detected as significant outliers in the data, such as extremely high cost of fuel per gallon or exceptionally high fuel economy. Errors of omission involve several types. Either an entire fueling stop can be missed, which means that the gap in the mileage between the previous and next reported refueling is large while the next reported fuel cost is small relative to the distance driven. These are detectable as jumps in fuel economy that far exceed the laws of physics for a vehicle this size. For an individual fueling data point, columns can be missing. For instance, the mileage will be reported but the total cost of refueling is missing. In these cases, the fuel economy information can be recovered by combining data from the previous complete fueling point with the next one (adding the total number of gallons used and subtracting the two mileage numbers) to estimate the fuel economy.

3.2 How anomalies in the data were handled

The author had to make choices along the way in order to handle the data. These can be summarized as follows

- If a fuel economy estimate from a given entry exceeded 40 MPG, the data were assumed flawed (due to transcription, or due to a missing entry) and excluded from the analysis. There is no way a vehicle with a 22 MPG rating can spontaneously exceed 40 MPG in fuel economy, barring a strong tailwind or significant drafting during that period. Given the conservative nature of the drivers involved in the study (avoiding bad weather conditions, such as high-wind storms, or tailgating vehicles for long periods of time), I exclude either of these possibilities and justify ignoring this data.

[1] The existence of the record book for this information was only brought to the attention of this author in August of 2008, making my previous errors those of omission

Table 1: Error estimates on fuel economy, determined from several groups of "like points". Their likeness is determined from the distance driven since the last refueling stop.

Distance Range (miles)	Number of Data Points	Standard Dev. in Fuel Economy (MPG)
100 ± 10	16	4.0
150 ± 10	59	2.8
200 ± 10	27	4.2

- Data swapping of the total fuel cost and total fuel pumped were detected as anomalously low gas prices ($0.50 per gallon), which have not occurred in a very long time. By reversing the data points in the fuel cost ratio, a more reasonable gas price (~$2) was achieved. These entries were corrected by swapping the two numbers.
- For entries missing mileage information, but with all other information intact, the previous complete point was combined with the next one and the fueling data from the current one in order to get a fuel economy estimate. Fuel cost is intact in these entries.
- Several points exhibit anomalously high fuel cost. No obvious error in the data can be found (e.g. fuel economy looks reasonable). I assume that the total cost was not transcribed correctly, but I leave that data in the analysis.

3.3 Deriving an uncertainty on the fuel economy data

The fuel economy data are expected to vary from entry to entry due to climate conditions, driving conditions, driving habits, average speed during drive time, etc. Therefore, I developed a procedure to estimate the typical uncertainty due to these effects for a given entry. I group entries with a similar total distance traveled since the last fuel stop. For instance, I find all entries where (200 ± 10) miles were driven before refueling. I then assume variations for like points are Gaussian and find the standard deviation of these entries. I used the standard deviation as the measure of uncertainty on a single point.

I repeated this procedure for several central driving distances, well-separated from one another. The results are reported in Table 1. I find that for the sample distances there is some variation in the standard deviation. To be conservative, I take the error to be 4.2 MPG, with a systematic uncertainty due to the inability to control for driving conditions as the difference between the largest and smallest standard deviation, 1.4 MPG. Combining these in quadrature, I arrive at a total uncertainty on a given fuel economy number of ± 4.4MPG.

4 Analysis of the Fuel Economy Data

The fuel economy data, filtered for errors as described earlier, is shown in Fig. 1. A few things are clear upon inspection of this plot. There is more sparse data in the worst part of the winter (for instance, between 300 days and 380 days after 1/21/2003, which corresponds to late November and early February). This is particularly true in the winter months spanning 2003-2004 and 2004-2005.

I perform a fit to the data using the function

$$F(t) = a + b\cos(c \times t)$$

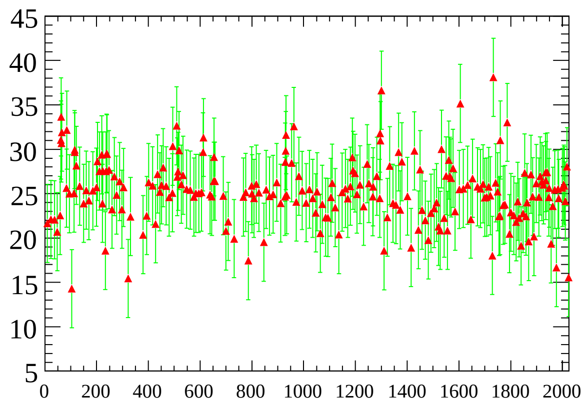

Figure 1: The fuel economy data (MPG vs. days since 1/21/2003) after correcting for errors in the data.

Table 2: Results from the fit to the fuel economy data

Coefficient	Fitted Value
a (constant)	25.2 ± 0.3 MPG
b (amplitude)	$-(16.1 \pm 0.4)$ MPG
$2\pi/c$ (period)	(369.4 ± 0.5) days

where a, b, and c are coefficients to be determined by the fit and t is the time (in days) since 1/21/2003. I find the fit converges reliably if I set the period ($2\pi/c$) to a number above 300 days. The binned χ^2/DOF is 0.6, which suggests that the errors are underestimated (or, perhaps more accurately, that my assumption of Gaussian uncertainties on these points is not an adequate assumption). The period determined by the fit is 369 days, which is relatively close to a one-year periodicity in the data. The fit result is shown in Fig. 2, and the fitted coefficients are given in Table 2.

From this analysis of the data, I conclude that the average fuel economy is in excellent agreement with the manufacturer's report (about 22 MPG). This average is maintained with great stability across many years and in the face of yearly variations in weather conditions that can have adverse effects on the vehicle.

5 Analysis of the fuel cost data

The popular media spend a lot of time in the current economic climate decrying the cost of gasoline. While each summer in the past decade has led to similar complaints with varying degrees of amplitude, this year (2008) signifies the loudest public outcry over the cost of fuel in about 30 years. The question I want to address with this analysis is as follows: should we have seen this fuel cost coming, at least based on constant upward change in the price of gas even despite seasonal market variations?

The fuel cost data is shown in Fig. 3. Data have been removed when there was not enough information to

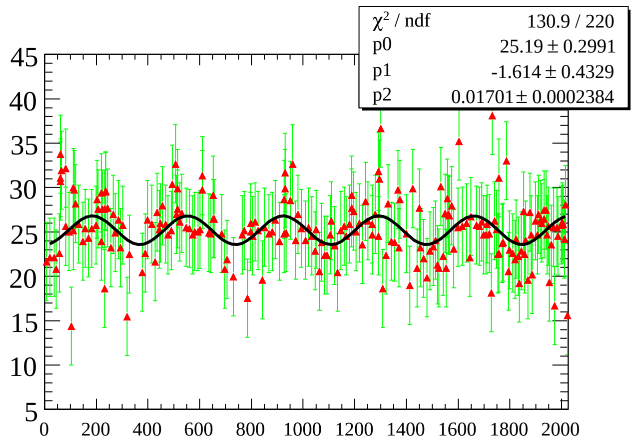

Figure 2: The fuel economy data (MPG vs. days since 1/21/2003) after correcting for errors in the data and fitting with a cosine-based fit function.

compute the fuel cost per gallon. Data points with no obvious problems, but which are clearly not smoothly varying with respect to their neighbor points (either too expensive or too cheap) remain in the data. Several interesting conclusions can be drawn from the data.

5.1 Correcting the data for inflation

The value of the dollar, relative to its value on 1/21/2003, varies over time. I correct the fuel cost data for inflation. I do this by using the Consumer Price Index to recompute the fuel cost in constant 2003 dollars, using the tool from Ref. [1]. I then replot the data in constant 2003 dollars in Fig. 4. Any dollar amount quoted henceforth in this analysis is in constant 2003 dollars.

A final cross-check of this data is performed by comparing the average CPI-corrected fuel cost in each year to data from the U.S. Department of Energy. Using the data in reference [2], I compute the ratio of next year's average price to the previous year. I perform the same calculation using the data from the single vehicle. The results are compared in Table 3. The rates of increase from year-to-year compare favorably between the two samples, except in 2006 when the rate of change in our data set was slightly larger than in the national average data. Based on this, I conclude that the data shows no long-term average trends that differ significantly from national averages determined by the U.S. Department of Energy.

5.2 Analysis of the corrected fuel cost data

A microscopic analysis of the data (looking at the feature of small subsets of the data) shows that the highest prices occur during summer periods (July-August of 2003, 2004, 2005, 2006, 2007, and 2008 occur at days 162-223, 527-588, 892-953, 1257-1318, 1622-1683, and 1987-2048). The peak-to-valley ratio in 2003 and 2004 was much smaller (by almost a factor of 2-3) than it was in all subsequent years. The large price spikes started in 2005 and have continued steadily since. While it is nearly impossible to make valid market

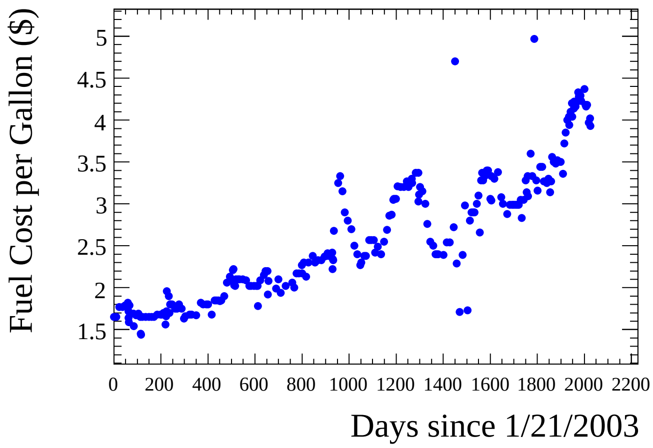

Figure 3: The fuel cost data, excluding points with not enough information to compute the cost per gallon but including several clearly inconsistent data points with no obvious external problems

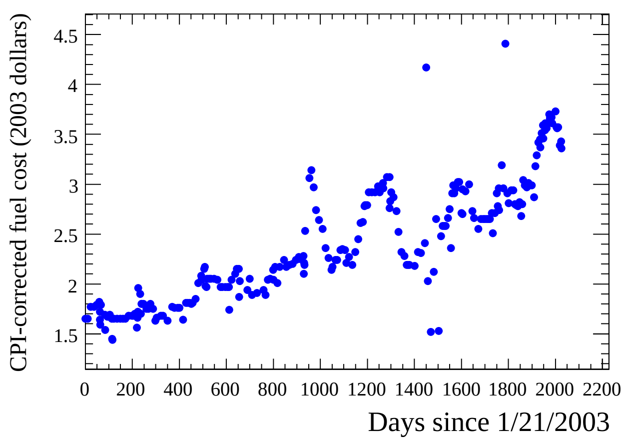

Figure 4: The fuel cost data in constant 2003 dollars, excluding points with not enough information to compute the cost per gallon but including several clearly inconsistent data points with no obvious external problems

Table 3: Data from the U.S. Department of Energy for the average national unleaded regular fuel cost, corrected to 2000 dollars, and the vehicle. The relative average price increase from year to year is computed from each data sample.

Year	Average Fuel Cost (2000 dollars)	Current-to-previous year ratio	Average Fuel Cost for the Vehicle (2003 dollars)	Current-to-previous year ratio
2003	1.50	N/A	1.70	N/A
2004	1.72	1.15	1.94	1.14
2005	2.03	1.18	2.29	1.18
2006	2.22	1.09	2.6	1.13
2007	2.34	1.05	2.74	1.05

predictions on the price of gasoline in a given year [2], based on the data no one should have expressed surprise after 2006 that the summer months were significantly more expensive than the winter months. This should have allowed for preventative budget planning for government agencies, business, and families.

Another microscopic observation is that in 2005, 2006, and 2007 the maximum price per gallon never exceeded a "ceiling" of about $3.00 per gallon (remember, this data is primarily from Connecticut). In the current year, this ceiling has been shattered. This may be the origin of the media and public outcry over the cost of fuel in the summer of 2008. While prices spiked in the summer consistently from 2005-2007, they never exceeded what might have been seen as an arbitrary "price ceiling". The year 2008 breaks this ceiling, and could be the origin of the stress on the economy and the public mind.

However, this same microscopic analysis - combined with the macroscopic observation that after spiking, prices tended to return to a curve that marked a steady upward trend - also suggests that the summer of 2008 should not have been a real surprise. The data in the time between summer 2007 and summer 2008 is the key.

First, note that after the spikes the data never returned to the pre-spike price. Instead, the post-spike price lies on a slowly rising curve that continues to rise under the spike (the "slow-rise" trend in the data). Increases of the average non-spike price continued steadily from 2003-2008. The maximum of the 2007 summer spike was reached by the increasing curve of the slow-rise curve just about 4-6 months after the summer of 2007 (in the winter of 2008). The data tells us that a 2008 summer spike that exceeded the old "ceiling" of $3.00 was not only possible (based on the existence of previous spikes), but very likely (given that the slow-rise curve already exceeded the old ceiling).

6 Conclusions

I have analyzed fuel economy and fuel cost data from a single vehicle, purchased new in 2002 and driven primarily in the Connecticut and New England areas. The fuel economy data suggests a seasonal variation that is roughly cosinusoidal, with errors (assigned by me) which likely underestimate the true uncertainty point-by-point. The average fuel economy of 25.2 MPG is consistent with the manufacturer's reported fuel economy, and is stable over the period of the data taking. The fuel cost data are consistent with trends in national fuel costs (albeit with an absolute scale relevant to the New England area), and suggest that

[2] The number of factors driving gas price - supply, demand, geopolitics, speculation - are largely unpredictable

there are two components to the fuel cost changes. One component is a slow-rising curve, which at no point between 2003 and 2008 stalled significantly. The second component are price spikes that correlate with the summer months. The spikes had a ceiling of about $3.00 (constant 2003 dollars) per gallon in 2005-2007, a ceiling which was exceeded by the slow-rise curve in winter of 2008 ahead of the summer spike of 2008. I conclude that significant fuel economy data from a single vehicle are an excellent means to perfect data-handling techniques and learn about vehicle performance and economic trends in the price of fuel.

References

[1] http://www.minneapolisfed.org/Research/data/us/calc/

[2] http://www.eia.doe.gov/emeu/aer/txt/ptb0524.html

[3] http://www.fueleconomy.gov/FEG/noframes/17542.shtml