HARVARD UNIVERSITY
DEPARTMENT OF ECONOMICS

James H. Stock
Harold Hitchings Burbank Professor
of Political Economy
http://scholar.harvard.edu/stock/home

Cambridge, MA 02138

Littauer Center M-27
Telephone: 617-496-0502
Fax: 617-384-8362
james_stock@harvard.edu

To: The Honorable Elaine L. Chao, Secretary, Department of Transportation
    The Honorable Andrew Wheeler, Acting Administrator, EPA

From: James H. Stock, Professor, Department of Economics, Harvard University
    Kenneth Gillingham, Associate Professor, Yale University
    Wade Davis, Yale University

Date: October 26, 2018

Re: Comments on Notice of Proposed Rulemaking for The Safer Affordable
    Fuel-Efficient (SAFE) Vehicles Rule for Model Years 2021-2026 Passenger
    Cars and Light Trucks

    Docket number EPA–HQ–OAR–2018–0283 and NHTSA–2018–0067

Disclosure Statement

None of the authors of this comment have received any compensation for the work undertaken in this comment. Kenneth Gillingham has worked as a paid consultant for the California Air Resources Board on other analysis relating to fuel economy standards.

Executive Summary

This comment examines the modeling undertaken to forecast new vehicle sales in the proposed rulemaking "The Safer Affordable Fuel-Efficient (SAFE) Vehicles Rule for Model Years 2021-2026 Passenger Cars and Light Trucks" (83 Fed. Reg. 42,986 Aug. 24, 2018). The August 2018 Notice of Proposed Rulemaking (NPRM) and Preliminary Regulatory Impact Analysis (PRIA) describes the time series analysis used to project vehicle sales under different alternative standards, including the previous "augural" standards and the proposed rollback alternative, which hold the standard fixed at 2020 levels through 2026. This time series analysis uses an autoregressive distributed lag (ARDL) model to quantify the relationship between new vehicle prices and new vehicle sales. The analysis then takes an increase in new vehicle prices due to fuel economy standards from the U.S. Department of Transportation Volpe Center CAFE Model and uses the estimated coefficients from the ARDL model to estimate the change in new vehicle sales each year in the future.

We have five major concerns about this methodology. We were able to quantify the effect of the first three of these concerns. **Taken together, doing so reduces the estimated effect**

on light duty sales of a given price increase by approximately 87%, compared to the estimates in the NPRM/PRIA.

1. We uncovered a simple calculation error in the interpretation of the ARDL model. Specifically, the estimation uses quarterly data, so the estimated dynamics play out at a quarterly level. However, it appears that the NPRM mistakenly uses the quarterly coefficients for annual calculations. The NPRM provides an illustration of the implied dynamics, in which a $1,000 price increase results in 170,000 lost sales of light duty vehicles in the initial year of the increase and 600,000 lost sales over the 10 subsequent years. This calculation is in error because the calculation incorrectly assumes the quarterly coefficients apply to annual dynamics. **Correcting this single error reduces the NPRM lost sales estimate by approximately 70%**. For the NPRM/PRIA example of a $1,000 price increase, the correct values are 115,000 lost sales in the first year and 120,000 over the next ten years.

2. The ARDL model being used in the NPRM approach is inappropriate for estimating the demand for new vehicles due to the fact that supply and demand are simultaneously determined. The ARDL model requires all price movements, in the past, present, and future, have nothing to do with unobserved changes in demand. This assumption makes no sense in the market for light duty vehicles; for example, if sales were weak last quarter, a dealer might mark down its prices or hold a "blowout sale" to move inventory. **Using an approach (a vector autoregression) that allows prices to react to past demand disturbances further reduces the estimated effect of a price increase on sales.** Fully addressing the problem that prices are simultaneously determined – for example, a dealer marking down prices if sales in the *current* quarter are unexpectedly weak – would require a different method, specifically the use of instrumental variables.

3. The ARDL model assumes a specific mathematical structure for the dynamic effects of a price increase, as embodied in the number of lags of sales and prices used in the model. When this restrictive assumption is relaxed, both within the ARDL model and within the vector autoregression, the dynamics show greater mean reversion. As a result, **relaxing the NPRM/PRIA lag assumptions in a way consistent with standard econometric practice further reduces the estimated effect of a price increase.**

4. New vehicle fuel economy standards reduce the operating cost of a new vehicle. If consumers fully value fuel economy (as the NPRM argues they do), then this reduction in operating cost is, in effect, a reduction in the price to the potential buyer. **Thus, estimates of the effect on sales of a price increase using the model in the NPRM overestimates the effect on sales of a price increase that**

specifically arises from fuel economy technology. While the NPRM discusses attempts to address this concern by including fuel economy as a control variable, we are concerned that the time series data simply do not have enough variation to estimate this effect, especially for a highly persistent series such as average fuel economy. This is an important area that deserves further consideration.

5. The ARDL model in the NPRM is also misspecified because it does not adjust for improvements in the quality of new cars. If the task were, for example, modeling the demand for Brent oil, one would need to address the previous three concerns, but one would not need to worry that the commodity being purchased changes over time. Not so for automobiles, which have seen substantial and ongoing improvements in quality along the dimensions of safety, reliability, comfort, amenities, handling, etc. Thus, unlike Brent oil, the product being studied is changing over time. To address this problem requires adjusting for quality improvements. Doing so is difficult but has been the subject of ongoing research by academics and within the federal government Bureau of Economic Analysis (BEA), which employs quality adjustments in its personal consumption expenditure-new motor vehicle price index. We suggest using a quality-adjusted time series of new vehicle prices. It is not obvious to us what direction the results are biased due to this change, but it would unquestionably change the results.

In addition, we have some technical comments. One is that the levels specification implies a time-varying elasticity. We have tested for a time-varying elasticity, however, and do not find evidence of variation over time and thus recommend the far more conventional econometric approach of using logarithms of quantities, prices, and employment. A second is that the model assumes cointegration between sales and employment, and we confirmed that the data (strongly) affirm this modeling assumption, both in levels and in logs, indicating that this is not currently an issue, but it is something the Agencies should keep in mind for future modeling. These technical comments do not substantially affect the estimated effect of a price change.

These findings have important implications, as the magnitude of any delayed scrappage effect of older vehicles due to fuel economy standards scales with the effect on new vehicle sales. If the effect of fuel economy standards on new vehicles sales is negligible, there would be a negligible adjustment in used vehicle prices, and thus only a negligible change in fleet turnover and scrappage. As the NPRM argues that the augural standards would lead to thousands of crash fatalities due to delayed scrappage of older vehicles, it is crucial that the relationship between fuel economy standards and new vehicle sales is modeled correctly.

We strongly encourage the Agencies to entirely revise their approach to estimating the effect of fuel economy standards on new vehicle sales to correct the issues in the current approach.

Preliminaries: The NPRM Approach to Modeling New Car and Light Duty Truck Sales

This comment is on the August 2018 proposed rule titled the "The Safer Affordable Fuel-Efficient (SAFE) Vehicles Rule for Model Years 2021-2026 Passenger Cars and Light Trucks" (83 Fed. Reg. 42,986 Aug. 24, 2018). In a Notice of Proposed Rulemaking (NPRM), the U.S. Environmental Protection Agency (EPA) and U.S. Department of Transportation (DOT) proposed several alternative standards to the previous so-called "augural" standards that were analyzed in depth in a 2016 Draft Technical Assessment Report (TAR). The preferred standard in the NPRM holds new vehicle fuel economy or greenhouse gases constant at 2020 levels through 2026. For ease of readability, this comment will use the terminology "fuel economy standards" to refer to both the U.S. DOT fuel economy standards and the U.S. EPA vehicle greenhouse gas standards.

The August 2018 NPRM contains a new analysis that examines the effect of an increase in vehicle prices on the number of new vehicles sold. This analysis is used in the broader NPRM analysis to model how fuel economy standards increase the price of vehicles, which then can affect the total fleet size (and aggregate vehicle-miles-traveled or VMT). In the previous 2016 TAR, the Agencies assumed that the increase in vehicle prices from adding new technology to vehicles due to fuel economy standards is exactly balanced by the improvements in the quality of vehicles from the improved fuel economy, so that overall new vehicle sales would be constant.

If new vehicle sales do change, there may be ripple effects on used vehicle prices, and thus on the number of used vehicles that are scrapped. For example, if the higher vehicle prices due to fuel economy standards reduce new vehicle sales, this means that households not buying new vehicles will demand used vehicles, increasing used vehicle prices, and implying that some older used vehicles will be repaired rather than scrapped. The August 2018 NPRM aims to model this set of cascading effects, and in the NPRM modeling, many used vehicles are assumed to stay on the road longer under the augural standards, and through simple calculations, the NPRM argues that this leads to thousands of additional crash fatalities because the older vehicles are less safe than newer vehicles.

Consequently, it is important to understand how fuel economy standards affect new vehicle sales. This comment examines the econometric model used in the NPRM and uncovers several important issues. These comments are organized in the order in which they are listed in the Executive Summary. For the first three issues, along with the technical comments referred to following the enumerated items in the Executive Summary, we implement improved methods and provide the STATA code in an appendix.

We begin with a brief summary of the NPRM/PRIA ADL model. We then turn the main comments, followed by some additional technical comments and the conclusion. As explained in the technical comments, the use of levels of sales and prices results in a time-varying elasticity. This is nonstandard and when we formally test for a constant elasticity

we do not reject the assumption of a constant elasticity. A constant elasticity is implied by a specification that uses logarithms of sales and prices, so we undertake the analysis using both the levels specification and our preferred logarithmic specification.

Summary of the NPRM/PRIA sales model. A key observation in the NPRM, with which we agree, is that it is important to estimate the dynamic effect on sales of a price increase, that is, the causal effect on current and future demand of a price increase. This dynamic causal effect is a central object of interest in time series econometrics and differs from a static causal effect because it allows the response to an intervention – here, a one-time price increase or sequence of such increases – to evolve over time (Stock and Watson (2015, Chapter 15), Stock and Watson (2018)).

The NPRM uses an autoregressive distributed lag (ARDL) model to estimate the dynamic response of new vehicle sales to an autonomous increase in price, specifically,

$$S_t = \theta_1 S_{t-1} + \theta_2 S_{t-2} + \beta[P_t - P_{t-1}] + \gamma_1 GDPgrowth_t + \gamma_2 E_t + \gamma_3 E_{t-1} + u_t. \tag{1}$$

Here, S_t is the new vehicle sales in quarter t (millions of units at an annual rate), P_t is the average new vehicle transaction price (2016 dollars), $GDPgrowth_t$ is the percentage growth of GDP from quarters *t-1* to *t* (at an annual rate, seasonally adjusted), E_t is employment (thousands, seasonally adjusted)[1], and ϵ_t is an error term.

While it is not exactly clear how the NPRM analysis calculates the elasticity, the standard way to calculate a short-run price elasticities of vehicle sales would be to using the equation $\eta = \beta \frac{\bar{P}}{\bar{S}}$, where the bar over the variable refers to the mean taken over all values in the dataset.

Our analysis uses the data provided as attachment B in the letter from DiMarsco to Peters (October 23, 2018), pursuant to the Freedom of Information Act Request ES18-003395 by the California Air Resources Board. The data in attachment B provide the time series for *S*, *P*, *GDP* and *E* in the ARDL equation (1). The data are quarterly and cover the time frame 1967-2016. Henceforth we refer to this as the DOT data set.

The original source of the data in the DOT data set is the National Automotive Dealers Association (NADA) for annual average vehicle transaction prices, the Bureau of Economic Analysis (BEA) for automobile sales and GDP, and the Bureau of Labor Statistics (BLS) for employment. The NADA annual data are converted to quarterly data using an interpolation approach based on BEA new vehicle expenditure data. Specifically, in the letter from

---

[1] The NPRM refers to this as the "labor force participation rate", however the data description in the letter from DiMarsico to Peters (10/23/2018) states that the data are quarterly aggregates of FRED (https://fred.stlouisfed.org) series LNS12500000, which is total employment from the household survey (CPS). The standard definition of the labor force participation rate (FRED series CIVPART) is the fraction of the population age 16+ that is either employed or looking for work (unemployed). In this comment, we use the correct terminology and refer to the household employment data (LNS12500000) as employment.

DiMarsco to Peters (October 23, 2018), it states that the time series is "Annual (real dollar) transaction price, 1967-2016, transformed to quarterly price by using average per-vehicle expenditure variation from BEA to introduce quarterly variation into the NADA average annual price, normalized by quarterly sales to ensure annual average matched NADA annual series" (typos are verbatim from the letter). We have acquired NADA data for most of the times series, but are unable to replicate the interpolation approach used to go from the raw NADA data to the DOT data series for new vehicle prices.[2]

Using the DOT data, we replicated the NPRM ARDL model coefficient estimates to at least two significant digits for all coefficients except the coefficient on employment.[3] This allows us to more closely examine the model and perform a series of analyses to better understand the NPRM analysis.

1. Quarterly-to-Annual Calculation Error

The first issue in the NPRM analysis is a simple calculation error. To understand this error, we begin with a statement on page 43075: "Based on the model, a $1,000 increase in the average new vehicle price causes approximately 170,000 lost units in the first year, followed by a reduction of another 600,000 units over the next ten years." In the Volpe CAFE model used to calculate the costs and benefits of CAFE standards, the increase in average new vehicle prices in any given year is endogenously determined based on the technology that is being added by automakers in that year. However, we are able to confirm that this statement accurately characterizes how the CAFE model is using the coefficients from the ARDL model.

To find this result, it appears that the NPRM analysis confuses annual effects and quarterly effects. Table 1 column (a) reports the effects on Sales of a one-time increase in price of $1,000, based on our replication of the incorrect calculation in the NPRM ARDL model. The first column provides the estimated change in sales obtained by feeding the $1,000 price increase forward through the model. The initial effect is a decrease of roughly 170,000 units, just as reported in the NPRM. In the next time period, because of mean reversion in sales, the decline in sales is somewhat less, at 105,000 units. Taken together, the total loss in sales in quarters 2-11 is 600,000 units at an annual rate. These values – roughly 170,000 and 600,000 – match those in the NPRM.[4]

---

[2] An issue is that the vehicle expenditure data for light trucks does not go back to 1967. Thus, we are not sure how DOT managed to develop a per-vehicle expenditure measure. More transparency on this methodology would be welcome.

[3] Our coefficient on employment is of the same order of magnitude as in the NPRM but is not quite the same. We are not sure why this would be – perhaps there was a copy-and-paste error in putting the coefficients into the NPRM.

[4] Our estimates are 172,000 and 610,000, but we ascribe these minor differences to rounding in the NPRM.

To correctly compute the effect on total car sales annually requires (1) converting the sales data, which are quarterly sales at an annual rate, to quarterly sales, and (2) adding up the quarters within a year.

First we do (1). A standard convention in quarterly economic statistics is to report data at an annual rate, for ease of reference. Sales of, say, 16 million vehicles quarterly at an annual rate means that, if vehicle sales were to continue at that same rate for each quarter in the year, then annual sales would be 16 million. Said differently, actual sales in the quarter in this example were 4 million; if 4 million vehicles were sold in each of the next 4 quarters, then that year there would be 16 million vehicles sold. Because the data used to estimate the NPRM ARDL are quarterly sales at an actual rate, to obtain the actual number of sales lost in a given quarter, one must divide the estimates in column (a) by 4. The result of doing so is given in column (b).

Next we do (2), adding up quarters within a year. This calculation is done in column (c), for the first two years. In the first year, 115,000 sales would be lost. In the second year, 57,000 sales would be lost, and so forth. Over years 2-11 (quarters 5-44), a total of 120,000 sales would be lost. These values – 115,000 in the first year, and 120,000 in the next 10 years, are the correct values for the reduction in sales from the posited one-time permanent $1,000 price increase.

Our understanding is that the calculation in the NPRM that provided the 170,000 and 600,000 numbers is illustrative, and that the actual calculation is done using the Volpe CAFE model. However, our understanding is also that the quarterly coefficient estimates from the NPRM ARDL were included in the annual Volpe CAFE model without adjustment. Thus, this quarterly-annual conversion error illustrated in Table 1 applies directly to the calculations in Volpe CAFE model. This implies that the effects on vehicles sales in the Volpe CAFE model are substantially overstated, which also should reduce the importance of the fleet turnover and scrappage effect in the CAFE model.

Table 1. Corrected conversion of quarterly to annual estimates using our replication of the NPRM ARDL model.

Quarters after price increase (0 is the quarter of the price increase)	(a) Effect on sales at annual rate (thousands of units at an annual rate)	(b) Effect on quarterly sales (thousands of units)	(c) Effect on sales, sums over 4 quarters within a year (thousands of units)
0	-172	-43	
1	-105	-26	
2	-100	-25	
3	-82	-21	-115
4	-71	-18	
5	-60	-15	
6	-52	-13	
7	-44	-11	-57
8	-37	-9	
9	-32	-8	
10	-27	-7	
Sum quarters 1-10	-610	-153	
Sum quarters 0-10	-782	-196	
Sum 0-3 (first year)	-459	-115	-115
Sum 4-43 (next 10 years)	-480	-120	-120
Sum 0-43 (first 11 years)	-939	-235	-235

2. The ARDL Model is Inappropriate for Estimating the Demand Curve

2a. The ARDL model assumes that unexpected changes in demand have no effect on current or future prices.

A fundamental observation of economics is that prices and quantities are determined by the interaction of supply and demand. An increase in demand will, all else equal, put upward pressure on price which will bring forth more supply. Similarly, if an increase in production costs pushes up prices, then the higher price will reduce demand to the point that the quantity demanded at the higher price equals the quantity supplied. In general, this implies that observed data on prices and quantities reflect variations in both supply and demand. The econometric jargon for this situation is to say that both price and quantity are endogenous, that is, are jointly determined by market forces. It has long been recognized that this creates a problem for estimation of demand curves. If prices were set randomly – or, more generally, if prices are set in a way that is unrelated to demand – then a regression of quantity on price (or log quantity on log price) would estimate a demand curve. If, however, prices respond to shifts in demand, then the regression of quantity on price will not provide an unbiased estimate of the demand curve or (in the logarithmic specification) of the demand elasticity.

It is useful to distinguish two types of endogenous behavior of prices: the response of prices to past demand disturbances, and the response of prices to current demand disturbances. For the ARDL model to estimate the dynamic causal effect of a price change on demand, both these conditions need to be true, along with an additional assumption that prices not anticipate future demand disturbances. This trio of assumptions – past, contemporaneous, and future exogeneity – is referred to as strict exogeneity. Strict exogeneity is a key assumption for the validity of the ARDL model.[5]

The assumption of strict exogeneity of prices is not plausible in the context of car and light duty truck sales. For example, a dealer who had disappointingly weak sales in the previous quarter, and had thus built up an inventory on her lot, might decide to hold a sale in the current quarter. This behavior violates the past exogeneity condition. If the dealer holds a sale in the *current* quarter to counteract disappointing demand in that same quarter, then doing so violates the contemporaneous exogeneity condition.

The ARDL model (1) includes two drivers of demand, GDP growth and employment. For that model, the exogeneity condition is that prices must be strictly exogenous conditional on GDP growth and the current and first lag of employment. While these two variables control for some important drivers of demand, other demand disturbances remain, for

---

[5] Because the ARDL model includes lagged sales and prices as endogenous variables, current prices must be exogenous with respect to lagged sales and thus to lagged demand disturbances. For a complete discussion see Stock and Watson (2015, Section 5.5).

example driven by unexpected changes in operating costs (e.g. gasoline prices), in tastes, or for economic reasons not fully captured by national GDP growth and employment.[6]

It is useful to separate out the two questionable assumptions required for the ARDL model: the assumption that past demand disturbances does not affect prices (past exogeneity) and the assumption that contemporaneous demand disturbances does not affect prices (contemporaneous exogeneity). We address these in turn.

2b. The assumption of no dynamic response of prices to past demand disturbances can be relaxed using a vector autoregression (VAR) instead of an ARDL model.

The assumption that past demand shocks do not affect prices can be relaxed by explicitly allowing feedback from past demand disturbances to prices. This is done by adding an equation that explicitly models current prices as a function of past demand shocks, which are in turn captured by past prices and sales data. Because cars and light trucks are a major part of the economy, it is plausibly the case that there is some feedback from auto demand to overall macroeconomic conditions in the future. This logic suggests also including equations for GDP growth and employment that allow for this feedback.

The system of these four equations – one each for E_t, $GDPgrowth_t$, P_t, and S_t, – in which lags of each of these variables appears in each equation, is called a vector autoregression (VAR). Because of these multiple equations and the dynamic feedback in the VAR, the VAR resolves the problem of past exogeneity.

The VAR still needs an assumption similar to the contemporaneous exogeneity condition. To keep as close as possible to the NPRM ARDL model, we make the following assumptions: (a) GDP growth and employment do not respond to unexpected movements in auto sales or prices within a quarter; (b) conditional on GDP growth and employment, prices are contemporaneously exogenous for the purpose of estimating demand; and (c) sales can respond to price changes within a quarter. Together, these imply a Wold causal ordering for the series in the order of GDP growth, employment, the change in price, and sales. This VAR structure is conceptually close to the NPRM ARDL model. Specifically, like equation (1), the sales equation in the VAR includes lags of prices and sales, and includes contemporaneous prices, GDP growth, and employment. The major difference between the ARDL specification of this equation and the VAR specification is that the VAR specification allows for lags of GDP growth, and has the same number of lags of all the variables. This

---

[6] The strict exogeneity requirement can be violated for reasons other than the feedback from demand disturbances to prices described here. In fact, it requires that any omitted factor that affects sales be uncorrelated with past, present, and future prices. It is easy to see how this requirement could be violated. For example, vehicle sales could be influenced by a major quality improvement in vehicles in a given year. But such a quality improvement in vehicles would also be correlated with vehicle prices, as automakers would be expected to raise their prices along with the quality improvement. Thus, vehicle prices would not be strictly exogenous.

VAR specification no longer requires past exogeneity, although it does require contemporaneous exogeneity of prices (assumption (b) above). [7]

We estimated the VAR described in the preceding paragraph using the DOT data set, using the same four variables (E_t, $GDPgrowth_t$, ΔP_t, S_t). Because the ARDL has two lags of sales, we use 2 lags in the VAR (thus the VAR has two lags of all the variables, in each of the four equations). We did this both for the levels specification akin to the NPRM ARDL, and in our preferred logarithm specification for ($\log(E_t)$, $GDPgrowth_t$, $\Delta\log(P_t)$, $\log(S_t)$).

The estimated response of sales to a $1,000 one-time permanent price increase is plotted in Figure 2, for the (a) levels and (b) log specifications. The units are millions of vehicles sold at an annual rate. The shaded area represents a 67% confidence band (67% is conventional for VAR impulse response functions).[8]

Figure 2. Response of quarterly sales (SAAR) to a $1,000 one-time permanent price increase, estimated by a VAR with 2 lags of all variables, levels specification
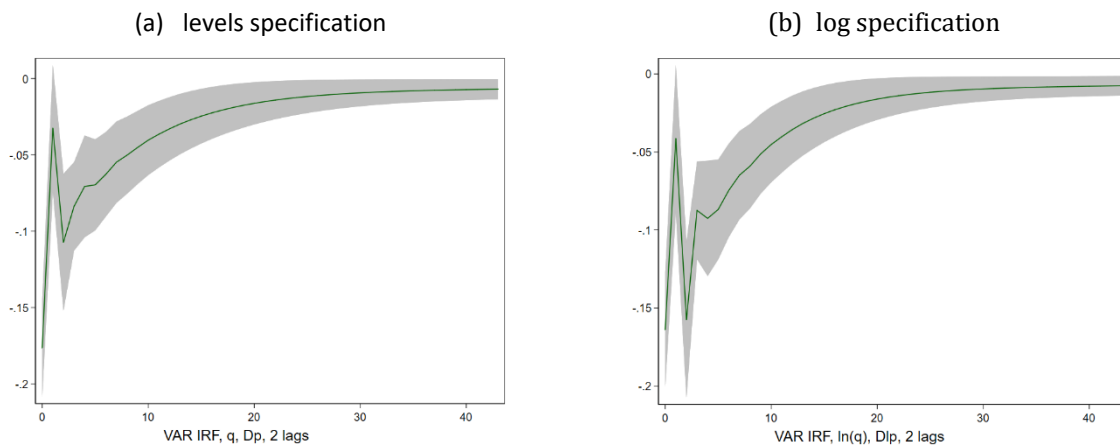
Notes: Shaded areas are 67% confidence bands. The VAR is estimated using the Wold causal ordering ($GDPgrowth_t$, E_t, ΔP_t, S_t).

Two findings are noteworthy. First, allowing for dynamic feedback, including the feedback of prices to previous demand disturbances, does not materially change the estimated effect on sales in the initial quarter, however it does change the dynamic path of effects. In particular, the dynamic response of sales to the price increase is less in the VAR than in the ARDL in Table 1 (column (a)).

Second, the results for the levels and logarithmic specifications are numerically similar.

---

[7] For the purpose of estimating the dynamic response of sales to a price disturbance, it does not matter whether GDP growth or employment comes first in the causal ordering.

[8] The logarithmic specification estimates impulse responses in percentages. These were computed to sales using a $32,000 base average price and 17.8 million vehicles of quarterly sales at an annual rate.

2c. Relaxing the assumption of no contemporaneous response of prices to demand disturbances requires using a different econometric framework, instrumental variables regression.

The problem that contemporaneous simultaneous determination of prices and quantities invalidates standard regression estimates of supply and demand elasticities is a classic problem in econometrics. This so-called simultaneous causality bias is a well-known problem that is described in undergraduate textbooks.[9]

The standard solution to simultaneous causality bias is the use of instrumental variables. In this case, the instrumental variable for estimation of the effect of a change in price on demand (that is, of the impact elasticity of demand) would require an instrumental variable that is correlated ("relevance") with price changes but is uncorrelated with unobserved demand disturbances ("exogeneity"). Work over the past decade has extended the concept of instrumental variables regression in time series data to estimation of the dynamic effects of structural shocks, such as an increase in vehicle price arising from higher technology costs stemming from fuel economy improvements. This extension of instrumental variable methods to vector autoregressions, and the assumptions needed to justify these methods, are summarized in Stock and Watson (2018).

Developing an instrumental variable for the purpose of estimating the impact elasticity of demand for vehicle sales constitutes a new research program that goes beyond the scope of this comment. We see this as a challenging research project but one worth pursuing given the importance of this task.

In the interim, we recognize that some estimates are needed of the effect on sales of a price increase, for preliminary use until improved estimates based on instrumental variables methods are obtained. We therefore make this important methodological note but proceed with our analysis and attempts to improve upon the NPRM ARDL model, while maintaining its core identification assumption: the contemporaneous exogeneity of prices conditional on current GDP growth and employment.

2d. Good one-quarter ahead and dynamic forecasts from the ARDL model (PRIA Figures 8-4 and 8-5) do not imply that price is exogenous or that the price coefficient estimates a demand elasticity

The PRIA provides two plots that are claimed to provide evidence that the ARDL model fits the data well and, by implication, that the price coefficient correctly estimates the effect of price on demand.

The first of these plots (PRIA Figure 8-4) is of the predicted values (one quarter ahead) from the model. Because the ARDL includes lagged sales, as well as current-quarter values

---

[9] For example, see Angrist and Pischke (2009, Ch. 4), Stock and Watson (2015, ch. 12), and Wooldridge (2016, ch. 15 and Sections 16.1-16.3).

of employment and prices, it is to be expected that this model fits the data well. In layperson's terms, if you know sales for this quarter and last quarter, and somehow you know GDP growth and employment next quarter, it is not hard to "predict" sales next quarter.

The second plot (PRIA Figure 8-5) is a dynamic forecast of sales, which uses the realized values of prices, employment, and GDP growth, but not actual sales. This is a substantially more demanding test of model fit, and as Figure 8-5 shows the model fits reasonably well. However, this does not imply that the price variable plays a meaningful role in that fit.

This point is illustrated in Figure 3. Figure 3 presents the quarterly (not annual, as in PRIA Figure 8-5) dynamic forecast of sales, computed two ways. The first is using the NPRM ARDL model in Equation (1). The second omits new vehicle prices as a covariate but is otherwise the same as model (1). Evidently, the dynamic forecast paths from the two models – the NPRM ARDL with and without new vehicle prices – are very close; indeed there is no evident reason to prefer one or the other based on their dynamic forecasts. Thus, the fairly good fit of the dynamic forecasts is entirely due to the use of GDP growth and employment in the ARDL model, and in particular does not provide confirmatory evidence on whether the new vehicle price coefficient in pinned down.

Figure 3. Dynamic forecast of quarterly sales from the NPRM ARDL model (levels), with and without vehicle prices.
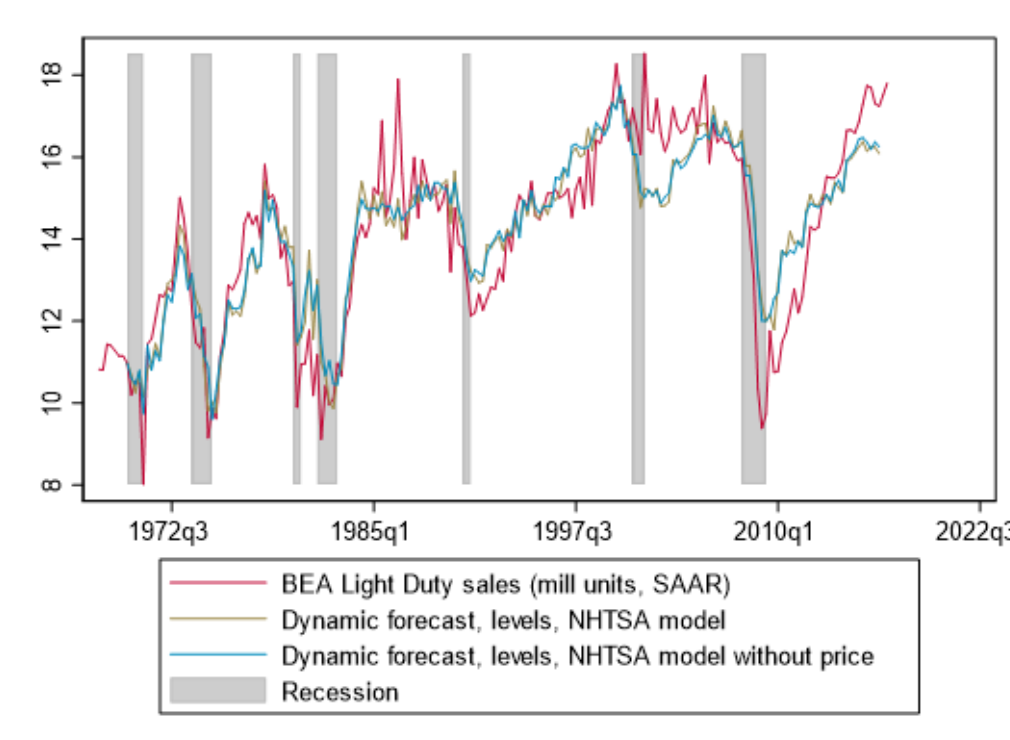
The key point here is that forecasting performance, either one-step ahead or dynamic, is not the primary way to ascertain whether the model is a reasonable model. Instead, the focus should be on unbiased estimation of the dynamic causal effect on demand of an autonomous price increase.

3. Relaxing the Short-lag Assumptions in the NPRM ARDL Model Reduces the Dynamic Effect on Sales of a Price Increase

3a. The short lag restriction of NPRM ARDL misses mean-reverting dynamics that reduce the effect on sales of an autonomous price increase.

In general, the autoregressive distributed lag model has p lags of the dependent variable (S_t) and q lags of the driving variable (ΔP_t). The NPRM ARDL model has short lags, $p = 2$ and $q = 1$ (a single contemporaneous value of ΔP_t). Such short lags restrict the shape of the impulse response function and can lead to biased estimates of the impulse response function if there is in fact a more complicated dynamic response. In the case of light duty vehicles, some new vehicle buyer might plausibly track price changes for a few quarters before ultimately deciding that now is the right time to buy. If so, sales might depend on more than just the current value of price; sales might depend on price trends over the past few quarters.

A-priori, it is not clear whether such longer-lagged effects are present or not, and it is not clear whether incorporating such longer-lagged effects would increase or decrease the cumulative effect of a price change on sales. Whether longer lags have an effect on the bottom-line estimates is an empirical question. The NPRM and PRIA do not explore the sensitivity of the estimates to choices of lags, so we do so here, first in the context of the NPRM ARDL, then in the context of our preferred specification, the VAR.

Figure 4 presents the response of sales to a $1,000 one-time permanent price increase for ARDL models of the form (1), using p lags of S_t and q lags of ΔP_t, for $p = 2, 3, 4$ and $q = 1, 2, 3, 4$, using the levels specification (Figure 3(a)) and logs specification (Figure 3(b)). The smallest values of p and q (2 and 1, respectively) correspond to the NPRM ARDL model, and that dynamic response is plotted in Figure 3 in solid black. The remaining dynamic responses are in patterned colors.

Figure 4. Impulse response function of sales to a permanent $1,000 price increase: ARDL(p,q) models, for lag lengths p = 2, 3, 4 and q = 0, 1,..., 5.

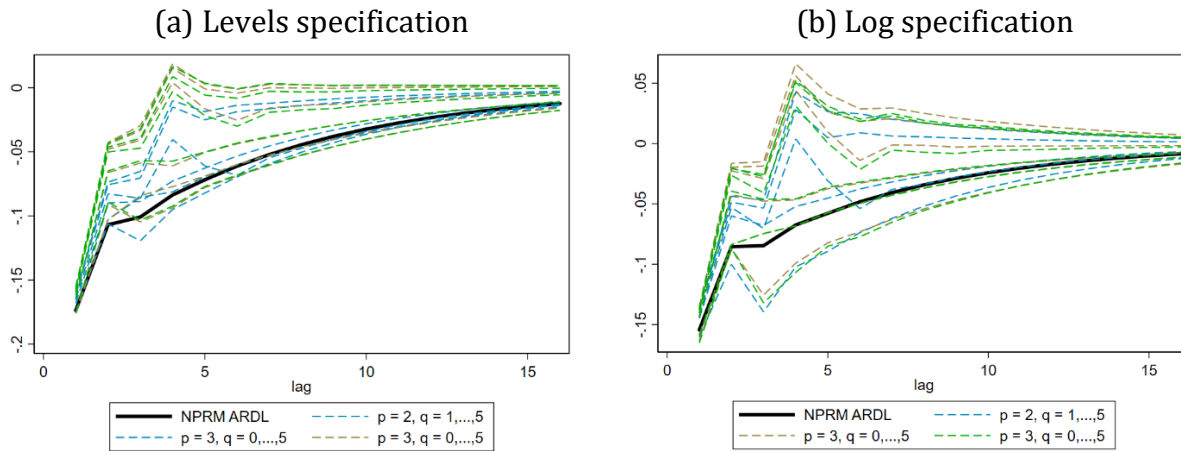(a) Levels specification

(b) Log specification

Figure 4 shows that increasing the number of lags in the ARDL model tends to reduce the dynamic effect of a price increase on sales. As we mentioned, the effect of relaxing the lag length assumption could go either way. Evidently, there are mean-reverting dynamics in the data that are not accounted for in the NPRM ARDL model because of its short-lag restrictions, that is, the NPRM ARDL model results are not robust to using longer, less restrictive lag assumptions.

Some of the responses in Figure 4 estimate that, after a few quarters, the dynamic effect of a price increase is to *increase* sales. We view this "wrong" sign in two ways. First, Figure 4 does not report standard errors so these positive effects may or may not be statistically significant. Second and more importantly, the ARDL model requires strict exogeneity which is implausible, so we might expect there to be issues with the ARDL model. As discussed above, we prefer the VAR specification to the ARDL model because the VAR does not require strict exogeneity of prices.

3b. Increasing the VAR lag length is supported by the data and reduces the effect on sales of an autonomous price increase.

We therefore explore the sensitivity of the VAR impulse response function estimates to longer lags. One guideline for determining the number of lags in a VAR is to use the Akaike Information Criterion (AIC). Applied to these data, the AIC chooses a VAR with 3 lags. We therefore consider a VAR with 3 lags and, to check on robustness to additional lags, also a VAR with 4 lags. Except for these longer lags, the VAR specification is identical to that described above.

Figure 5 presents the dynamic effect on sales of a $1,000 price increase for the VAR with 3 lags, and Figure 6 presents this dynamic effect for the VAR with 4 lags. Compared to the 2-lag VAR (Figure 2), the 3- and 4-lag VARs show more mean reversion in the dynamic

response. This is not an automatic effect of adding more lags; *a-priori*, the effect of adding lags could be to increase or decrease the dynamic effect. As it turns out, adding lags decreases the dynamic effect on sales of a price change. The dynamic effect is similar for the 3- and 4-lag VARs, that is, the results are robust to increasing the number of lags beyond those selected by the AIC. In this sense, the results of relaxing the lag length assumption in the VARs is qualitatively similar to relaxing the lag length assumption in the NPRM ARDL and reflect apparent mean reversion in the data that is not captured in the short lag specifications. The dynamic effects in Figures 4 and 5 have some quarters in which there is a puzzling positive sales response to prior price changes, however this positive effect is quantitatively small and, as can be seen by the shaded error bands, is not statistically significant.

Figure 4. Impulse response function of sales to a permanent $1,000 price increase: VAR model with 3 lags

(a) Levels specification   (b) Log specification
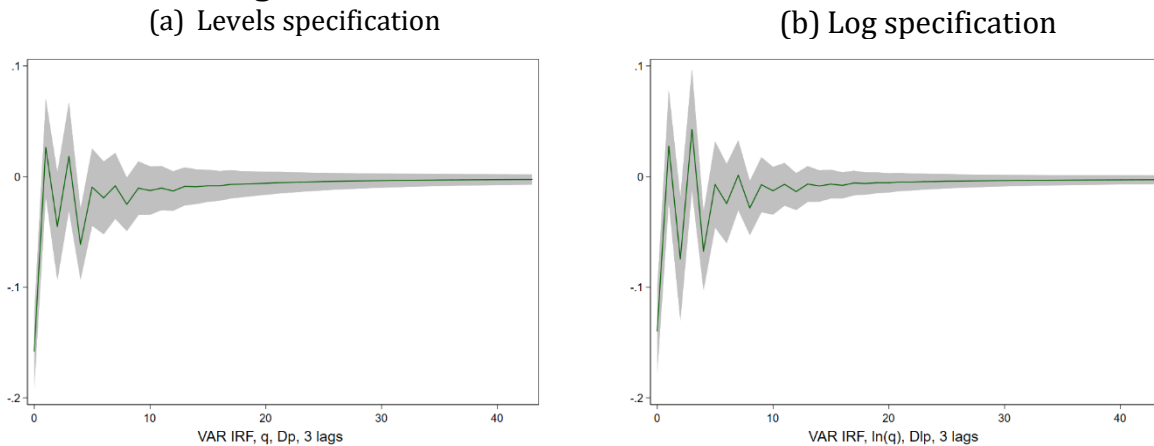
Notes: See the notes to Figure 2.

Figure 5. Impulse response function of sales to a permanent $1,000 price increase: VAR model with 4 lags

(a) Levels specification   (b) Log specification
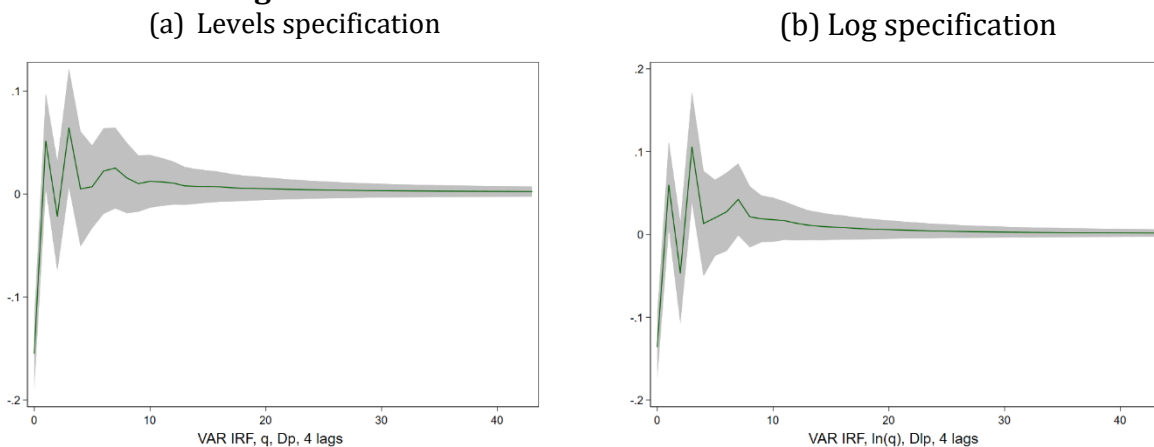
Notes: See the notes to Figure 2.

3c. Taken together, correcting the quarterly-to-annual conversion error, relaxing the strict exogeneity requirement by using a VAR, and using the longer lag VAR specification reduces the total effect on sales of an autonomous price change by an order of magnitude.

Table 2 summarizes the effects of the foregoing corrections and improvements to the estimated effect on sales of a one-time permanent $1,000 price increase. The total effect of correcting the conversion error, using a VAR to replace the strict exogeneity assumption with a contemporaneous exogeneity assumption, and relaxing the lag length assumption to use the AIC choice of 3 lags for the VAR, is to reduce the estimated 11-year total sales reduction from 770,00 vehicles in the NPRM (782,000 in our replication) to 111,000 vehicles (levels specification) or 101,000 vehicles (log specification). That is, the estimated effect on sales is, in our preferred VAR(3) log specification, 87% less than stated in the NPRM.

Using a VAR with four lags further reduces the total effect on sales because of even faster mean reversion, and in fact gives an unexpected positive effect on sales from the price reduction. These cumulative estimates do not include standard errors and we suspect these are not statistically significantly different from zero. In any event, we rely on the 3-lag VAR, which was selected by the AIC, with the log specification, as our preferred estimates. These are highlighted in Table 2.

Table 2. Summary of Model Estimates of the Effect on Sales of a $1000 one-time price increase, in thousands of units

Model	First quarter (at annual rate)	First year	Years 2-11	Total, years 1-11
NPRM ARDL, as reported in NPRM	a	-170	-600	b
NPRM ARDL, our replication	a	-172	-610	-782
NPRM ARDL, correcting quarterly-to-annual error	-172	-115	-120	-235
VAR(2), levels	-176	-99	-210	-309
VAR(2), logs	-162	-111	-233	-344
VAR(3), levels	-158	-39	-72	-111
VAR(3), logs	-139	**-34**	**-67**	**-101**
VAR(4), levels	-155	-15	+64	+48
VAR(4), logs	-135	-4	+86	+82

Notes: Entries might not add due to rounding.
[a] First quarter is incorrectly reported in the NPRM as first year. Our replication of the NPRM calculation in the second line does not correct this error.
[b] 11-year total not reported in NPRM
[c] Computed using our replication of the ARDL NPRM but correcting for the quarterly-to-annual conversion error per Table 1, column (c).

4. Fuel Economy Standards Reduce the Operating Cost of Vehicles

New vehicle fuel economy standards improve the fuel economy of the vehicle at the same time that they increase the price of the vehicle. Improved fuel economy lowers the operating cost of the vehicle, which is a quality improvement. Standard economic logic suggests that higher quality vehicles are more likely to be sold. Thus, there are two forces underway with fuel economy standards: the improved quality of the vehicles should increase sales and the higher prices should decrease sales.

Models like the ARDL or VAR are intended to look at the effect of exogenous changes in vehicle prices, rather than the effects of changes in vehicle prices that occur along with changes in the quality of the vehicle (e.g., improved fuel economy). Thus, they are only looking at one of the two factors – the factor that decreases sales. The NPRM recognizes

that this is an issue and discusses including average fuel economy as another variable in the ARDL model, and that the coefficient on the average fuel economy is not statistically significant.

While we appreciate this effort, we believe that this is a deeper issue that should be considered more carefully. If consumers fully value fuel economy – as the NPRM argues – then economic theory suggests that the quality improvement from lower operating costs would encourage more sales of new vehicles that may partly or entirely offset the change in new vehicle price. We are concerned that an approach that cannot model this effect is inherently going to be biased towards showing a larger effect of fuel economy standards on vehicle sales, and thus is in appropriate for regulatory analysis.

5. The NPRM Uses Non-Quality-Adjusted Data

As mentioned above, the NPRM approach uses the vehicle transaction prices from NADA. These transaction prices are not quality-adjusted. This is a concern because new vehicles today are dramatically different than new vehicles just a few decades ago. Automobiles have seen substantial and ongoing improvements in quality along the dimensions of safety, reliability, comfort, amenities, handling, etc. Thus, the unit of new vehicle sales is fundamentally different in 1970 than in 2010. Accordingly, it is inappropriate to examine the impact of non-quality-adjusted new vehicle prices on units of new vehicle sales. $20,000 (in real terms) is buying much more in terms of vehicle services in 2010 than in 1970!

Thus, the ARDL model in the NPRM is misspecified for not including quality-adjustment. Addressing this problem requires adjusting for quality improvements. Doing so is difficult but has been the subject of ongoing research by academics and within the federal government Bureau of Economic Analysis (BEA), which employs quality adjustments in its personal consumption expenditure-new motor vehicle price index. Thus, we would suggest using data on vehicle transaction prices that adjusts for changes in vehicle quality. The NPRM argues that one would not want to adjust for new vehicle quality because fuel economy standards are intended not to change quality. However, we do not agree with this assessment. If the goal is to estimate the demand for new vehicles, it is highly relevant that the transaction prices used in the analysis are buying different qualities of vehicle over time. However, it would take further exploration to determine the direction of the bias from this change, and the direction of this bias is not obvious to us *a priori*.

Additional Technical Comments

We make two additional technical comments.

Levels v. logs. The use of levels of the series implies that the demand elasticity depends on the values of sales and prices and thus varies over time. A more standard approach in econometrics is to specify demand curves so that the elasticity is constant, then to test

whether the demand curve varies either over time or perhaps in a way that depends on economic conditions or other variables. A demand curve with a constant elasticity is obtained by estimating the model using the logarithm of sales and the logarithm of prices, not their levels

To this end, we re-estimated Equation (1) using a logarithmic specification, specifically using $\log(S_t)$, $\log(P_t)$, and $\log(E_t)$ instead of S_t, P_t, and E_t. The estimated impact elasticity (that is, the coefficient on $\Delta\log(P_t)$) in this specification is -0.27, within the range of -0.2 to -0.3 given in the PRIA (p. 949). We then performed three tests of the hypothesis that the elasticity is constant, against the alternative that it varies with economic conditions. The first test allows it to vary with the state of the economy through the growth rate of GDP and/or the growth rate of employment (two interaction terms). The second test allows it to vary with the four-quarter growth rate of GDP and/or the unemployment rate. Both tests fail to reject the null hypothesis that the elasticity is constant at the 10% significance level. In markets that are subject to large qualitative changes, such as the market for smart phones, one might think that the elasticity changes over time as the uses of the product changes. For automobiles, however, the primary use of which has long been to provide personal transportation services through driving. Absent theoretical or empirical reasons to treat the elasticity as time-varying, we prefer the standard approach of assuming the elasticity is constant. The preferred specifications therefore use the logarithms of sales, prices, and employment.

Cointegration. The model relates the *level* of sales to the *first difference* of prices, the *growth rate* of GDP, and the *level* of employment. This use of the *levels* of sales and employment is justified only if sales and employment are cointegrated; if they are not cointegrated, then this regression will yield a spurious correlation between the two variables (Engle and Granger 1987; Stock and Watson (2015, Ch. 16.4). Thus this is an important assumption that needs to be verified. We therefore undertook Engle-Ganger Augmented Dickey Fuller tests of the null hypothesis of non-cointegration. We rejected the null hypothesis, against the alternative of cointegration, at the 1% significance level for both the levels and log specifications. Visual inspection of the variables, and of their estimated cointegrating residual (error correction term), confirms the presence of a shared stochastic trend between the logs of sales and employment. This analysis confirms the use of the levels of sales and employment in the NPRM ARDL model and in subsequent modeling of these variables, whether in levels or logs, so we follow the NPRM approach in this regard in our analysis.

Concluding Remarks

To summarize, our examination of the modeling of the effect of fuel economy standards on vehicle prices in the NPRM strongly indicates that the current modeling is inappropriate for regulatory analysis. We uncover a calculation error that reduces the effect of vehicle prices on sales modeled in the NPRM by 70%. Additional corrections using a model that allows prices to respond to prior demand shocks and relaxes the restrictive lag assumptions in the

NPRM model, further reduce the total sales effect, so that it is approximately 13% of what is stated in the NPRM.

We demonstrate that the current ARDL approach is flawed due to the very likely violation of the strict exogeneity condition necessary for this approach. We urge the DOT to modify the approach to use instrumental variables to identify the contemporaneous effect. Further, the current approach plausibly overestimates the effect on sales of a price increase associated with fuel economy improvements because those improvements reduce the lifetime user cost of the vehicle, improving the quality of the vehicle. Finally, we urge DOT to consider using quality-adjusted data to address the problem that vehicles are not a homogeneous good but rather evolve over time as quality improves.

References

Angrist, Joshua D. and Jorn-Steffen Pischke (2009). *Mostly Harmless Econometrics*. Princeton: Princeton University Press.

Stock, James H. and Mark W. Watson (2015). *Introduction to Econometrics*, 3rd Edition - updated. Boston: Pearson

Stock, James H. and Mark W. Watson (2018). "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *The Economic Journal* 128 (May), 917-948.

Wooldridge, Jeffrey M. (2016). *Introductory Econometrics*, 6th Edition. Boston: Cengage.

Appendix

STATA .do file for results in this comment

```
clear
set scheme s1color
set more off
cap log close
log using "newsales.txt", text replace
pause on
*******************************************
*  newsales.do
*  user needs to:
*     create a "fig" directory for output
*     load dependencies if necessary
*     ssc install egranger
*******************************************
import excel using "data\Attachment_B.xlsx", firstrow
drop Time
gen time = q(1967q1)+_n-1
format time %tq
tsset time
gen obsno = _n
* Variable nomenclature. All variables from FOIA
rename GDPGrowthRate gdp_growth
rename LDSalesAdj sales
rename NADAPriceAdj price
rename LaborParticipation emphh
rename Quarter quarter
rename Year year
desc
*
* transformations
foreach yy of varlist sales price emphh {
 gen d`yy' = D.`yy'
 gen l`yy' = ln(`yy')
 gen dl`yy' = D.l`yy'
}
*
label var sales "Light Duty sales (mill units, SAAR)"
label var price "Real vehicle price (2016$)"
label var emphh "Employment, HH survey"
label var lemphh "Log employment, HH survey"
label var gdp_growth "GDP growth (%, SAAR)"
*
******************************************************
* Replication of NPRM ARDL
******************************************************
reg sales L(1/2).sales D.price gdp_growth L(0/1).emphh
*
*----------------------------------------------------------------*
* Technical issue #1: log or level specification?
* Test null of constant elasticity vs. alternative that elasticity varies over time,
* specifically that the elasticity varies cyclically
*  Cannot reject null so prefer conventional (log) specification
```

```
*  Note: log specification uses log of employment to align with treatment of sales
*-----------------------------------------------------------------*
*
* Constant elasticity specification
reg lsales L(1/2).lsales D.lprice gdp_growth L(0/1).lemphh, r
*
* (1a) Test the hypothesis that the elasticity depends on gdp growth and/or employment
gen dlp_dlrgdp = (D.lp)*gdp_growth
gen dlp_dlemphh = (D.lp)*D.lemphh
reg lsales L(1/2).lsales D.lprice dlp_dlrgdp dlp_dlemphh gdp_growth L(0/1).lemphh, r
 test dlp_dlrgdp dlp_dlemphh
*
* (1b) Test the hypothesis that the elasticity depends on 4-qtr growth of gdp growth and employment
gen d4lrgdp = (gdp_growth + L.gdp_growth + L2.gdp_growth + L3.gdp_growth)/4
gen d4lemphh = 100*ln(emphh/emphh[_n-4])
gen dlp_d4lrgdp = (D.lp)*d4lrgdp
gen dlp_d4lemphh = (D.lp)*d4lemphh
reg lsales L(1/2).lsales D.lprice dlp_d4lrgdp dlp_d4lemphh gdp_growth L(0/1).lemphh, r
 test dlp_d4lrgdp dlp_d4lemphh
*
*-----------------------------------------------------------------*
* Technical issue #2: is there cointegration between (log) employment and (log) sales (yes)
*-----------------------------------------------------------------*
* levels
twoway (tsline sales, yaxis(1)) (tsline emphh, yaxis(2)), legend(cols(1))
 graph export "figs/cointegration_levels.png", replace
reg  sales emphh
 predict e_civ, resid
 tsline e_civ
 egranger sales emphh, lags(6)
egranger sales emphh, trend lags(6)
* DOLS estimator
newey sales emphh L(-4/4).d.emphh, lag(12)
*
* logs
twoway (tsline lsales, yaxis(1)) (tsline lemphh, yaxis(2)), legend(cols(1))
 graph export "figs/cointegration_logs.png", replace
reg  lsales lemphh
 predict le_civ, resid
 tsline le_civ
 egranger lsales lemphh, lags(6)
egranger lsales lemphh, trend lags(6)
* DOLS estimator
newey lsales lemphh L(-4/4).d.emphh, lag(12)
*
*-----------------------------------------------------------------*
*  Compute ARDL IRFS for NRPM lag specification and for additional lags.
*-----------------------------------------------------------------*
*
* (4a) NHTSA levels specification, ADL(p,q)
local cirflist ""
forvalues p = 2/4 {
 forvalues q = 0/5 {
  cap drop ytmp dptmp x1tmp x2tmp x3tmp f99*
  gen ytmp = sales
  gen dptmp = D.price
  gen x1tmp = gdp_growth
```

```
gen x2tmp = emphh
gen x3tmp = L.emphh
var ytmp, exog(L(0/`q').dptmp x1tmp x2tmp x3tmp) lags(1/`p')
 sca m99 = _b[_cons]/(1-_b[L.ytmp]-_b[L2.ytmp])
 replace ytmp = m99
 replace dptmp = 1000*(_n==101)
 replace x1tmp = 0
 replace x2tmp = 0
 replace x3tmp = 0
 fcast compute f99_, dynamic(80) step(100)
 gen cirf_adl`p'`q' = f99_ytmp - f99_ytmp[100]
 if(`p'==2 & `q'==0){
          gen cirf_adl_t = cirf_adl*1000
          /*qui sum cirf_adl_t if (_n ==101)
          di "q1 sales effect at annual rate: " r(sum)
          qui sum cirf_adl_t if (_n >= 102 & _n <= 111)
          di "sum of q2-q11 sales effects at annual rate: " r(sum)
          qui sum cirf_adl_t if (_n >= 101 & _n <= 111)
          di "sum of q1-q11 sales effects at annual rate: " r(sum)
          qui sum cirf_adl_t if (_n >= 101 & _n <= 104)
          di "sum of q1-q14 sales effects at annual rate: " r(sum)
          qui sum cirf_adl_t if (_n >= 105 & _n <= 144)
          di "sum of q5-q44 sales effects at annual rate: " r(sum)
          qui sum cirf_adl_t if (_n >= 101 & _n <= 144)
          di "sum of q1-q44 sales effects at annual rate: " r(sum)*/
          outsheet cirf_adl_t if (_n >= 101 & _n <= 144) using "figs/adl_irf.csv", comma replace
        }
 local cirflist = "`cirflist' cirf_adl`p'`q'"
 }
}
 preserve
 drop in 1/100
 replace obsno = _n
 tsset obsno
 list `cirflist' in 1/16
 tsline `cirflist' in 1/16, lc(black ebblue = = = = = brown = = = = = midgreen = = = = =) ///
   lp(solid dash ...) lw(thick medium ...) xtitle("lag") ///
          legend(cols(2) order(1 "NPRM ARDL" 2 "p = 2, q = 1,...,5" 7 "p = 3, q = 0,...,5" 13  "p = 3, q = 0,...,5"))
 graph export "figs/IRF_`pvar'_level_adlpq.png", replace
 restore
 tsset time
*
*  (4b) log specification, ADL(p,q)
cap drop cirf_adl*
local cirflist ""
forvalues p = 2/4 {
 forvalues q = 0/5 {
  cap drop ytmp dptmp x1tmp x2tmp x3tmp f99*
  gen ytmp = lsales
  gen dptmp = D.lprice
  gen x1tmp = gdp_growth
  gen x2tmp = lemphh
  gen x3tmp = L.lemphh
  var ytmp, exog(L(0/`q').dptmp x1tmp x2tmp x3tmp) lags(1/`p')
  sca m99 = _b[_cons]/(1-_b[L.ytmp]-_b[L2.ytmp])
  replace ytmp = m99
  replace dptmp = (1000/32000)*17.8*(_n==101)
```

```
 replace x1tmp = 0
 replace x2tmp = 0
 replace x3tmp = 0
 fcast compute f99_, dynamic(80) step(100)
 gen cirf_adl`p'`q' = f99_ytmp - f99_ytmp[100]
 if(`p'==2 & `q'==0){
          gen cirf_adl_t = cirf_adl*1000
          /*qui sum cirf_adl_t if (_n ==101)
          di "q1 sales effect at annual rate: " r(sum)
          qui sum cirf_adl_t if (_n >= 102 & _n <= 111)
          di "sum of q2-q11 sales effects at annual rate: " r(sum)
          qui sum cirf_adl_t if (_n >= 101 & _n <= 111)
          di "sum of q1-q11 sales effects at annual rate: " r(sum)
          qui sum cirf_adl_t if (_n >= 101 & _n <= 104)
          di "sum of q1-q14 sales effects at annual rate: " r(sum)
          qui sum cirf_adl_t if (_n >= 105 & _n <= 144)
          di "sum of q5-q44 sales effects at annual rate: " r(sum)
          qui sum cirf_adl_t if (_n >= 101 & _n <= 144)
          di "sum of q1-q44 sales effects at annual rate: " r(sum)*/
          *outsheet cirf_adl if (_n >= 101 & _n <= 144) using "figs\adl_irf_log.csv", comma replace
          }
 local cirflist = "`cirflist' cirf_adl`p'`q'"
 }
}
 preserve
 drop in 1/100
 replace obsno = _n
 tsset obsno
 list `cirflist' in 1/16
 tsline `cirflist' in 1/16, lc(black ebblue = = = = brown = = = = = midgreen = = = = =) ///
   lp(solid dash ...) lw(thick medium ...) xtitle("lag") ///
          legend(cols(2) order(1 "NPRM ARDL" 2 "p = 2, q = 1,...,5" 7 "p = 3, q = 0,...,5" 13  "p = 3, q = 0,...,5"))
 graph export "figs/IRF_`pvar'_log_adlpq.png", replace
 restore
 tsset time
*/
*-----------------------------------------------------------------*
* VARs
* NOTE: STATA cholesky ordering convention (from official STATA blog)
*  Suppose we have a VAR with three variables: inflation, the unemployment rate, and the interest rate.
*  With the ordering (inflation, unemployment, interest rate), the shock to the inflation equation can
*  affect all variables contemporaneously, but the shock to unemployment does not affect inflation
*  contemporaneously, and the shock to the interest rate affects neither inflation nor unemployment contemporaneously.
* SOURCE: https://blog.stata.com/2016/09/20/structural-vector-autoregression-models/
*-----------------------------------------------------------------*
*
*define program to estimate var irf using logs or levels
*  arguments are nsteps, plags, price_var, sales_var, emp_var and trans, an indicator for log transformation
*  make sure that trans = 1 for logs and 0 for levels
capture program drop do_var
program define do_var
          args nsteps plags price_var sales_var emp_var trans
          var gdp_growth `emp_var' `price_var' `sales_var', lags(1/`plags')
                  mat ss = e(Sigma)
                  mat sschol = cholesky(ss)
                  mat list ss
                  mat list sschol
```

```
            if (`trans'== 0){
                    sca qscale = 1000/sschol[3,3]
                    local xaxislab "VAR IRF, q, Dp, `plags' lags"
                    local trans_string "level"
            }
            else if (`trans'== 1){
                    sca qscale = (1000/32000)*17.8/sschol[3,3]
                    local xaxislab "VAR IRF, ln(q), Dlp, `plags' lags"
                    local trans_string "log"
            }
        cap drop qtmp ptmp etmp
        qui gen qtmp = qscale*`sales_var'
        qui gen ptmp = `price_var'
        qui gen etmp = `emp_var'
        qui var gdp_growth etmp ptmp qtmp, lags(1/`plags')
         qui irf create set1, set(vartmp1, replace) step(`nsteps') replace
         irf table oirf, std noci impulse(ptmp) response(ptmp qtmp)
         irf graph oirf, impulse(ptmp) response(qtmp) level(67) ///
                  byopts(title("") note("") legend(off)) xtitle("`xaxislab'") subtitle("")
         graph export "figs/IRF_`pvar'_`trans_string'_var`plags'.png", replace
        preserve
                qui use vartmp1.irf, clear
                qui keep if impulse == "ptmp" & response == "qtmp"
                * Note: sales are in millions at an annual rate, convert to thousands of quarterly sales
                gen oirf_t = oirf*1000/4
                di "year 1 effect annual rate: " r(sum)
                qui sum oirf_t if (step <= 3)
                di "year 1 effect: " r(sum)
                qui sum oirf_t if (step > 3)
                di "year 2-11 effect: " r(sum)
                qui sum oirf_t
                di "year 1-11 effect: " r(sum)
                outsheet oirf_t using "figs/var`plags'_irf_`nsteps'_`trans_string'.csv", comma replace
        restore
        qui tsset time
end
*
* (6a) VAR(2), VAR(3), VAR(4), levels
varsoc gdp_growth emphh dprice sales, maxlag(6)
do_var 43 2 dprice sales emphh 0
do_var 43 3 dprice sales emphh 0
do_var 43 4 dprice sales emphh 0
*
* (6a) VAR(2), VAR(3), VAR(4), logs
varsoc gdp_growth lemphh dlprice lsales, maxlag(6)
do_var 43 2 dlprice lsales lemphh 1
do_var 43 3 dlprice lsales lemphh 1
do_var 43 4 dlprice lsales lemphh 1
*
*------------------------------------------------------------------*
* Dynamic forecast plot
*  The apparently good fit of the dynamic forecasts in Figure 8.5
*  is due entirely to the gdp and employment variables, not price
*------------------------------------------------------------------*
*
cap drop f1_* f2_*
local n1 "1970q1"
```

```
local dn = q(2016q3) - q(`n1')
var sales, exog(D.price gdp_growth L(0/1).emphh) lags(1/2)
 fcast compute f1_, dynamic(q(`n1')) step(`dn')
 label var f1_sales "Dynamic forecast, levels, NHTSA model"
var sales, exog(gdp_growth L(0/1).emphh) lags(1/2)
 fcast compute f2_, dynamic(q(`n1')) step(`dn')
 label var f2_sales "Dynamic forecast, levels, NHTSA model without price"
local yy "sales"
 qui su `yy'
 local r99 = r(max)
 local r98 = r(min)
 twoway function y=`r99',range(39 43) recast(area) color(gs12) base(`r98') || ///
 function y=`r99',range(55 60) recast(area) color(gs12) base(`r98') || ///
 function y=`r99',range(80 82) recast(area) color(gs12) base(`r98') || ///
 function y=`r99',range(86 91) recast(area) color(gs12) base(`r98') || ///
 function y=`r99',range(122 124) recast(area) color(gs12) base(`r98') || ///
 function y=`r99',range(164 167) recast(area) color(gs12) base(`r98') || ///
 function y=`r99',range(191 197) recast(area) color(gs12) base(`r98') || ///
 tsline `yy' f1_sales f2_sales if tin(1967q1,2016q4), xlabel(,format(%tq)) ///
 legend(cols(1) order(8 9 10 1 "Recession")) lcolor(cranberry brown ebblue)
 graph export "figs/tsplot_`pvar'_nberdates.png", replace
*
*------------------------------------------------------------------*
log close
```