A Data Appendix

Given the comprehensiveness and richness of the dataset used in this study, we include this appendix to describe the data in more detail and elaborate on the sample selection.

A.1 Development of the Final Dataset

As a general rule, we focused our data cleaning efforts on avoiding dropping observations to maximize our dataset coverage. Our data cleaning proceeded as follows. First, we converted the dataset into one where the observation is a driving period and all negative driving observations or other observations with critical missing data are dropped. This dataset has 10,994,333 observations. We further restrict the data to have driving periods that began after July 1, 1998 (since tests were not mandatory prior to this date) and before January 1, 2008 (this ensures that our sample is not biased away from new vehicles). These restrictions leave us with 7,254,893 observations. After this, we delete observations where the length of the driving period, which we call *years to test*, is not either between 1 and 2.5 years or between 3.5 and 4.5 years. This is chosen to balance not getting too many observations with unexplained lengths of the driving periods while also accounting for early phase-in, which led to a number of 1-year periods in 1999 and 2000. This leaves us with 6,877,185 driving periods. We have missing demographic variables for 277,294 observations, which brings us to 6,599,891 observations. We drop 178 outlier observations with VKT greater than 10,000 km/day. Finally, because we use household fixed effects, we drop 744,267 households that are only observed once in the dataset. This brings us to our final sample size of 5,855,446.

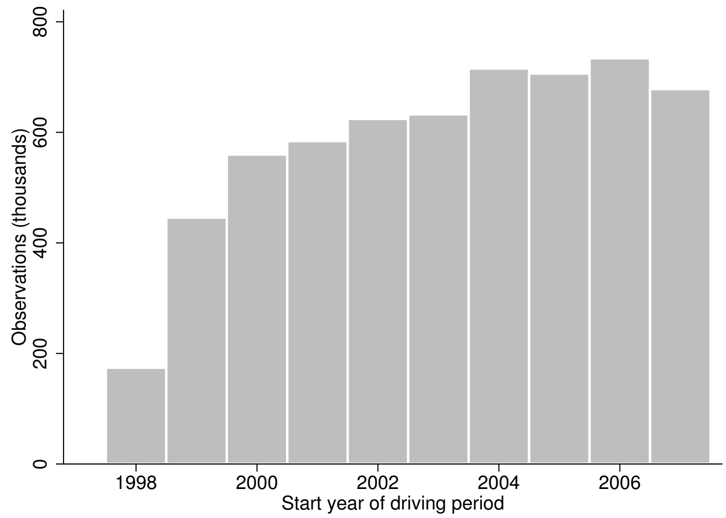
To clarify how these observations are distributed over time, Table 10 shows a histogram of the start year of the driving period. The low number in the first year is due to the sample selection criterion keeping only periods starting after July 1, 1998.

The following sections provide more detail on the sources and cleaning of the data.

A.1.1 Car Ownership

The data we use on car ownership comes from the Danish Central Motor Register. This register contains the license plate, vehicle identification number (VIN), and personal identification number (i.e., CPR numbers, which allow us to merge these data in with other public registers). In the raw data, we observe some problematic observations. When we observe a car with a car ownership period for one owner that does not end and a car ownership period for a different owner at a later date, we know that the transaction was not properly recorded. In this case, we assign the ending of the ownership period for the first owner at the date when the second owner is first observed with the vehicle. We also do the same for the

Figure 10: Observations by start year



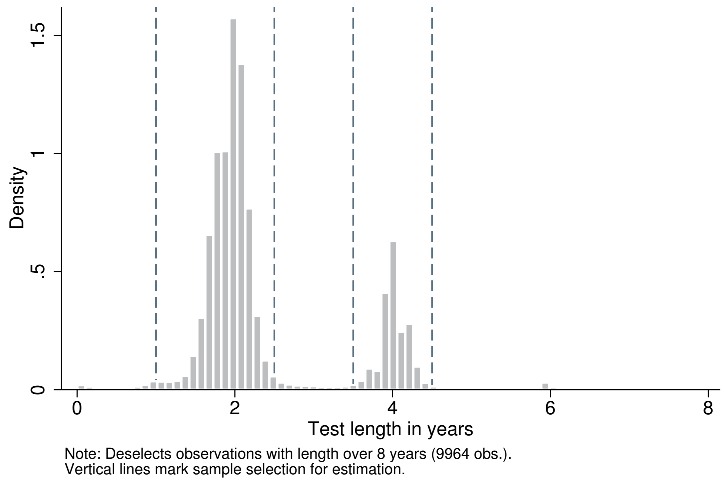
reverse scenario. We also occasionally see problematic observations where there is an overlap of owners. In that case, we have no way of discerning which person truly owns the car and according to the data documentation such an observation should be impossible so we drop them from the dataset.

A.1.2 Driving Periods

The data on driving periods come from the Ministry of Transportation (MOT) tests that were introduced in 1997. These inspections are mandatory and must be performed at car ages 4, 6, 8, 10, 12, etc. This means that we have two types of driving periods; The *first* driving period is 4 years long (that is, it has *4 years to test*) and any *subsequent* driving period will be only 2 years long. The inspection date is set based on the date of the first registration of the car in Denmark. In practice, the years to test may deviate with plus or minus three months around these designated years. A person may choose to take the car in for inspection *earlier* than the set date if he or she wishes.

MOT tests were originally performed by public authorities directly but in more recent years, they have been performed by private companies approved by the MOT. The goal of the test is to verify taht the car is in safe condition for driving on the roads. As a part of the test, the odometer of the car is recorded. A test may have four outcomes; 1) The car can be *approved*. 2) The car can be *conditionally approved*, meaning that certain repairs must be performed for the car to be in legal driving order but that no extra test will be required. 3) The car can be *approved after a re-inspection*, implying that repairs must be made and then the car must return for another test before 33 calendar days. Finally, 4) the car can be declared *not approved* in which case it will be illegal to drive the car and the police will withdraw the license plates. Some drivers may take their vehicle in for an inspection early prior to selling the car in order to give the buyer a signal that the car is in proper working order. Figure 11 shows the distribution of the driving period length. The vertical lines mark the sample selection described above.

Figure 11: Years to Test Distribution



The sample selection criteria mentioned above for the timing of the driving periods can also be seen in Figure 12. We have selected the sample for a period when the years to test is relatively constant, thus helping to alleviate any concerns of sample selection bias based on this variable.

A.2 Detailed Variable Description

Table 6 lists of all the variables used in this paper with details on each.

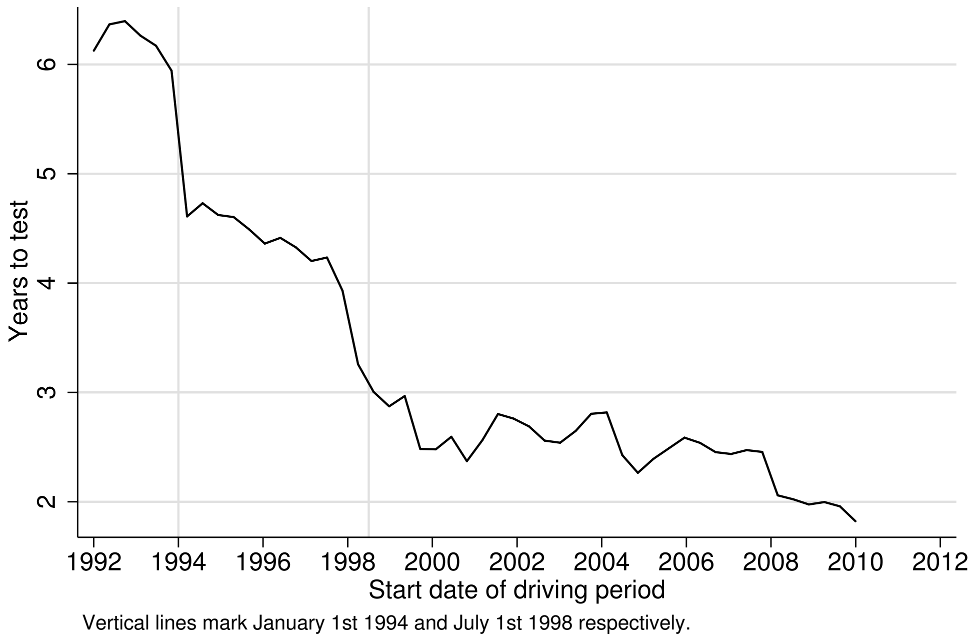
Table 6: Variables used in the paper

Variable	Description
VKT	Vehicle-kilometers-traveled in km per day. The variable is constructed by taking first-differences of the odometer readings from the dataset with vehicle inspections. For the first inspection we observe for a car, we assume that the odometer was zero at the time of the car's first registration in Denmark. This will be incorrect if the car was imported from abroad. However, then the car must have had a toll inspection, which we observe, so we can run a robustness check on this assumption. We find that this does not impact our results.
Couple	Dummy for there being two members of the household (married or co-habiting, of opposite genders and having at less than 15 years of age difference).

Real gross income	The sum of gross incomes for the member(s) of the household. The variable comes from the income tax registers. The variable includes all government transfers such as pension payments, unemployment benefits, etc.
Real gross income (couples)	As above but equal to zero for singles.
Real gross income (singles)	As above but equal to zero for couples.
WD	Work distance. The variable is based on the Danish deduction for work distance. Any working household having further than 12 km each way to work can deduct a fixed amount per km. Thus, the measure will be equal to zero if the individual lives closer than 12 km from his or her work. Between 12 and 25 km, there is a rate and above 25 km, the rate drops to half. The rate changes over the period. The total deduction is the daily rate times the number of days worked. The variable is self-reported but the tax authorities have access to both the home and work addresses for the individual. The deduction is the rate times the distance times the number of days worked. We do not observe the number of days worked so we assume 225 work days, which corresponds to the number of days in a typical Danish work year. For example, the official number of work days were 224 in 2007, 226 in 2008, 225 in 2009 and 228 in 2010. Most unions follows these, as do most public sector employees. Figure 16 shows the density of the work distance variable. Note that there is a positive mass on the interval $(0; 12)$ km even though the deduction is only given if the actual work distance is above 12 km; this is due to the assumption about 225 work days per year. If an individual works part-time, say 110 days, but has a distance of 20 km to work, then the variable will be equal to 10. The positive mass will therefore contain many part time employees. For validity, we can compare it to the continuous WD measure, available for a subset of the period (see Appendix A.3.3).
WD non-zero	Dummy for the WD measure being observed. Thus, this is essentially a dummy for the individual living further than 12 km from the work place.
WD (actual distance)	This is the actual distance from home to work. The variable comes from the Danish Technical University's Department of Transportation. It is calculated using a shortest-path algorithm and the National Transport model with GIS data on households and their work places. The variable is only observed for households where the work place is observed and not for 1998 or 1999. In total, it is observed for 76.17% of our estimation sample (79.61% of the observations between 2000 and 2008). We use this measure to validate the tax-based WD variable.
# of children	The number of children aged less than 18 years living with the household.
Urban (dummy)	Dummy equal to one if the household lives in either Copenhagen, Frederiksberg, Aarhus, Aalborg og Odense municipalities, which constitute the major Danish urban areas.
Company car	Dummy equal to one if at least one member of the household has paid the tax penalty for having access to a company car. The use of company cars is restricted to avoid making it an alternative to buying your own car privately. The size of the tax depends on the value of the car. We collapse the variable to a dummy for having any car available to any of the members of the household. Individuals may have access to a company car and not pay this tax if the car is a "yellow license plate" car. These cars can have at most two seats and are typically vans used by craftsmen. The police enforce this very strictly and an individual caught using a company car privately and not paying the penalty is fined and some times forced to pay the registration tax.
Self employed	Dummy equal to one if the household has at least one self employed individual. This information comes from the tax registers.
# of periods observed	The number of driving periods observed for the household. Note that the other driving periods may be with different cars and that our sample selects only households with at least two driving periods.

Bus/Train stops per km^2	The number of public transport stops in the municipality in 2013 divided by the area of the municipality of residence at the start of the driving period in km^2. The data for this comes from the Travel Planner (`http://rejseplanen.dk`), which is a search engine for planning trips using public transportation. The data are only available for a cross-section in 2013. The highest number of stops is 79.9 stops per km^2 for Odense municipality and the lowest is Aaskov municipality with 0.3 stops per km^2.
Weight (ton)	The gross weight of the car in metric tonnes. This is the maximum allowed weight of the vehicle including cargo. The variable comes from the vehicle type approval documents.
Diesel	Dummy equal to one if the car uses diesel fuel. Note that the fuel price will then be based on the diesel price.
Van	Dummy equal to one if the vehicle is a van.
Percent owned of period	The fraction of the driving period where the car was owned by this household. That is, if the driving period starts on Jan 1st, 2001 and ends on Jan 1st 2003, but the car changed owner on Jan 1st 2002, this variable will be equal to 0.5 for both the observations of the two households driving the car.
Driving period length	The length of the driving period in years. For new cars, this will be 4 years and for older cars, it will be 2 years, both plus or minus 3 months and with some exceptions. Note that our sample selects on driving periods being either 1.0 to 2.5 years long or 3.5 to 4.5 years long.
Car age	Car age in years at the start of the driving period. Car age is defined as the time since the car's first registration in Denmark since we do not observe the actual production year of the vehicle. This will be very close to the number of years since the model year for most vehicles, but will be off for the small number of imported vehicles.
# cars / vans / motorcycles / mopeds / trailers owned	Continuous measure of the number of vehicles of the given type owned by the household. For example, if for a given household i and driving period t, the household owns another car for the entire duration of the period, then # of cars owned will be 2.0. If that other car is only purchased half-way through the driving period t, then it is equal to 1.5. That is, the variable is equal to the fraction of this driving period overlapping with the ownership of other vehicles.
First driving period	Dummy equal to one if it is the car's first driving period, i.e., the driving period's start date is equal to the first registration date of the car.
Fraction owned	For household i and driving period t, this is the percent of the driving period where houseohld i is the owner. That is, if the car changes owner midway through, there will be an observation in the dataset for each of the two households owning the car and they will both have this variable set to 0.5.
Years to test	The length of the driving period in years (continuous variable). Due to our sample selection, this will be in $[1.0; 2.5]$ or in $[3.5; 4.5]$.
% of each month	This is a set of variables for each month equal to the % of the driving period taking place in each of the 12 months. Thus, if a driving period is precisely 2 or 4 years long, these will all be equal to $\frac{1}{12}$. We omit April as the reference group in regressions since the fractions will always sum to 1.
Year controls	These are variables for each year, 1998, ..., 2011, each equal to the % of the driving period falling in that year. In the preferred specification, we exclude year 2003 as the reference year and include an additional full set of year controls interacted with the diesel dummy to allow a separate time trend for diesels.

Figure 12: Years to Test by Start Date of the Driving Period



A.3 Additional Descriptives

A.3.1 Driving and Demographics

Figure 13 shows the distribution of vehicles kilometers traveled (VKT). The figure is cut at 200 km/day for clarity. Note that there is still positive mass for very low VKT. This may be explained by vehicles such as vintage or specialty cars.

A.3.2 Additional Spatial Descriptives

Figure 14 shows the number of observations (i.e., driving periods) by municipality. The four major urban areas clearly stand out: Copenhagen (east), Odense (center, on the island of Fyn), Aarhus (midway up on the eastern side of Jutland) and Aalborg (Northern part of Jutland).

Figure 15 shows a map of Denmark where municipalities are colored by the average work distance of the households. We see that the households with high work distances tend to be in the outskirts of the major urban areas with a few exceptions. Note that this figure plots observations in the estimation sample, so it should be interpreted recognizing that it conditions on households owning a car.

Figure 13: The Distribution of Vehicle Kilometers Traveled

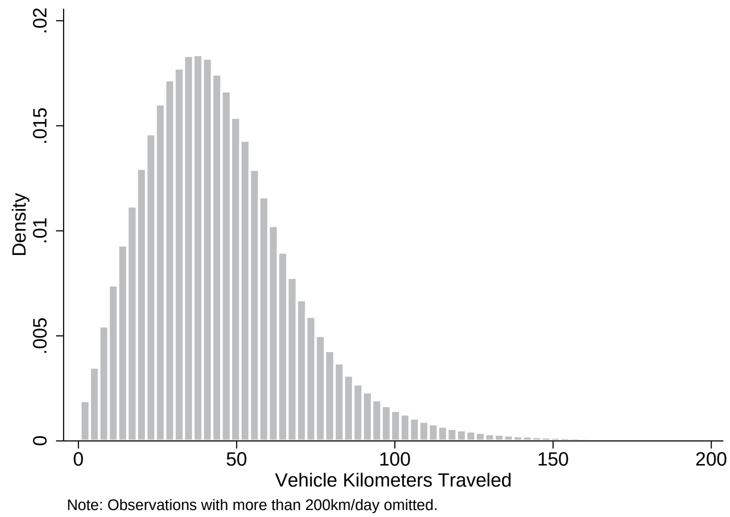


Figure 14: Observations in the Estimation Sample by Municipality

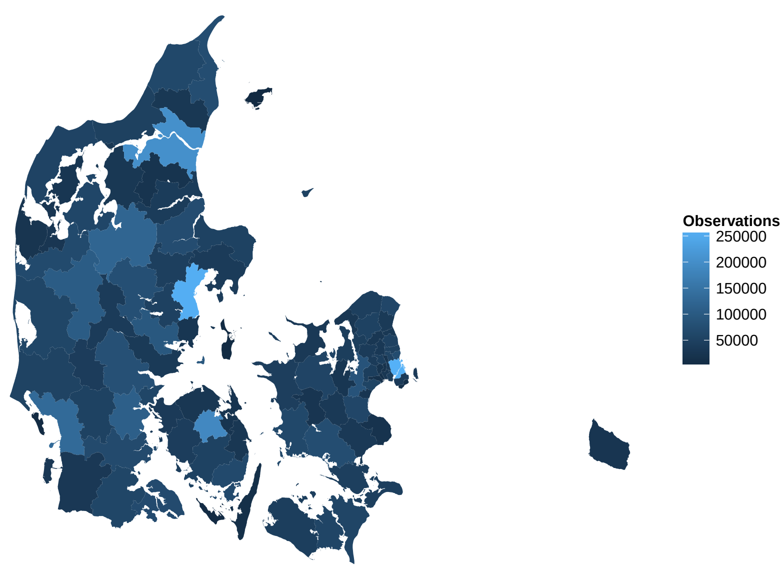


Figure 15: Average Work Distance by Municipality

A.3.3 Work Distance

In this subsection, we discuss the validity of the work distance variable. Table 7 shows summary statistics for work distances of males, females and singles. It shows both the measure based on the tax deduction for work distance as well as the "actual work distance" variable, which measures the distance using GPS coordinates. The tax deduction is a deduction from taxable income and it is given as a fixed amount per kilometer per day but is equal to zero if the distance is shorter than 12 km. The number of days worked is not observed so we assume that all individuals work 225 days a year, which is very common in Denmark. Hence, if the individual actually worked fewer days, we will be undershooting the measure (which explains why the variable can take values below 12 km) and vice versa. The per km rate varies over time and there is a kink in the schedule at 50 km where it falls to half the rate.[29]

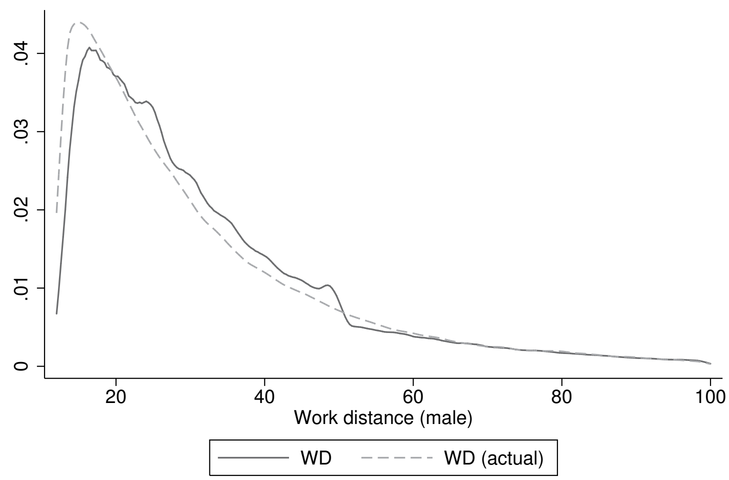
To explore the validity of the work distance variable, we exploit the aforementioned *actual* work distance, which is based on the address of the home and work location. In that sense, it is a better variable for measuring commuting and driving than the work distance reported on the tax returns. However, it is not available for the full sample. We compare the distribution

[29] In some years, a small number of *fringe municipalities* (Danish: *udkantskommuner*) also had the full rate after the 50 km threshold.

Table 7: Work Distance (WD) Variables

	count	mean	sd	p1	p10	p25	p50	p75	p90	p95	p99
WD, male	4550411	9.5932	18.51	0.0	0.0	0.0	0.0	15.5	32.2	44.7	80.7
WD, female	4550411	6.9385	13.77	0.0	0.0	0.0	0.0	10.7	24.8	33.7	58.3
WD, single	1305035	7.6966	16.87	0.0	0.0	0.0	0.0	9.0	27.5	39.3	75.1
WD non-zero, male	4550411	0.3493	0.48	0	0	0	0	1	1	1	1
WD non-zero, female	4550411	0.3137	0.46	0	0	0	0	1	1	1	1
WD non-zero, single	1305035	0.2917	0.45	0	0	0	0	1	1	1	1
Actual WD, male	3343884	20.3157	34.36	0.0	0.0	2.7	9.8	23.7	46.5	71.8	196.3
Actual WD, female	3094025	14.3657	22.45	0.0	0.6	2.8	8.1	18.1	32.1	45.0	99.3
Actual WD, single	813453	18.6009	32.87	0.0	0.1	2.6	8.6	21.1	42.0	66.1	183.4

Figure 16: Comparing the Two Work Distance Measures



Selection: males' work distance should be in [12; 100] km with both measures (1013011 obs.).
WD is based on the tax deduction and WD (actual) on GPS coordinates.

of driving according to the two variables to validate the measure. To make the comparison sensible, make the comparison for the subsample where both measures fall in the range [12 km ; 100 km]. The lower bound ensures that the tax-based measure is also observed, while the upper bound makes the graph easier to read. Figure 16 shows the comparison, demonstrating the comparability of the two work distance variables.

B Additional Regression Results

This appendix contains number of econometric results supplementing the primary results from section 5. To begin, Table 8 shows the coefficients pertaining to car characteristics and the driving period that were suppressed in the primary results table in our paper. Table 9 shows the fuel price elasticities used in figure 6.

Table 10 shows the coefficients for the demographic variables for the quantiles 1, 50 and 99 in the panel quantile regression estimates. They show that many of the coefficients do not vary over the conditional distribution of VKT. However, the fuel price elasticity, work distance, company car dummy, and transit stop density variables change.

For the results with interactions, not all the estimated coefficients were shown in Table 4 in the main results section. Table 11 shows some of the omitted coefficients, namely the ones pertaining to car characteristics.

Table 8: Main results — Car and Period Controls

	OLS		Household FE	
	(1)	(2)	(3)	(4)
	No demo	Base	FE	Main
$\log p^{\text{fuel}}$	-0.866***	-0.298***	-0.515***	-0.304***
	(0.00509)	(0.0143)	(0.00722)	(0.0154)
New car	-0.00350*	0.0128***	0.00838***	0.0394***
	(0.00148)	(0.00148)	(0.00160)	(0.00164)
Percent owned of period	-0.189***	-0.112***	-0.0537***	-0.0154***
	(0.000826)	(0.000862)	(0.00106)	(0.00110)
Driving period length	-0.0507***	-0.0541***	-0.0465***	-0.0242***
	(0.000634)	(0.000645)	(0.000681)	(0.000725)
Weight (ton)	0.00214***	0.00169***	0.00166***	0.00167***
	(0.00000523)	(0.00000506)	(0.00000799)	(0.00000798)
Weight squared	-0.000000471***	-0.000000369***	-0.000000354***	-0.000000354***
	(1.35e-09)	(1.30e-09)	(2.00e-09)	(2.00e-09)
Diesel	0.316***	0.311***	0.228***	0.259***
	(0.000918)	(0.00557)	(0.00139)	(0.00545)
Van	-0.236***	-0.199***	-0.204***	-0.205***
	(0.00117)	(0.00115)	(0.00171)	(0.00170)
Car age	-0.0302***	-0.0275***	-0.0284***	-0.0293***
	(0.0000932)	(0.0000911)	(0.000140)	(0.000141)
# cars owned	0.0482***	-0.0202***	-0.0581***	-0.0501***
	(0.000593)	(0.000759)	(0.00114)	(0.00109)
# vans owned	0.0111***	-0.0470***	-0.0711***	-0.0654***
	(0.00124)	(0.00122)	(0.00183)	(0.00179)
# motorcycles owned	0.0319***	-0.00420***	0.0102***	0.0118***
	(0.00101)	(0.000905)	(0.00178)	(0.00178)
# mopeds owned	0.136***	0.0415***	0.0232***	0.0204***
	(0.00138)	(0.00131)	(0.00218)	(0.00217)
# trailers owned	0.0123***	0.0258***	0.00334**	0.00595***
	(0.000519)	(0.000983)	(0.00106)	(0.00106)
Year controls	No	Yes	No	Yes
Household FE	No	No	Yes	Yes
R^2	0.20	0.34	0.18	0.18
N	5,855,446	5,855,446	5,855,446	5,855,446

Dependent variable is the log VKT. An observation is a driving period. All specifications have all of the other variables and controls in Table 8. Robust standard errors clustered at the household level in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 9: Fuel Price Elasticity by Conditional Quantile

Quantile	$\log p^{\text{fuel}}$	std error
99	-0.609***	(0.0592)
95	-0.369***	(0.0223)
90	-0.289***	(0.0145)
85	-0.256***	(0.0114)
80	-0.238***	(0.00988)
75	-0.234***	(0.00890)
70	-0.228***	(0.00831)
65	-0.230***	(0.00796)
60	-0.231***	(0.00758)
55	-0.233***	(0.00742)
50	-0.233***	(0.00739)
45	-0.244***	(0.00755)
40	-0.250***	(0.00779)
35	-0.267***	(0.00808)
30	-0.284***	(0.00868)
25	-0.309***	(0.00942)
20	-0.330***	(0.0106)
15	-0.357***	(0.0124)
10	-0.402***	(0.0156)
5	-0.490***	(0.0233)
1	-0.559***	(0.0663)
For all quantile regressions:		
Demographics		Yes
Year controls		Yes
% of each month		Yes
Car		Yes
Period		Yes
Household FE (Canay, 2011)		Yes
N		5,855,446

Table 10: Panel Quantile Regression for P01, P50 and P99: Demographics

	(1) Linear	(2) P01	(3) P50	(4) P99
$\log p^{\text{fuel}}$	-0.304***	-0.559***	-0.233***	-0.609***
	(0.0154)	(0.0663)	(0.00739)	(0.0592)
Work Distance (WD) controls				
WD, male	0.00242***	0.00137***	0.00260***	0.00347***
	(0.0000336)	(0.000106)	(0.0000118)	(0.0000947)
WD non-zero, male	0.0329***	0.0797***	0.0320***	-0.00688
	(0.00107)	(0.00427)	(0.000476)	(0.00382)
WD, female	0.00303***	0.00245***	0.00328***	0.00338***
	(0.0000443)	(0.000151)	(0.0000168)	(0.000135)
WD non-zero, female	0.0257***	0.0950***	0.0247***	-0.0327***
	(0.00111)	(0.00461)	(0.000513)	(0.00412)
WD, single	0.00419***	0.00360***	0.00448***	0.00562***
	(0.0000835)	(0.000216)	(0.0000241)	(0.000193)
WD non-zero, single	0.0724***	0.138***	0.0713***	-0.0174*
	(0.00243)	(0.00832)	(0.000927)	(0.00743)
Age controls				
Age, male	0.0212**	0.0224***	0.0213***	0.0199***
	(0.00813)	(0.00133)	(0.000148)	(0.00119)
Age, female	0.0468***	0.0534***	0.0469***	0.0403***
	(0.00813)	(0.00132)	(0.000148)	(0.00118)
Age, single	0.0598***	0.0631***	0.0604***	0.0549***
	(0.000939)	(0.000971)	(0.000108)	(0.000868)
Age squared, male	-0.0000930***	-0.000118***	-0.0000943***	-0.0000705***
	(0.0000112)	(0.0000128)	(0.00000143)	(0.0000115)
Age squared, female	-0.000195***	-0.000275***	-0.000197***	-0.000117***
	(0.0000115)	(0.0000134)	(0.00000149)	(0.0000120)
Age squared, single	-0.000206***	-0.000275***	-0.000213***	-0.000119***
	(0.00000767)	(0.00000933)	(0.00000104)	(0.00000834)
Other demographic controls				
log gross inc (couple)	-0.0242***	-0.0156***	-0.0176***	-0.0278***
	(0.00162)	(0.00304)	(0.000339)	(0.00272)
log gross inc (single)	0.0200***	0.0268***	0.0195***	0.0144***
	(0.00288)	(0.00226)	(0.000252)	(0.00202)
Urban (dummy)	-0.0249***	-0.0392***	-0.0254***	-0.0146**
	(0.00284)	(0.00519)	(0.000578)	(0.00463)
# of kids	-0.0168***	-0.0145***	-0.0169***	-0.0145***
	(0.000650)	(0.00146)	(0.000163)	(0.00130)
Company car	-0.0977***	-0.312***	-0.102***	0.0601***
	(0.00216)	(0.00695)	(0.000775)	(0.00621)
Self employed	0.000712	-0.0818***	0.00334***	0.0694***
	(0.00136)	(0.00426)	(0.000474)	(0.00380)
Bus/Train stops per km^2	0.0000419	-0.0000421	0.0000173	0.000300**
	(0.0000548)	(0.000103)	(0.0000114)	(0.0000916)
Year controls	Yes	Yes	Yes	Yes
% of each month	Yes	Yes	Yes	Yes
Car	Yes	Yes	Yes	Yes
Period	Yes	Yes	Yes	Yes
Linear Fixed Effects (FE)	Yes	No	No	No
Canay (2011) FE	No	Yes	Yes	Yes
N	5855446	5855446	5855446	5855446

Standard errors in parentheses. FE are at the household level.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Additional Coefficients for the Estimations in Table 4

	(1)	(2)	(3)
Mean elasticity	-0.253	-0.288	-0.238
$\log p^{\text{fuel}}$	-0.879***	-3.847***	-4.698***
	(0.240)	(0.0862)	(0.283)
Weight (ton) $\times \log p^{\text{fuel}}$		0.00316***	0.00299***
		(0.0000780)	(0.0000820)
Weight squared $\times \log p^{\text{fuel}}$		-0.000000633***	-0.000000603***
		(1.90e-08)	(1.98e-08)
Diesel=1 $\times \log p^{\text{fuel}}$		-0.389***	-0.439***
		(0.0321)	(0.0325)
Van=1 $\times \log p^{\text{fuel}}$		0.346***	0.413***
		(0.0185)	(0.0200)
Car age $\times \log p^{\text{fuel}}$		0.0318***	0.0295***
		(0.00104)	(0.00111)
# cars owned $\times \log p^{\text{fuel}}$		-0.232***	-0.238***
		(0.0339)	(0.0378)
# vans owned $\times \log p^{\text{fuel}}$		-0.189***	-0.177***
		(0.0487)	(0.0466)
# motorcycles owned $\times \log p^{\text{fuel}}$		-0.0404*	-0.0448**
		(0.0164)	(0.0164)
# mopeds owned $\times \log p^{\text{fuel}}$		0.145***	0.0768***
		(0.0227)	(0.0227)
# trailers owned $\times \log p^{\text{fuel}}$		0.0497***	0.0399***
		(0.0135)	(0.0120)

Robust standard errors clustered on household in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C Robustness Checks

C.1 Stratifying on Time

Tables 12 and 13 shows the implications for the estimated fuel price elasticity of dropping certain years from the sample. These results demonstrate considerable robustness.

Table 12: Robustness: dropping earlier years

	(1) Full	(2) 1999-	(3) 2000-	(4) 2001-
$\log p^{\text{fuel}}$	-0.304***	-0.326***	-0.384***	-0.402***
	(0.0154)	(0.0165)	(0.0149)	(0.0153)
Year controls	Yes	Yes	Yes	Yes
% of each month	Yes	Yes	Yes	Yes
Car	Yes	Yes	Yes	Yes
Period	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes
N	5,855,446	5,681,226	5,235,440	4,675,560
R^2	0.180	0.182	0.188	0.198

Note: In each column (2)–(4), data before year 97, 98, 99 are dropped respectively.
Robust standard errors clustered on household in parantheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: Robustness: dropping later years

	(1) Full	(2) -2006	(3) -2005	(4) -2004
$\log p^{\text{fuel}}$	-0.304***	-0.258***	-0.308***	-0.279***
	(0.0154)	(0.0156)	(0.0171)	(0.0187)
Year controls	Yes	Yes	Yes	Yes
% of each month	Yes	Yes	Yes	Yes
Car	Yes	Yes	Yes	Yes
Period	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes
N	5855446	5177147	4443035	3736630
R^2	0.180	0.173	0.166	0.161

Note: In each column (2)–(4), data after year 06, 05, 04 are dropped respectively.
Robust standard errors clustered on household in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.2 Stratifying on Couples or Singles

Table 14 shows the results when estimating on the sample consisting exclusively of couples or singles, again demonstrating considerable robustness.

Table 14: Robustness: dropping couples or singles

	(1) Base	(2) Only couples	(3) Only singles
$\log p^{\text{fuel}}$	-0.304***	-0.318***	-0.250***
	(0.0154)	(0.0176)	(0.0323)
Year controls	Yes	Yes	Yes
% of each month	Yes	Yes	Yes
Car	Yes	Yes	Yes
Period	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
Household FE	Yes	Yes	Yes
R^2	0.180	0.200	0.108
N	5855446	4550410	1305036

Note: columns (2) and (3) contain only couples or singles respectively.
Robust standard errors clustered on household in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.3 Stratifying on the Length of the Period

In table 15, we drop the driving periods that have years to test (length of the driving period) more than 3 months away from either 2 or 4 years. Recall that a normal test period will be 4 years for a new car and 2 years for a used car. However, during the phase-in of the inspections, cars were summoned for inspection for the first time and therefore did not necessarily drive the normal length early on. The results show that when we remove these driving periods with non-standard lenght we find a numerically lower elasticity of -0.275. In column (2), we include a dummy to control for the non-standard length, but this does not change the fuel price elasticity much at all (-0.304).

Table 15: Robustness: length of the driving period

	(1) Base	(2) Dummy	(3) Subsample
$\log p^{\text{fuel}}$	-0.304***	-0.298***	-0.275***
	(0.0154)	(0.0154)	(0.0158)
Non-standard test length		-0.00278***	
		(0.000596)	
Year controls	Yes	Yes	Yes
% of each month	Yes	Yes	Yes
Car	Yes	Yes	Yes
Period	Yes	Yes	Yes
Demographics	Yes	Yes	Yes
Household FE			
R^2	0.180	0.180	0.192
N	5855446	5855446	4535353

Note: Standard test length: years to test is ± 3 months from either 2 or 4 years.
Elsewhere, sample selection requires VKT in [1;2.5] or [3.5;4.5] years.
Robust standard errors clustered on household in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.4 Year and Seasonality Controls

Table 16 shows the results when we change the way we control for time effects in decreasing complexity over the columns. The results show that even if we simplify down to a specification with only a linear time trend, our mean elasticity is nearly unchanged. However, if we remove time controls entirely, the elasticity changes substantially.

Table 16: Robustness: year controls

	(1) Preferred	(2) No Year	(3) No month	(4) Linear	(5) None
$\log p^{\text{fuel}}$	-0.304***	-0.309***	-0.303***	-0.313***	-0.517***
	(0.0154)	(0.0124)	(0.0123)	(0.00691)	(0.00722)
Linear time trend				-0.0414***	
				(0.000360)	
Year controls (gas)	Yes	No	No	No	No
Year controls (diesel)	Yes	No	No	No	No
% of each month	Yes	Yes	No	No	No
Car	Yes	Yes	Yes	Yes	Yes
Period	Yes	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes	Yes
Household FE	Yes	Yes	Yes	Yes	Yes
N	5855446	5855446	5855446	5855446	5855446
R^2	0.180	0.180	0.180	0.177	0.174

Col (2) has no driving year controls, Col (3) also drops month controls.

Col (4) has a linear time trend, Col (5) has no time controls.

Robust standard errors clustered on household in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In table 17, we change the main specification to use the *number* of months covered by the driving period rather than the *fraction* of each month covered (as we use in the main specification). Our mean elasticity is almost unchanged (from –0.373 to –0.372).

Table 17: Robustness: month controls

	(1) Fraction	(2) Sum
$\log p^{\text{fuel}}$	-0.304***	-0.304***
	(0.0154)	(0.0154)
Feb	-0.152***	-0.00366***
	(0.0394)	(0.000866)
Mar	-0.0973	-0.000226
	(0.0513)	(0.000826)
May	-0.0312	0.00143
	(0.0517)	(0.000852)
Jun	0.0515	0.00344***
	(0.0404)	(0.000862)
Jul	0.231***	0.00791***
	(0.0429)	(0.000890)
Aug	-0.0445	0.00114
	(0.0421)	(0.000862)
Sep	0.00780	0.00255**
	(0.0410)	(0.000820)
Oct	-0.0541	0.00105
	(0.0412)	(0.000829)
Nov	-0.141***	-0.00183*
	(0.0423)	(0.000841)
Dec	-0.174***	-0.00257**
	(0.0440)	(0.000937)
Apr		0.00199*
		(0.000851)
Year controls	Yes	Yes
Car	Yes	Yes
Period	Yes	Yes
Demographics	Yes	Yes
Household FE		Yes
N	5855446	5855446
R^2	0.180	0.180

(1): The share of the driving period falling in each month.

(2): The number of months covered by the driving period.

Robust standard errors clustered on household in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.5 Fuel Type

Table 18 explores heterogeneity in the fuel price elasticity by the fuel type of the car. Note that when we have household fixed effects, removing one or more rows will drop households entirely if they end up with one or zero remaining periods. Thus, we are removing some of the "switchers" who have responded on the extensive margin of choosing a different vehicle, which we do not model separately in this paper. We showed in a separate robustness check that this sample selection does not appreciably change the results, but it should be kept in mind in interpreting these results.

We see that allowing the elasticity to vary by fuel type results in a lower (in absolute value) mean estimate (–0.257), while the positive coefficient on the interaction of the diesel dummy and the log fuel price implies a higher elasticity for the diesel drivers (–0.392). Estimating only on the subsamples of each fuel type confirms these results, yielding a lower elasticity for gasoline drivers (–0.268) and a higher for diesel drivers (–0.541). Note that diesel cars generally cost more up-front but are cheaper to use due to a higher fuel efficiency and a lower price per litre of fuel (see e.g. Munk-Nielsen, 2015). Therefore, it is perhaps not surprising that the diesel sample appears to be more price responsive. Note also that the diesel sample is much smaller than the gasoline sample.

Table 18: Robustness: elasticity by fuel type

	(1) Base	(2) Interaction	(3) Gas only	(4) Diesel only
$\log p^{\text{fuel}}$	-0.304***	-0.257***	-0.268***	-0.541***
	(0.0154)	(0.0191)	(0.0194)	(0.0260)
Diesel=1 × $\log p^{\text{fuel}}$		-0.135***		
		(0.0279)		
Year controls	Yes	Yes	Yes	Yes
% of each month	Yes	Yes	Yes	Yes
Car	Yes	Yes	Yes	Yes
Period	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes
Household FE		Yes	Yes	Yes
R^2	0.180	0.180	0.140	0.135
N	5855446	5855446	5018019	837427

In columns 3 and 4, only a single set of time controls is included.
Robust standard errors clustered on household in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.6 Instrumental Variable Estimation

Here we present results from instrumenting for the fuel price. Our primary instrument is the WTI crude oil price in USD per barrel. The price is converted to DKK using the exchange rate from June 18, 2015 and then deflated using the Danish CPI. Figure 17 shows the oil price together with the Danish real fuel prices, illustrating the high correlation.

Figure 17: Danish Fuel Prices and the WTI Oil Price

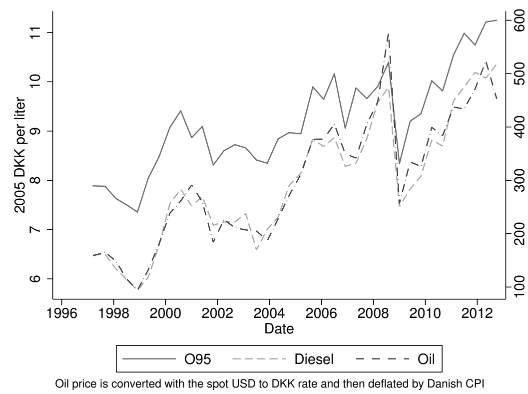


Table 19 shows the main two-stage least squares results, instrumenting log real fuel price with log real WTI oil price.

Table 19: Instrumental Variables Results

	(1) OLS	(2) FE	(3) 2SLS	(4) 2SLS FE
$\log p^{\text{fuel}}$	-0.298***	-0.304***	-0.511***	-0.368***
	(0.0143)	(0.0154)	(0.0148)	(0.0160)
Year controls	Yes	Yes	Yes	Yes
% of each month	Yes	Yes	Yes	Yes
Car	Yes	Yes	Yes	Yes
Period	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes
Observations	5855446	5855446	5855331	5855296

Robust standard errors clustered on household in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20 shows the first stage results. Note that the very high R^2 of 98% is partially due to the fact that overlapping periods are repeated. These results indicate that the log oil price is a very strong instrument. The F-statistic for both columns is well above 100.

Table 20: Instrumental Variables Results: First Stage

	(1) Simple	(2) Full
diesel	-0.973***	-0.855***
	(0.000299)	(0.000700)
log_oil	0.176***	0.177***
	(0.0000187)	(0.0000552)
diesel_log_oil	0.147***	0.124***
	(0.0000519)	(0.000129)
All controls	No	Yes
Household FE	No	No
N	5855331	5855331
R^2	0.972	0.982

Robust standard errors clustered on household in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C.7 Fuel Efficiency and Car Price

In this section, we argue why our estimate of the fuel price elasticity is not biased by our inclusion of most, but not all, vehicle characteristics. First, we show that adding the fuel economy as a control (in the subsample where the variable is observed) does not change the fuel price elasticity.

In Table 21, we show the results of our primary estimation only including fuel economy and car price (manufacturer's suggested retail price, MSRP). One major reason why these variables are not included in the main specifications is that they are only available for a subset of the period. The data source for these variables is the Danish Automobile Dealer Association (DAF). This dataset has been merged to the VINs used by the Motor Register.[30]

The results in Table 21 show how the sample where the characteristics are observed is different from the estimation sample used throughout this paper; switching to this subsample changes the fuel price elasticity from –0.30 to –0.59 (see column (2)). This can be at least partly explained by there being more households with newer cars in the subsample; from the interaction results, we saw that households who have newer cars tend to also be more price sensitive. Including the fuel efficiency variable in column (3) only very slightly changes the elasticity from -0.59 to -0.58. Further including the MSRP in column (4) leaves this almost entirely unchanged (-0.58). We take this as an indication that the included car characteristics are so highly correlated with these variables, that we have little to worry about by excluding

[30] The authors gratefully acknowledge Ismir Mulalic at DTU Transport for his assistance with this.

them.

Table 21: Robustness: controlling for fuel efficiency and MSRP

	(1)	(2)	(3)	(4)
$\log p^{\text{fuel}}$	-0.304***	-0.591***	-0.582***	-0.584***
	(0.0154)	(0.0166)	(0.0166)	(0.0166)
Fuel efficiency in km/l			-0.00249***	0.00166***
			(0.000338)	(0.000343)
price_new				0.000000478***
				(9.93e-09)
Weight (ton)	0.00167***	0.00196***	0.00193***	0.00176***
	(0.00000798)	(0.0000125)	(0.0000131)	(0.0000133)
Weight squared	-0.000000354***	-0.000000383***	-0.000000380***	-0.000000365***
	(2.00e-09)	(3.30e-09)	(3.33e-09)	(3.34e-09)
Diesel	0.259***	0.214***	0.228***	0.195***
	(0.00545)	(0.00766)	(0.00793)	(0.00787)
Van	-0.205***	-0.229***	-0.232***	-0.136***
	(0.00170)	(0.00225)	(0.00228)	(0.00278)
Car age	-0.0293***	-0.0195***	-0.0201***	-0.0178***
	(0.000141)	(0.000233)	(0.000247)	(0.000248)
# cars owned	-0.0501***	-0.0258***	-0.0259***	-0.0271***
	(0.00109)	(0.00135)	(0.00135)	(0.00137)
# vans owned	-0.0654***	-0.0798***	-0.0796***	-0.0836***
	(0.00179)	(0.00220)	(0.00220)	(0.00223)
# motorcycles owned	0.0118***	0.00889***	0.00883***	0.00859***
	(0.00178)	(0.00216)	(0.00216)	(0.00217)
# mopeds owned	0.0204***	0.0161***	0.0161***	0.0162***
	(0.00217)	(0.00272)	(0.00272)	(0.00271)
# trailers owned	0.00595***	0.00905***	0.00900***	0.00909***
	(0.00106)	(0.00128)	(0.00128)	(0.00128)
Year controls	Yes	Yes	Yes	Yes
% of each month	Yes	Yes	Yes	Yes
Period	Yes	Yes	Yes	Yes
Demographics	Yes	Yes	Yes	Yes
Household FE		Yes	Yes	Yes
R^2	0.180	0.202	0.202	0.205
N	5855446	3035301	3035301	3035301

Robust standard errors clustered on household in parentheses.

(2), (3) and (4) restricts the sample to fuel efficiency and car MSRP being observed.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

D Semi-parametric Estimation Methodology

This section provides more detail on the methodology used for the semi-parametric regression. Our semiparametric model begins with this equation:

$$\log \mathrm{VKT}_{it} = m(\log p_{it}^{\mathrm{fuel}}) + X_{it}\beta + u_{it},$$

where X_{it} is a $1 \times K - 1$ vector of household demographics and time dummies. This specification is just like the main OLS regression considered in this paper with two exceptions. First, it does not control for household-specific fixed effects. Second, it allows $\log p^{\mathrm{fuel}}$ to enter nonlinearly with out placing any functional form restrictions on $m(\cdot)$. The function $m(\cdot)$ is estimated using the double-residual method by Robinson (1988) as is described below.

D.0.1 The Robinson (1988) Double Residual Method

Consider the semiparametric regression,

$$y_i = m(z_i) + x_i'\beta + u_i,$$

where y_i, z_i, u_i are scalars, x_i is $K \times 1$ and it is assumed that $\mathbb{E}(u_i|x_i, z_i) = 0$. The estimator proposed by Robinson (1988) proceeds in three steps;

Step 1: Compute

$$\begin{aligned}\tilde{y}_i &= y_i - \hat{m}_y(z_i),\\ \tilde{x}_{ik} &= x_{ik} - \hat{m}_{x_k}(z_i),\end{aligned}$$

where the functions $\hat{m}_y(\cdot), \hat{m}_{x_k}(\cdot)$ are the orthogonalized y_i and x_{ik} respectively, defined by

$$\begin{aligned}\hat{m}_y(z) &= \sum_{i=1}^{N} w_i(z) y_i,\\ \hat{m}_{x_k}(z) &= \sum_{i=1}^{N} w_i(z) x_{ik},\\ w_i(z) &= \frac{K_h(z - z_i)}{\sum_{i=1}^{N} K_h(z - z_i)}.\end{aligned}$$

Step 2: Estimate the linear coefficients, β, as

$$\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y},$$

where $\tilde{X}$ and $\tilde{Y}$ are the the stacked versions of $\tilde{x}_i$ and $\tilde{y}_i$.

Step 3: The unknown function, $m(\cdot)$, can now be estimated as the usual nonparametric estimator of

$$\tilde{v}_i = m(x_i) + u_i,$$

where $\tilde{v}_i = y_i - x_i\hat{\beta}$.

Standard errors: The 95% CI bounds shown in the figures around $\hat{m}(\cdot)$ are found by ignoring the first-stage estimation of $\hat{\beta}$ — in other words treating $\tilde{v}$ as data.