Modeling Fuel Economy Benefits of Manual Transmissions Using Data From Automotive Enthusiast Publications

Executive Summary

Over the decades, *Motor Trend* has accumulated an archive of design characteristics and performance measurements of thousands of auto models. We can use these data to estimate the effect of automotive design choices on performance. In this report, I use data ($n = 32$) from the 1974 issues of *Motor Trend* to examine the effect that selection of an automatic or manual transmission had on fuel economy in the era immediately following the first major oil crisis. I develop a multiple linear regression model on transmission type and 3 other predictors of fuel economy, having $R^2 =$ 0.8975, and find that the best predictors are related to weight and power-to-weight ratio, not transmission type. The model indicates that selecting a manual transmission-equipped car improves fuel economy by about 0.7605 mpg, with 95% confidence interval (-1.2913 mpg, 2.8124 mpg)–implying that a fuel-economy *decrease* cannot be ruled out. The large variance in the estimate is unsuprising because of the relatively large variance of fuel economy observations, the small size of the dataset, and its bias towards unusual, editorially interesting cars.

Available Data

The dataset used here ships with the R statistical programming environment. It reports 11 variables for 32 1973 and 1974 model year cars. The variables are *Motor Trend*'s observed fuel economy and quarter-mile time, as well as the transmission type, number of forward gear ratios, final drive ratio, vehicle weight, gross horsepower, engine displacement, number and V or straight configuration of cylinders, and number of carburetor barrels.

I also added a factor to the dataset to indicate engine type. The Mercedes 240D had a diesel engine, which has fundamentally different fuel economy characteristics than a gasoline engine, and the two Mazda RX-4 models were powered by Wankel rotary engines rather than piston engines. I also removed the spurious, incorrect data regarding RX-4 cylinder configuration and number, treating both as `NA`.

Analysis of Confounders

As can be seen in Figure 1, the automatic-transmission cars tested by *Motor Trend* were more powerful and much heavier than the manuals. (The high-powered manuals are exotic sports cars, the Maserati Bora and DeTomaso “Ford” Pantera.) Weight and power and known determinants of fuel economy, so the effect of these confounders must be accounted for in order to correctly estimate the fuel economy response that is due solely to transmission type.

Figure 2 shows that all variables (except engine type) have strong correlations with fuel economy. However, many are correlated with fuel economy primarily because of confounding influences. For example, cylinder count: powerful, large-displacement engines happen to be built with more cylinders to reduce vibration.

Development of Multiple Regression Model

I accounted for the confounders by generating a multiple regression model using stepwise addition of predictors. I used physical reasoning to devise candidate predictors and used R's `add1` method to evaluate the candidates, choosing only predictors that reduce the AIC (Akaike Information Criterion), which rewards goodness of fit while penalizing overfit models with too many predictors.

I start by invoking `add1` on a model with *no* predictors:

```
add1(lm(mpg ~ 1, data = mt), names(mt))  # mt has cleaned data
```

According to the results, among all the variables in the dataset, adding vehicle weight (`wt`) to the model would best reduce the the unexplained variance and the AIC.

However, note that over a set course and with a fixed time profile of acceleration, the mechanical work done by the engine of a car is proportional to the force applied by the drive wheels, which is proportional to the mass of the car. Assuming roughly similar engine and drivetrain efficiency, this means the amount of fuel consumed over the course is roughly proportional to the mass of the car. In other words, we would expect `mpg` to be proportional to `1/wt` rather than `wt`. In fact, mpg is more strongly correlated with inverse weight, and in Figure 3 one can arguably discern a slighly nonlinear relationship between weight and fuel economy. `1/wt` reduces the AIC better than any of the other variables, so I selected it as the first predictor

I found that quarter-mile time (`qsec`) was a good candidate for the second predictor. However, `qsec` is a proxy for acceleration; assuming constant acceleration (an idealization) acceleration should be proportional to `1/qsec^2` (by $\Delta x = 1/2at^2$). Another factor proportional to acceleration is power-to-weight ratio hp/wt. Again because energy consumed is in the *denominator* of mpg, I added `qsec^2` and `wt/hp` and their inverses to the candidate predictors. I found that `wt/hp` best reduced the AIC and chose it as the second predictor.

An aditional slight in improvement in the AIC was obtained by adding `wt` (rather than its inverse) as the third term. Finally, I added transmission type (`am`) as fourth term in order to estimate its effect. The resulting $R^2 =$ 0.8975; the residual plot is shown in Figure 4. There is little pattern in the residual plot, though the two largest residuals correspond to the two cars with the highest fuel economy, the Toyota Corolla and Fiat 128. It stands to reason that engineering factors not represented in the dataset account for the high fuel economy of these pioneering “econoboxes”. As shown in Figure 5, the large residuals correspond to observations with low leverage, so the fit is little affected by the large residuals.

The expected increase in fuel economy when choosing a manual is the coefficient of the dummy variable corresponding to transmission type = `manual`: 0.7605 mpg. The standard error of the coefficient is 1.1361 mpg; with 27 degrees of freedom, the 95% confidence interval is (-1.2913 mpg, 2.8124 mpg).

Appendix

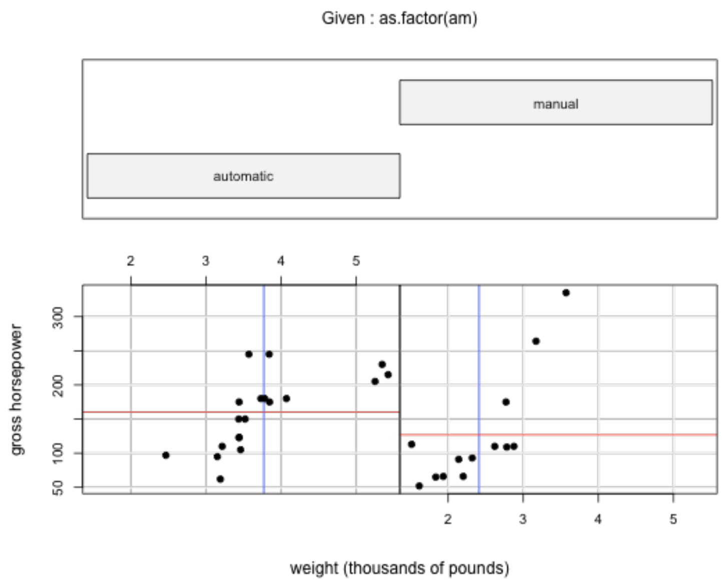

Figure 1

(*Horizontal and vertical lines denote the mean weight and mean power for the two cases*)

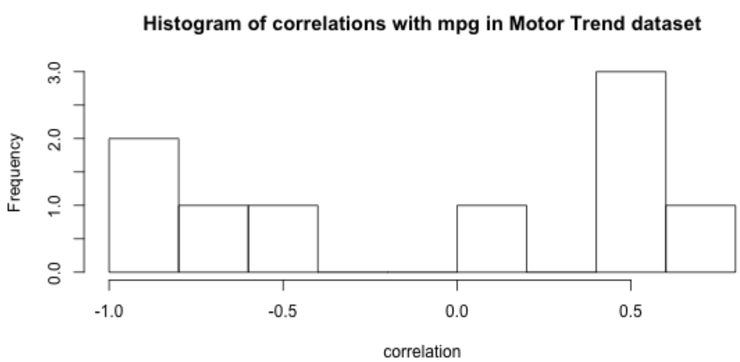

Figure 2

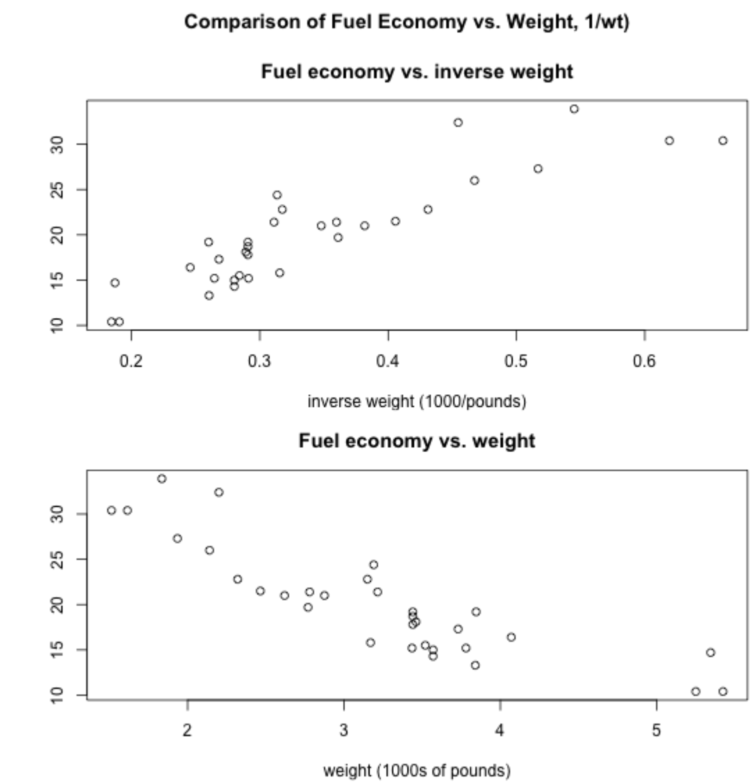

Figure 3

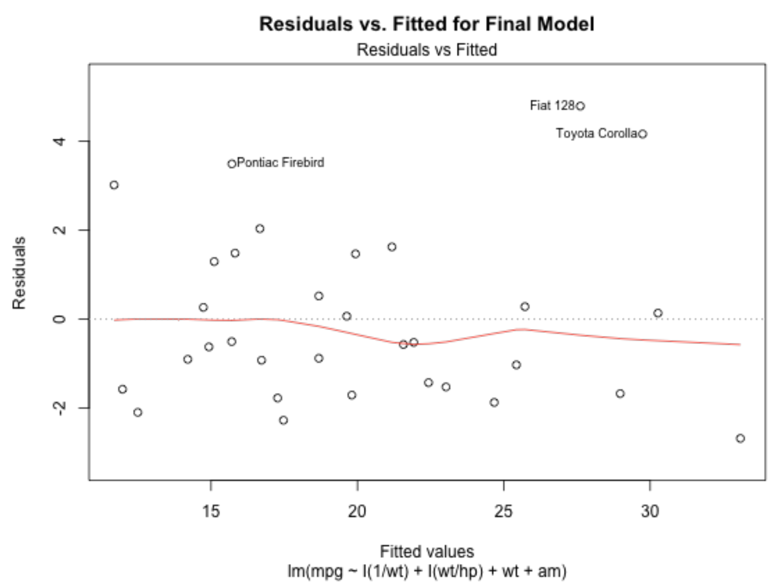

Figure 4

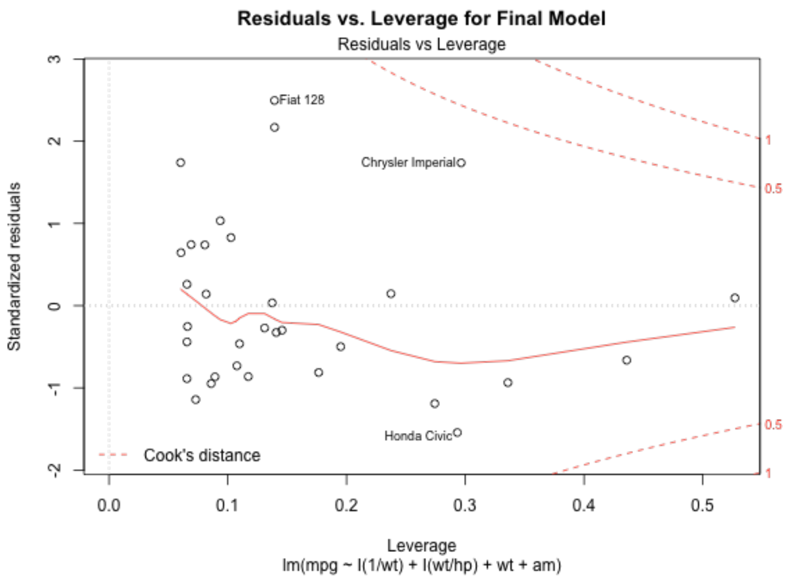
Residuals vs. Leverage for Final Model
Residuals vs Leverage
Standardized residuals
Leverage
lm(mpg ~ I(1/wt) + I(wt/hp) + wt + am)
Fiat 128
Chrysler Imperial
Honda Civic
Cook's distance
0.0
0.1
0.2
0.3
0.4
0.5
-2
-1
0
1
2
3
1
0.5

Figure 5