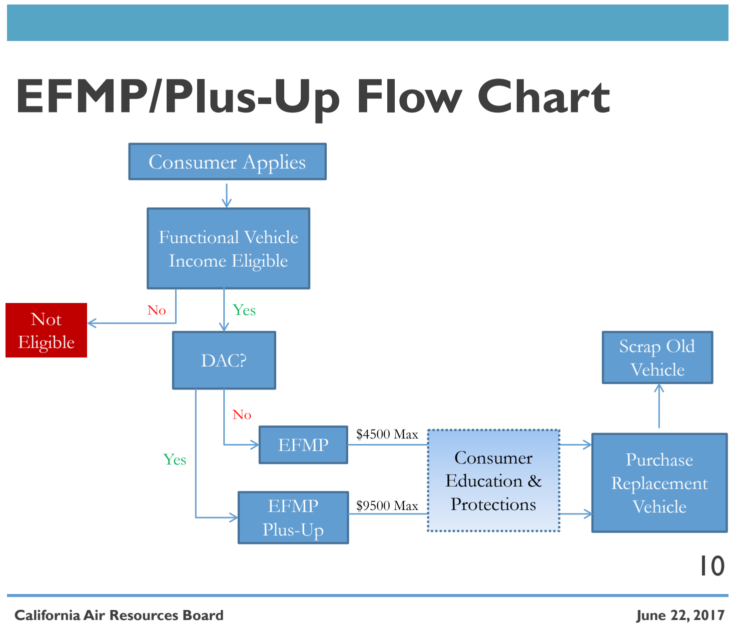
A Appendix

A.1 Supplementary Figures and Tables

Figure A1: EFMP Eligibility Flowchart



Source: https://www.arb.ca.gov/board/books/2017/062217/17-6-1pres.pdf

Figure A2: CalEnviroScreen Components

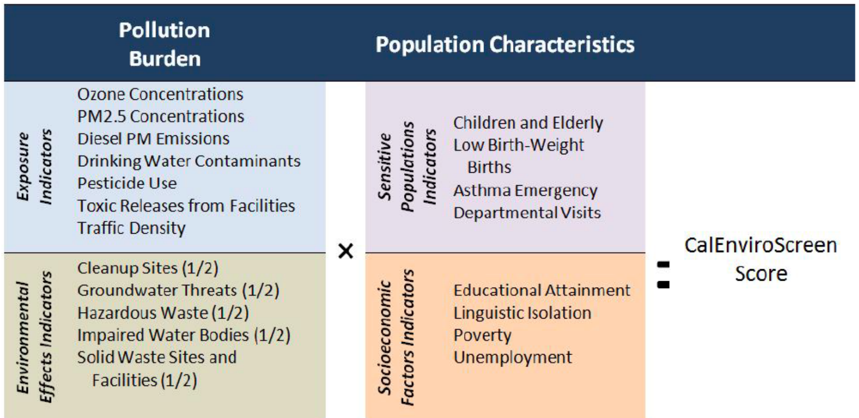
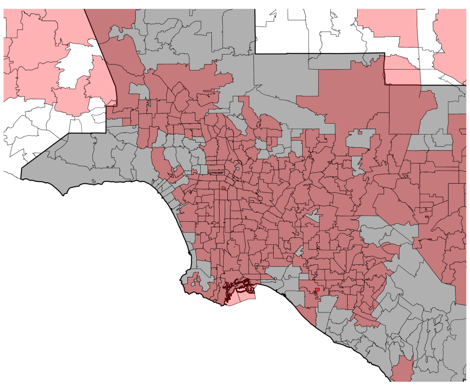
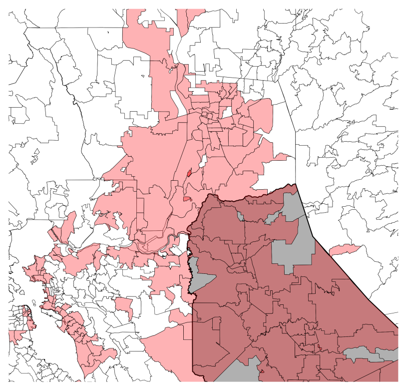


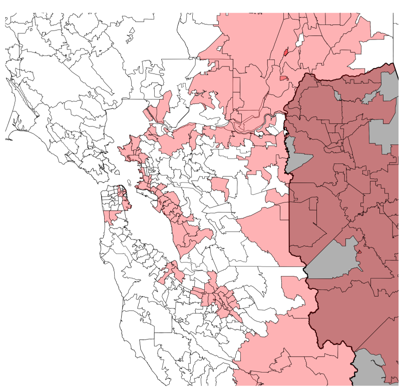
Figure A3: DACs and AQMD borders, Major Metro Areas



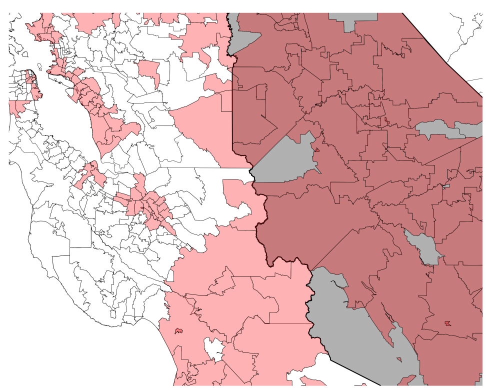
(a) Los Angeles



(b) Sacramento



(c) San Francisco



(d) San Jose

Table A1: Effects by Air District

	DepVar = Price				Dep Var = Log Q			
	(1) Unmatch	(2) Matched	(3) Unmatch	(4) Matched	(5) Unmatch	(6) Matched	(7) Unmatch	(8) Matched
% EFMP * South Coast	-2986.4	-3909.5**			0.84**	0.93**		
	(1824.6)	(1871.5)			(0.36)	(0.44)		
% EFMP * San Joaquin	-4778.0***	-5391.2***			0.70***	0.71***		
	(1278.8)	(1578.1)			(0.081)	(0.089)		
Avg. PU Subs. * South Coast			-0.62*	-0.79**			0.17**	0.19**
			(0.37)	(0.38)			(0.077)	(0.094)
Avg. PU Subs. * San Joaquin			-0.97***	-1.06***			0.14***	0.14***
			(0.26)	(0.31)			(0.016)	(0.018)
Observations	25139	22600	25139	22600	34477	27554	34477	27554
R-Squared	0.0012	0.0020	0.0012	0.0018	-0.0011	-0.00032	-0.00095	-0.00026

All specifications use the preferred set of instruments. Standard errors clustered by zip code.

A.2 Instrumental Variables

Our primary variables, EFMP-share of total transactions and the average subsidy across all transactions, are normalized by the total quantity of electric vehicles in a zip*quarter. This creates a structural endogeneity between the error term and the dependent variable, most clear in regressions where the dependent variable is log of total transactions. When constructing an instrument, the relevant exclusion restriction is that the error term is uncorrelated with the instrument. A necessary condition for the exclusion restriction to hold is that contemporaneous quantities in a zip code does not enter the construction of the instrument, either directly or indirectly.

Formally, denoting the number of post-period quarters as T, the quarter in which the EFMP program becomes active as $t*$ and the average number of transactions in zip z in quarter t as $Q_{zt} = \sum_i \mathbf{1}(zip = z, time = t)$, we construct our preferred instrument for EFMP-share as:

$$PreferredIV_{zt} = \frac{\sum_i \mathbf{1}(Subsidy_{izt} > 0, zip = z, time = t)}{\frac{\sum_{r \neq t, r \geq t*} Q_{zr}}{T-1} \frac{\sum_{x \neq z} Q_{xt}}{\sum_{r \geq t*} \sum_{x \neq z} Q_{xr}/T-1}} \tag{16}$$

The numerator of the instrument is identical to the numerator of EFMP-share. On the other hand, the first term in the denominator is the average number of total transactions in zip z in the post period, leaving out the current period. This captures largely cross-sectional variation across zip codes reflecting how many EVs are typically purchased in a location. The second term is the ratio of contemporaneous sales in all other zip codes in the district, to the average sales in all quarters except this one. This largely captures time-series variation with regards to EV sales in the air district. Note that this instrument excludes contemporaneous quantities in a zip code at time t. Absent autocorrelation or spatial correlation of preferences, which would lead contemporaneous quantities in a zip code to be either correlated with the former or latter, respectively.

We also construct three alternative instruments. The first two relax the assumptions of spatial correlation and autocorrelation of preferences, respectively. Formally,

$$AlternativeIV1_{zt} = \frac{\sum_i \mathbf{1}(Subsidy_{izt} > 0, zip = z, time = t)}{\frac{\sum_{r \neq t, r \geq t*} Q_{zr}}{T-1}} \tag{17}$$

$$AlternativeIV2_{zt} = \frac{\sum_i \mathbf{1}(Subsidy_{izt} > 0, zip = z, time = t)}{\frac{\sum_{x \neq z} Q_{xt}}{\sum_{r \geq t*} \sum_{x \neq z} Q_{xr}/T-1}} \tag{18}$$

Alternative IV 1 is identical to our preferred instrument, but excludes the time-series variation provided by average EV sales of other zip in district. If we worry that spatial correlation of sales invalidates our preferred instrument, alternative IV 1 does not rely on contemporaneous sales at all. In a similar fashion, alternative IV 2 excludes the cross-sectional variation provided by the average sales in the zip leaving out contemporaneous sales, allowing for autocorrelation in sales.

Finally, the third alternative instrument is a traditional shift-share instrument, interacting cross-sectional variation in the fraction of households in the zip code below 225% of the federal poverty line with time-series variation in either state-wide EFMP share or state-wide mean EFMP subsidy.

A.3 Income Distribution Estimation

Section 2 describes the EFMP the structure of discontinuities in income related to the EFMP's means-testing incentive structure, which for our regression discontinuity design, requires estimating the proportion of the population around the means-tested discontinuity. Below is the generalized equation for estimating the treatment effect at the discontinuity,

$$\tau_c = \underbrace{E\left[P_{c_j,1} - P_{c_j,0} \middle| Inc_i \sim c_j\right]}_{\text{treatment effect}} \underbrace{Pr\left[Inc_i \sim c_j\right]}_{\text{population weight}},$$

where c_j is the discontinuity in income leading to different subsidy levels, and Inc_i is the income of individual i around the discontinuity.

As researchers we cannot observe micro-level data of the proportion of individuals around the discontinuity at the same geographic resolution[26] as the program, and thus we have developed a method of using Census data to approximate tract-level income distributions. Following Salem and Mount's (1974)[27] use of the lognormal and generalized gamma distributions, we construct income distributions for each census tract using median income and its standard deviation from data provided by the Census Bureau. Then from the income distributions we then are able to calculate the population weight for the RDD.

A.3.1 Data

The data comes from the American Census Bureau's 5-year American Community Survey (ACS). The primary data set is census tract-level median income and the standard deviation for households separated by the number of occupants.[28] The data is primarily drawn from the 2010 ACS, which coincides with the decennial Census survey, providing coverage of 98.9% of

[26]We observe the subsidy and sales at the ZIP- and Census tract-level.

[27]Salem, Ali BZ, and T. D. Mount. "A convenient descriptive model of income distribution: the gamma density." *Econometrica: journal of the Econometric Society* (1974): 1115-1127.

[28]The set of household occupant sizes are separated into the set $\{1, 2, 3, 4, 5, 6, 7+\}$.

census tracts in California. Data will be compared to the 2015 ACS data, however 51.5% of data is missing for the 2015 vintage.

Furthermore, additional data will be included as constraints or tests, including the number of households in different regions of the federal poverty level (FPL), e.g. 100-125%, and the number of households in different income ranges, e.g. $40,000-$44,999.

Lastly, the Census Bureau releases the Integrated Public Use Microdata Series (IPUMS) at the county-level, in which we observe the true mass of individuals that are from 0%-500% of FPL. This data is used to compare the estimation procedures for the lognormal and generalized gamma distributions, which is necessary since we do not observe micro-level data at a higher resolution than the county-level.

A.3.2 Estimation

The estimation procedure occurs in three separate work flows to produce tract-level income distributions. First, the Census ACS data is collected and transformed from $2011 US Dollars to a corresponding percent of the FPL, and the IPUMS data is collected and binned by household size. Starting with a general check, we take each county within our study region and run a maximum likelihood routine to estimate the the lognormal and generalized gamma distributions on the IPUMS data. These distributions are used to determine the precision of fit, and to be used later to check the tract-level estimates.

Second, estimating tract level distributions for the the lognormal, we use the delta method, however the estimation procedure described below will be used to estimate the lognormal if necessary.

The last stage of the estimation procedure follows in four steps. (1) Using Salem and Mount (1974) we take the tract-level estimates and use a basic heuristic to transform the median and standard deviation parameters into a guess of the two (shape and scale) parameters for the generalized gamma distribution. (2) With the initial guess, we then take 10,000 random draws from the generalized gamma distribution and use a generalized method of moments (GMM) estimator with the moment conditions,

$$E\begin{bmatrix} med\left[ACS_i\right] - \hat{med}\left[\gamma_n(\alpha_n, \lambda_n)\right] \\ Var\left[ACS_i\right] - \hat{Var}\left[\gamma_n(\alpha_n, \lambda_n)\right] \end{bmatrix} = \mathbf{0}.$$

With $med\left[ACS_i\right]$ and $Var(ACS_i)$ as the median and variance estimates from the Census Bureau, and $\gamma_n(\cdot)$ as the iterate value $n \in \{0, 1, \ldots, N\}$ of the generalized gamma distribution. Then, (3) the previous step is repeated until the estimates of the shape and scale converge to a desired tolerance. The GMM procedure estimates the distribution function for all census tracts jointly to increase the stability of estimation. Lastly, (4) the distribution function is used to estimate the number of individuals at the tract-level whom possess income levels in the tiers of the EFMP.

Once all census tracts have been estimated, then we will use the additional data to estimate accuracy. One concern is that we cannot directly observe the income distribution at each census tract, and moreover do not have overlapping data in which to use as direct constraints on the estimation procedure. For example, at each tract we know the number of people in different regions of the FPL and the number of households in different income brackets, however we do not know the number of occupants of the households. Therefore, the data will be tested against the companion data sets, but techniques have not yet been developed to directly incorporate the constraints.

Once distributions have been estimated and checked against related data, we will then possess weights to precisely estimate the treatment effects across all the discontinuities in the Enhanced Fleet Modernization Program.

A.4 Subsidy Bill Calculation Details

In this section we describe how the subsidy bill estimates are calculated. First we calculate a net-of-subsidy growth rate of EVs in California and use this trend as a guide for what may happen in the absence of future subsidies. To the extent the projected cumulative EV registration count in 2025 using the net-of-subsidy growth rate falls short of 1.5 million, demand for these cars must be stimulated via subsidies. We use the mass-market demand elasticity estimates that are the central contribution of this paper to retrieve the subsidy bill that would allow California to reach the 1.5 million EVs by 2025 goal.

Table A2: EV Growth Rates: Subsidy and Net-of-Subsidy Estimates

Year	Cumulative EVs	Raw Growth Rate	Net-of-Subsidy Sales Estimate	Net-of-Subsidy ("Base") Growth Rate
2013	42,545		20,062	
2014	102,030	82.3%	28,050	65.9%
2015	164,247	46.7%	29,338	28.8%
2016	239,412	37.2%	35,444	21.6%
2017	333,114	32.7%	44,185	18.5%

Note: Net-of-subsidy sales calculated assuming a subsidy elasticity of -3.9, $10,000 in subsidies.

To calculate the baseline EV growth rate we begin with data on EV registration growth in California from 2013-2017, which is shown in Table A2 (the "Cumulative EVs" column). This growth combines a baseline growth rate (that would have occurred in the absence of subsidies) and the incremental demand that was induced by subsidies. The column "Net-of-Subsidy Sales Estimate" reflects our estimate of no-subsidy sales. For the purposes of this calculation we consider California and federal EV credits and rebates, which sum to $10,000 for most EVs during this period, and assume complete pass-through to consumers.[29] We then assume a subsidy elasticity of demand of -3.9 (our preferred estimate from this paper) and apply all of this to a $35,000 new EV price, reflecting the fact that the vast majority of EVs sold through 2017 were new. This allows us to net out subsidy-induced growth from baseline growth. Finally, we assume that 10 percent of the EV fleet is removed from the California fleet (e.g. via retirements and exports) each year beginning in 2020.

For the purposes of projecting the baseline growth rate into the future, it is natural to expect that it will continue to decline. This is consistent with an increase in the absolute number of EVs sold that is compared to an increasing cumulative fleet size. Since the rate of decline in the growth rate is not knowable, we present subsidy bill estimates for baseline growth rates ranging from 10 to 16 percent. Table 6 reflects the importance of this key parameter in determining the subsidy bill: moving from a 14 to 10 percent baseline growth rate more than doubles the required subsidy bill.

[29]There were other monetary and non-monetary subsidies during this period as well, including the potentially-large ZEV mandate credits. Quantifying these is difficult, and omitting them from our baseline growth rate calculation will have the effect of biasing the estimate of this rate upwards.