Documentation for Constructing Vehicle Service Flows, CE Survey 1980-2009

I. Data Cleaning

The following variables were cleaned for calculating vehicle service flows:

- **own_for**: the number of months that the CU has owned this vehicle (survey month & year minus purchase month & year).
- **vehageyr**: the age of the vehicle in years (survey year minus model year).
- **yrmade**: the model year for the vehicle.

In the full sample, the variables above are missing about 7 %, 10 %, and 10 % of the time, respectively. To replace missing or negative values for these variables we do the following:

1) If purchase month is missing but purchase year is observed, the purchase month is set to June. This affects 5.9 % of all observations.
2) Purchase year has been recoded in a small number of cases: 2022 recoded to 2002 (4 obs); 9996 recoded to 1996 (1 obs); 9999 recoded to 1999 (4 obs)
3) We impose a floor for **own_for** and **vehageyr** when these are negative (due to rounding). This affected less than 1% of all observations.
4) We replaced all missing values for **vehageyr** and **yrmade** whenever a) the car was purchased new and b) the purchase year is observed, using purchase year and survey year. This affects 4.8 % of all observations.
5) We replaced all missing values for **own_for** whenever a) the car was purchased new and b) **yrmade** is observed, using **yrmade** and survey year. This affects less than 1 % of all observations.
6) We assigned values for any remaining missing values for **vehageyr, yrmade,** and **own_for** as follows:
   a) We calculate the mean rounded to the nearest integer of **vehageyr (v_bar)** and **own_for** (**o_bar**) by survey quarter and whether purchased new/used across all observations where **vehageyr** and **own_for** are observed.
   b) A value for **vehageyr** is then imputed when missing as: **vehageyr** = MAX(**own_for**/12,**v_bar**)
   c) A value for **own_for** is then imputed when missing as: **own_for** = MIN(**vehageyr***12,**o_bar**)
   d) A value for **yrmade** is then imputed when missing as: **yrmade** = survey year – **vehageyr.**

Specifying make-model groups:

- The CE Survey includes a variable identifying unique make-model combinations for each vehicle. In some cases (about 8% of all vehicles between 1980 and 2005, but all vehicles 2006-present) only a make is reported. We treat these make-only cases as a separate model within that group of makes. For example, all Ford only vehicles are grouped together in a make-model group, but they are in a separate make-model group from Ford Escorts.
- All cars, vans, or trucks with a missing value for make-model or where make-model is designated as "other" are assigned to a vehicle-type specific make-model group. For

example, all cars with missing or "other" make-model are grouped together, while all vans with missing or "other" make-model are in a separate group.

- Make and model information is provided for cars, trucks, and vans, but not for "**other vehicles**," which include motorcycles, campers, boats, etc. For these other vehicles, all vehicles of a specific type are assigned to a single make-model group. For example, all motorcycles are grouped together.

II. Samples

For 1980-2009 we have 1,343,786 vehicles. 85% of these are cars, trucks, or vans. The other 15% are campers, motor cycles, boats, etc. We divide the vehicles into 5 samples:

1) Recently Purchased Vehicles—Estimation Sample (N = 189,787, 14 %): This includes all vehicles that have a reported purchase price, were purchased *within the 12 months* prior to the survey date and were not a gift. This sample will be used to estimate the relationship between vehicle characteristics, family characteristics, and vehicle market values. See Section III below.
2) Older Vehicles with Reported Purchase Price (N = 201,586, 15 %): This includes all vehicles that have a reported purchase price, were purchased *more than 12 months* prior to the survey date, and were not a gift. For this sample we will compute a service flow by converting the reported purchase price to survey year dollars and using an estimated depreciation rate as described in Section V below.
3) Prediction Sample 1: Vehicles with exact make-model and year match (N = 923,966, 69 %): This includes nearly all vehicles that do not have a purchase price, or do have a purchase price but the vehicle was a gift. NOTE: matches include vehicles that match on missing make. There are two groups of vehicles that are not included: a) vehicles where there is a make-model match in the estimation sample, but not a match on make-model *and year*, (see Prediction Sample 2) and b) vehicles where there is not a match in the estimation sample on any of the following: make, model or year (see Prediction Sample 3).
4) Prediction Sample 2: Vehicles without an exact make-model and year match (N = 21,966, 2 %): This includes vehicles for which there is an exact make-model match, but not a match on make-model and year. However, we can obtain a match on make (i.e. all Fords) and year.
5) Prediction Sample 3: Vehicles with obscure model years (N = 6,481, 0.5 %): This includes vehicles for which there is not a match on year in the estimation sample. Nearly all of the vehicles in this group are other vehicle types (motorcycles, campers, boats, etc.). Vehicles are in this sample because they were made in an obscure year. All vehicles in this sample were made between 1945 and 1976.

III. Vehicle Value Estimation

Using the estimation sample, we estimate three separate regressions, one for each of the prediction samples. All dollar values are converted to real 2005 dollars using the CPI-U-RS.

1) Estimates to predict values for Prediction Sample 1: Using the Estimation Sample we regress:

Log real purchase price$_{ivy}$ = B_1(car age$_{ivy}$) + B_2(car age$^2_{ivy}$) + B_3(car age$^3_{ivy}$) + B_4(car_characteristics_dummies$_{ivy}$) + B_5(family_characteristics$_{ivy}$) + B_6(non_missing_indicators$_{ivy}$) + γ_{vy} + ε_{ivy} (1)

where *ivy* refers to vehicle *i* of vehicle make-model group *v* with model year *y*, and γ_{vy} are vehicle group-year fixed effects. If we estimate this in levels, more than 5 % of the out of sample predicted values are negative. Vehicle purchase price is expressed in 2005 dollars using the CPI for new and used vehicles. Vehicle groups are constructed as follows:

- Specification of make-model groups are described in Section II above.
- Cars, trucks, and vans in the prediction sample for which we observe make-model, but there is no match on make-model in the estimation sample nor is there a match on make only, but there is a match on year are included in the "other" make-model group.

Other Covariates:

- car_characteristics_dummies: for all of these variables, the dummy = 1 if the car has this characteristic and 0 if not or if missing. These characteristics include: auto transmission, power brakes, power steering, air conditioning, diesel, sunroof, turbo charge, 4-wheel drive, and number of doors. Note, the last four are missing for all cars in the survey pre 1988.
- non_missing_indicators: these include a dummy indicating missing values for each of the vehicle characteristics listed above (1 = not missing). For about 89% of the estimation sample, we observe car characteristics such as power steering, ac, etc. For about 80% of the prediction samples, we observe car characteristics.
- family_characteristics: these include log expenditures (excluding vehicles and health), education dummies, family type dummies, family size, age of head, age of head squared, and region dummies. Expenditures are expressed in 2005 dollars using the CPI-U-RS.

2) Estimates to predict values for Prediction Sample 2 Using the Estimation Sample we estimate Equation (1), but the vehicle group constructed for the fixed effect (γ_{vy}) is all cars of the same make broadly defined (i.e. all Fords), rather than all cars of the same make and model. We do this so that we can predict the values of vehicles for this sample using vehicles of the same make, rather than using all unknown and obscure makes and models.

3) Estimates to predict values for Prediction Sample 3: Using the Estimation Sample we estimate Equation (1), but rather than vehicle group-year fixed effects, we include vehicle type fixed effects (i.e. car, truck motorcycle, boat, etc). These vehicle types are not interacted with year because vehicles in this residual sample have obscure model years, which is why they do not have a vehicle group-year match in the estimation sample. Thus, for this group, values are estimated off of age, vehicle type, and demographics.

IV. Vehicle Depreciation Estimation

From the Estimation Sample we construct a subsample of vehicles that are in a vehicle group-year with at least 2 vehicles that are not the same age. For estimating depreciation rates we exclude vehicles in the "other" make/model group. For this sample we estimate:

$$\text{Log real purchase price}_{ivy} = B_0 + B_1(\text{car age}_{ivy}) + \gamma_{vy} + \varepsilon_{ivy} \quad (2)$$

where *ivy* refers to vehicle *i* of vehicle group *v* with model year *y*, and γ_{vy} are vehicle group-year fixed effects. We also consider other specifications:

$$\text{Log purchase price}_{ivy} = B_0 + B_1(\text{car age}_{ivy}) + B_2(\text{car age}_{ivy})^2 + B_3(\text{car age}_{ivy})^3 + \gamma_{vy} + \varepsilon_{ivy} \quad (3)$$

$$\begin{aligned}\text{Log purchase price}_{ivy} = B_0 &+ B_1(\text{car age}_{ivy}) + B_2[I(3<=\text{car age}_{ivy}<=6)*(\text{car age}_{ivy}-3)] \\ &+ B_3[I(7<=\text{car age}_{ivy}<=10)*(\text{car age}_{ivy}-7)] \\ &+ B_4[I(\text{car age}_{ivy}>=11)*(\text{car age}_{ivy}-11)] + \gamma_{vy} + \varepsilon_{ivy} \quad (4)\end{aligned}$$

$$\text{Log purchase price}_{ivy} = B_0 + B_1(\text{car age}_{ivy}) + B_2(\text{car age}_{ivy})*(\text{make\_dummies}_{ivy}) + \gamma_{vy} + \varepsilon_{ivy} \quad (5)$$

$$\begin{aligned}\text{Log purchase price}_{ivy} = B_0 &+ B_1(\text{car age}_{ivy}) + B_2(\text{car age}_{ivy})*(\text{decade\_of\_model\_year\_dummies}_{ivy}) \\ &+ \gamma_{vy} + \varepsilon_{ivy} \quad (6)\end{aligned}$$

V. Calculating Vehicle Flows

Using Equation (2), we calculate an annual depreciation rate (δ): $\delta = 1 - \text{EXP}(B_1)$. The calculation of vehicle flows (vflow) will differ across samples:

1) <u>Estimation Sample</u>: vflow = (Reported Purchase Price)*δ.
2) <u>Other Vehicles with Reported Purchase Price</u>: First we convert reported purchase price to survey year dollars. Then, vflow = (Real Value of Reported Purchase Price)*$\delta(1-\delta)^t$, where t is the number of years since the car was purchased.
3) <u>Prediction Samples 1-3</u>: vflow = (Predicted Purchase Price)*δ.

A similar approach can be implemented to calculate vehicle flows using depreciation rate estimates derived from the alternative specifications in Equations (3) – (6).

A total flow from vehicles for each consumer unit-quarter observation is then calculated by summing vflow across all vehicles within a consumer unit-quarter.

DESCRIPTION FROM MEYER AND SULLIVAN (2009)

Appendix D: Estimating Vehicle Service Flows

Our measure of consumption replaces the purchase price of vehicles and vehicle maintenance costs with the service flow value from owned vehicles. We improve upon previous studies in how we calculate a flow that reflects the value that a consumer receives from owning a car during the period. Previous studies have imputed flows based only on recent spending on vehicles and descriptive characteristics of the family (Cutler and Katz 1991), recent spending on vehicles, vehicle age, and descriptive characteristics of the family (Meyer and Sullivan 2003, 2004), or reported purchase prices and vehicle age (Slesnick 1993). Our approach provides two important improvements upon previous work. First, in addition to vehicle age, our approach uses detailed information for each vehicle (such as make, model, year, automatic transmission, and other characteristics) to determine the market price. Second, we estimate depreciation rates by comparing the reported purchase prices for similar vehicles of different ages. We use the detailed expenditure data for owned vehicles from the 1980-2009 CE. A detailed explanation of the procedure used to estimate these service flows along with the data are available at www.nd.edu/~jsulliv4/vehicles.

D.1 Calculating the Market Price of a Vehicle

We determine a current market price for each of the 1.3 million vehicles in the data from 1980-2009 in one of three ways. First, for vehicles that were purchased within twelve months of the interview and that have a reported purchase price (the estimation sample), we take the current market price to be the reported purchase price. This estimation sample accounts for about 14 percent of all vehicles in the 1980-2009 surveys. Second, for vehicles that were purchased more than twelve months prior to the interview and that have a reported purchase price (about 15 percent of all vehicles), we specify the current market price as a function of the reported purchase price and an estimated depreciation rate as explained below.

Finally, for the remaining 71 percent of vehicles, we impute a current market price because the purchase price is not reported. Using the estimation sample, we regress the log real purchase price on a cubic in vehicle age, vehicle characteristics, family characteristics, and make-model-year fixed effects.[1] The vehicle characteristics include indicators for whether the vehicle has automatic transmission, power brakes, power steering, air conditioning, a diesel engine, a sunroof, four-wheel drive, or is turbo charged. Family characteristics include log real expenditures (excluding vehicles and health), family size, region, and the age and education of the family head. Coefficient estimates from this regression are then used to calculate a predicted log real purchase price for the i^{th} vehicle ($x_i\hat{\beta}$). The predicted current market value for each vehicle without a reported purchase price is then equal to $\hat{\alpha} * \exp(x_i\hat{\beta})$, where $\hat{\alpha}$ is the coefficient on $\exp(x_i\hat{\beta})$ in a regression of y_i on $\exp(x_i\hat{\beta})$ without a constant term.[2]

D.2 Estimating a Depreciation Rate and Service Flows

To estimate a depreciation rate for vehicles, we compare prices across vehicles of different age, but with the same make, model, and year. In particular, from the estimation sample we construct a subsample of vehicles that are in a make-model-year cell with at least two vehicles that are not the same age. Using this sample, we regress the log real purchase price of

---

[1] 75 percent of the vehicles without a reported purchase price can be matched to at least one vehicle in the estimation sample with the same make, model, and year. About three-fourths of the remaining 25 percent do not have a match because they are not a car, truck, or van so make and model are not observed. Starting in 2006, vehicles can be matched on make, but not model, because the CE stopped providing information on vehicle model after 2005. For those vehicles without a reported purchase price that do not have the same make, model, and year as at least one vehicle in the estimation sample, but do have the same make and year as a vehicle in the estimation sample, a separate regression is estimated that includes make-year fixed effects instead of make-model-year fixed effects.

[2] This adjustment is made because $\exp(x_i\hat{\beta})$ will tend to underestimate y_i.

the vehicle on vehicle age and make-model-year fixed effects.[3] From the coefficient on vehicle age (β), we calculate the depreciation rate (δ): $\delta = 1 - EXP(\beta)$. The service flow is then the product of this depreciation rate and the current market price. If the vehicle has a reported purchase price but was not purchased within 12 months of the interview we calculate the service flow as: (real reported purchase price)*$\delta(1-\delta)^t$, where t is the number of years since the car was purchased.

Although the 1972-1973 CE data files include an inventory of vehicles owned, we do not use these data to calculate service flows from vehicles for several reasons. First, we do not observe the year the car was manufactured, only whether it was manufactured before or after 1967. Second, we do not observe the model for vehicles manufactured during or before 1967, and for those manufactured after 1967 we only observe a broadly defined model group: subcompact domestic, compact domestic, etc. Thus, rather than using the vehicle inventory data, we impute service flows for owned automobiles using data on reported spending on new and used automobile purchases during the survey year and the reported number of automobiles owned during the year. Specifically, for a sample with positive spending on automobiles, we regress annual spending for new and used automobiles on a quadratic in total (non-automobile) spending and observable characteristics of the family including family income, family size, and the age, sex, and education of the family head. Parameter estimates from these regressions are used to predict spending on new and used car purchases for all families that own automobiles. We calculate the service flow from automobiles as the product of predicted automobile spending, the number of owned automobiles and a depreciation rate. This approach will understate total automobile flows for some families because the number of automobiles is topcoded at 2. This approach will overstate vehicle flows for families that dispose of an automobile during the survey year if this automobile is included in the total count of automobiles owned. This approach will also overstate vehicle flows for families that have owned their vehicles for an extended time, because we are predicting the value based on recent automobile purchases. Note that unlike our approach for 1980-2009, we calculate service flows only for automobiles, not for other vehicles such as trucks, motorcycles, campers, etc., because we do not have reliable information on the total number of each of these types of vehicles owned.

D.3 Validation

We validate our procedure for predicting the current market value of vehicles for those observations where we do not have a purchase price by comparing the predicted values to published values in National Automobile Dealers Association (NADA) guides. For a given year of the CE we take a random sample of 100 vehicles for which a purchase price was not observed. We then find the average retail price of the vehicle reported in the NADA Official Used Car Guide, using observable vehicle characteristics including make, model, year, number of cylinders, and number of doors. In cases where a unique match is not found in the NADA guide (for example, there might be multiple sub-models listed in the NADA guide), we use the midpoint of the range of prices for the vehicles that match the description of the vehicle from the

[3] The distribution of service flows does not differ noticeably when alternative specifications for depreciation are estimated. For example, specifications that allow the depreciation rate to vary by age of the vehicle (by including a cubic in vehicle age in the regression) yield similar results.

CE. For the sample of vehicles randomly drawn from the 2000 CE, the correlation between our imputed price and the 2000 NADA price was 0.88. Similarly, for a sample of 100 cars with a reported purchase price, the correlation between the reported price and the NADA price was 0.91.