Fixed Effects Regression Models

for

"Penalties for Speeding and their Effect on Moving Violations: Evidence from Quebec Drivers"

Vincent Chandler
Université du Québec en Outaouais

Lealand Morin
College of Business, University of Central Florida

Jeffrey Penney
University of Alberta

August 13, 2021

Introduction

I rearranged the raw data to create a new version of the dataset to allow us to address some of Arthur's comments. In particular, it allows us to estimate models with driver-specific fixed effects, even though it eliminates the effects of characteristics of the drivers, which we can show later in the empirical analysis. I propose that these results replace the first round of (misspecified) pooled regressions and the regressions by age group. This set of models presents a simpler analysis (for the reader) since the demerit point balances are the only remaining variables that are not annihilated by the fixed effects.[1]

Since the only possible variables are the demerit point categories and the interactions with the policy change, this analysis quickly establishes the result that the policy had a downward effect on the incidence of tickets but it was not statistically significant for the sample of all drivers. In contrast, for the sample of high-point drivers, the effects were all found to be significant. After accounting for individual driver-specific fixed effects, the results were not materially different by gender. However, tests of the difference between parameters for males and females yielded different results. The F-statistic strongly supports differences in parameter values by gender. So, from this point on in the analysis, we can still model separately by gender, which might very well be captured in the fixed effects but not in the other models.

Another consequence of these results, however, is that we must include or, at least, test policy interactions with demerit points groups in the regressions that follow. One way to do this is to create indicators for zero to six, seven to ten and ten and above categories for the full sample (see Figure 1). My plan is that we show tables of results that are similar to those we had in the last draft but we augment the list of control variables with any additional variables that show up in a simple model selection procedure.

I took Arthur's advice and split the dataset into training and testing samples, selecting by the driver serial number. I didn't use the testing sample for the fixed effects regressions because most of the predictive value of the model comes from the driver-specific fixed effects, so it doesn't seem appropriate there. This made the sample slightly smaller and more manageable to estimate several million parameters, of which the fixed effects regressions were calculated using the Frisch-Waugh-Lovell theorem in first-stage regressions (read: deviations from average by driver).

One more change that I made was that I changed the definition of the drivers with high balances. When I plotted the first versions of this graph, I saw a "missing tooth" in the 6- to 10-point range from our selection criteria, which was done using the sample before the policy change. Partly to satisfy Arthur's enthusiasm for the high-point drivers, I changed this definition to include drivers who have ever had a balance 6 points or higher during the period 01 April 2002 to 31 March 2006, which is the four-year period before the sample, so that none of the tickets that were counted in those balances were included in the dataset. I think this is an adjustment that had to be made, given what the old version of Figure

---

[1] To avoid the problem with the uninteresting age variation, we redefined the age categories to represent the age of a driver on the date of the policy change.

1 looked like in this model. It also mitigates Arthur's concern about selecting subsamples based on endogenous variables: we select based on past values before the sample period.

Overall, I think our new abstract should say something to this effect: "We find that the policy worked well overall, the effects were larger for the drivers most likely to engage in excessive speeding: males more than females, younger more than older drivers, drivers with prior demerit points more than those without, and especially for the drivers who have had high demerit-point balances." I think our model is still appropriate. We explicitly model it with gender labels but the results still hold if we were to swap "M" and "F" with labels for "young" and "old" or prior balance but those variables have more categories and would be more difficult to model; however, the intuition is the same.

Fixed Effects Regressions

Our dataset includes a small set of explanatory variables, all of which are categorical. Also, the dataset is very large: it comprises several billion driver days. As a consequence, many interesting and suitable modeling approaches would be computationally burdensome or infeasible. On the other hand, the dataset is very sparse, which lends itself to methods of aggregation and econometric models that can be computed with frequency-weighted observations. Drivers rarely get tickets, so that most observations represent zero tickets, resulting in many thousands of equivalent observations with many similar drivers receiving no ticket on a given day. Furthermore, we observe many instances in which several drivers get tickets with the same point value on a particular day, all of whom are in a single category, e.g. all males, aged 20-24, with two demerit points on their driving record.

In a later section, we will analyze the data after aggregating across individuals, by grouping over the date. In this format, the observations for each day will include a listing of the number of drivers with identical characteristics who get a ticket of the same point value, if any, on each day. This allows for an analysis of the changes in driving behaviour with respect to the characteristics of individual drivers, such as age and gender, which turn out to be important. This aggregation method also retains the effects of seasonality in the form of monthly and weekday indicators.

In contrast, our first approach aggregates over time and groups the data by individual driver number. With this data, we estimate fixed effects for the individual drivers, which accounts for all the variation explained by age, sex and other unobserved heterogeneity between the drivers. The demerit point level is the only remaining variable that changes over time for a particular driver. This categorical variable is defined as the sum of all points on tickets that a driver has received within the last two years. It is designed to reflect the demerit points that remain on a particular driver's record, which is taken into account when this balance reaches thresholds to warrant suspension or revocation of the driver's license. We calculate this balance including the two-year period before the sample start date of 01 April 2006, to ensure that the measurement is consistent across drivers and throughout the sample.

As an example, consider a driver who gets one ticket during the sample on day 101 of the

Sample	All Drivers		Male Drivers		Female Drivers	
Estimate	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Intercept	0.000	0.000	0.000	0.000	-0.000	0.000
policy	0.135	3.963	0.165	2.102	0.104	1.816
Demerit points group indicators:						
1 points	-0.580	2.277	-0.519	1.209	-0.666	1.044
2 points	-0.670	2.239	-0.616	1.187	-0.744	1.026
3 points	-0.673	2.279	-0.626	1.209	-0.741	1.044
4 points	-1.238	2.291	-1.146	1.215	-1.421	1.050
5 points	-1.232	2.290	-1.133	1.215	-1.450	1.049
6 points	-1.322	2.359	-1.239	1.251	-1.514	1.081
7 points	-1.683	2.319	-1.566	1.230	-2.034	1.062
8 points	-1.787	2.322	-1.686	1.232	-2.089	1.064
9 points	-1.696	2.375	-1.646	1.260	-1.764	1.088
10 points	-2.183	2.308	-2.066	1.225	-2.651	1.058
11-20 points	-2.517	2.321	-2.419	1.231	-2.916	1.063
21-30 points	-2.713	2.309	-2.626	1.230	-3.257	1.056
31-150 points	-2.054	2.370	-1.996	1.318	-2.102	1.064
Policy and points group interactions:						
1 points	-0.293	4.221	-0.334	2.239	-0.244	1.935
2 points	-0.236	4.202	-0.269	2.228	-0.201	1.926
3 points	-0.243	4.214	-0.266	2.235	-0.229	1.931
4 points	-0.408	4.264	-0.447	2.262	-0.355	1.954
5 points	-0.451	4.241	-0.510	2.249	-0.338	1.943
6 points	-0.472	4.303	-0.517	2.282	-0.394	1.971
7 points	-0.693	4.241	-0.745	2.249	-0.559	1.943
8 points	-0.692	4.239	-0.746	2.248	-0.558	1.943
9 points	-0.786	4.282	-0.834	2.271	-0.698	1.962
10 points	-0.636	4.212	-0.703	2.234	-0.375	1.930
11-20 points	-1.100	4.198	-1.167	2.226	-0.803	1.923
21-30 points	-2.233	4.136	-2.285	2.198	-1.876	1.893
31-150 points	-4.655	4.184	-4.669	2.284	-5.062	1.895
Drivers	1,928,782		1,236,868		691,915	
Driver days	5,589,694,528		3,042,120,572		2,547,573,956	
SSR	2,754,998		1,949,301		805,694	

Tab. 1: Fixed effects regression models, with CRVE (all demerit-point levels)
Fixed effects regression coefficients after estimating driver-specific intercept coefficients. Samples are drawn by randomly selecting seventy percent of the drivers. Standard errors were calculated using the cluster-robust covariance matrix estimator, clustering on the individual driver. For readability, all coefficients and standard errors were multiplied by 1,000.

sample period. This driver would have 100 days with zero demerit points, followed by one ticket with two points, and carry that balance of two points for 730 days, until the driver's point balance reverts to zero for the remainder of the sample. This driver will be recorded with zero demerit points and zero tickets with a frequency of 100, at the beginning of the sample. This is followed by one observation with zero demerit points and a two-point ticket. The driver then has two demerit points but no ticket observed with frequency 730.[2] Finally, the driver again has zero demerit points and zero tickets for the 630 days in the remainder of the sample. The calculation is slightly more complicated for drivers who receive several tickets over the sample, except that the observations are split into more distinct categories. In all cases, the demerit point balance is represented by a step function with many repeated observations. We record demerit points in integers from zero to ten and collect the drivers with higher balances into categories of 11-20, 21-30, and 30 or more demerit points.

All Drivers (with both high and low past demerit points)

We estimated the fixed effects model with indicators for each demerit point category and the interactions of these demerit point indicators with the period after the policy change. We performed the estimation using a training sample of seventy percent of the drivers in the sample.[3] The estimation results are shown in Table 1. Note that the indicator is missing for the category of drivers with zero demerit points, which is the benchmark group.

The coefficients on the indicators for demerit point categories are largely insignificant for all samples considered in Table 1, along with the coefficients for the interaction terms. The point estimates of these effects, however, are similar for male and female drivers, and the coefficients are largely increasing in the number of demerit points.

We calculated the F-statistic to test the restriction that the parameters are the same for both male and female drivers. The unrestricted sum of squared residuals was 2,754,995. The restricted sum of squared residuals was 2,754,998. The value of the F-statistic was 217, which corresponds to a p-value of nearly zero, since the F-statistic is much higher than the one percent critical value of 1.7242. This strongly suggests that male and female driving behaviour should be modeled separately.

Aside from the coefficients taken separately, the net effect of the policy indicator and the interaction of the policy indicator with the points group indicator is perhaps of more

---

[2] Strictly speaking, we also include zero-point dummy events in the sample to separate the number of days within certain important thresholds in the data. These events are excluded from the analysis but are used to indicate the dates 01 April 2006, 01 April 2008 and 31 March 2010. Using these markers on these dates ensures that the number of driver days are aggregated separately across different periods in the sample, to impose an equal number of days in the two-year periods before and after the policy change on 01 April 2008. The result is that the driver spends the first 265 of 730 days with two demerit points and zero tickets in the pre-policy period, and the remaining 465 days with two demerit points and zero tickets in the post-policy period.

[3] The remaining sample is reserved as a testing sample for further model specification searches, for the models with other variables besides the demerit point balances. Since the fixed effects represent most of the predictive value, and are only estimated for the particular drivers in the training sample, we did not validate the fixed effects model out of sample.
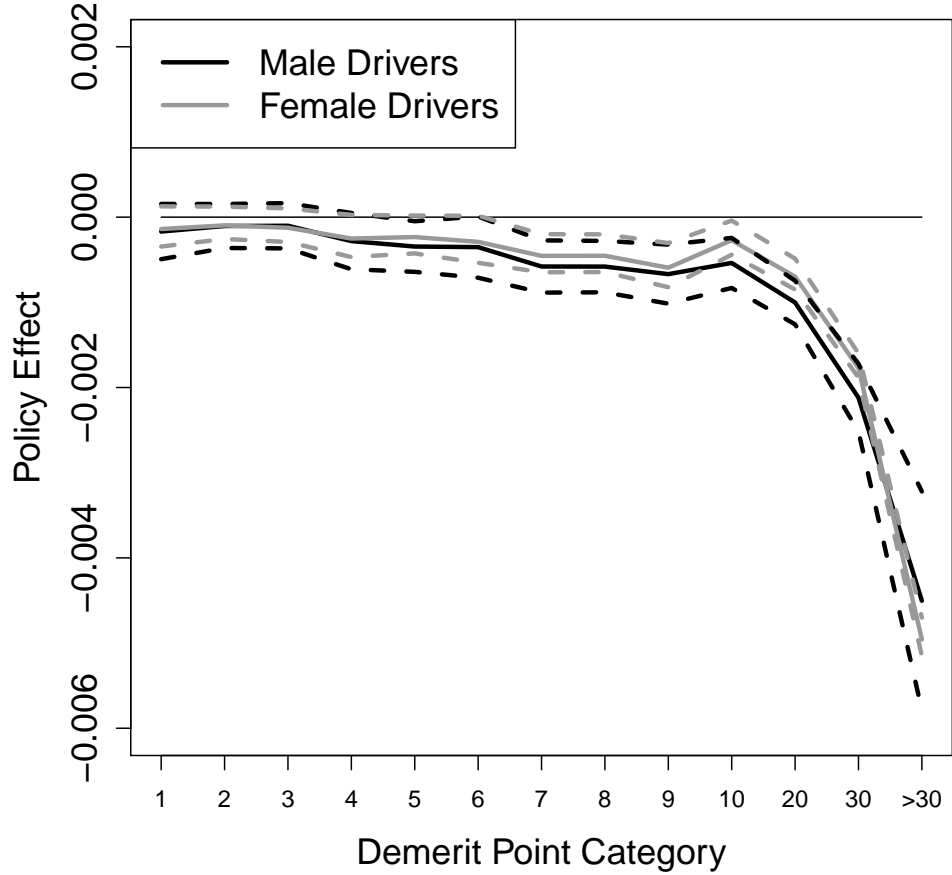
Fig. 1: Policy change and demerit points group interactions (all drivers). Solid lines represent the policy effect for drivers for each demerit-point category, calculated as the sum of the policy indicator and the interaction of the indicators for the policy change and the demerit point category. Males drivers are represented by black lines and female drivers with grey. Dashed lines represent 95% confidence intervals, calculated using cluster-robust standard errors, clustered on individual drivers. All estimates were calculated by fitting a fixed effects regression model, with intercept coefficients for each driver.

importance. In Figure 1, we plotted the sum of the coefficients for these interactions and the policy indicator for male and female drivers. We also plot 95% confidence intervals for each sum of coefficients, for which the standard errors are calculated using the standard formula to take into account the variance of both coefficients and the covariance of the two. The standard errors of the individual coefficients are calculated with the cluster-robust variance estimator, in which the clusters are taken to be the individual drivers.

For the net of the policy effect and the interaction, the relationship appears more clear. Not much difference is apparent, whether for male or female drivers, with six or fewer demerit points, and the effect is only slightly stronger for drivers with between seven and ten demerit points. For drivers with more than ten demerit points, however, the relationship is much more pronounced. Drivers with ten to twenty demerit points are expected to get two fewer tickets per thousand driver days under the excessive speeding legislation. For drivers with more than thirty demerit points, this number is as high as four tickets per thousand days—slightly more than one ticket per year—after the policy change.

To compare the difference in statistical significance with and without the cluster-robust standard error estimates, refer to the estimates in Table 2. The standard errors calculated without clustering are much smaller, indicating that there is substantial dependence on the particular driver. In other words, when considering the entire sample of drivers, the assumption of independence of trials between drivers is not supported by the data.[4]

---

[4] For this reason, when we analyze the data aggregated over drivers by time period, we will use elevated levels of significance. Because we're smart.

Sample	All Drivers		Male Drivers		Female Drivers	
Estimate	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Intercept	0.000	0.000	0.000	0.000	-0.000	0.000
policy	0.135	0.001	0.165	0.001	0.104	0.001
Demerit points group indicators:						
1 points	-0.580	0.005	-0.519	0.007	-0.666	0.006
2 points	-0.670	0.002	-0.616	0.003	-0.744	0.003
3 points	-0.673	0.002	-0.626	0.003	-0.741	0.003
4 points	-1.238	0.005	-1.146	0.007	-1.421	0.007
5 points	-1.232	0.004	-1.133	0.005	-1.450	0.006
6 points	-1.322	0.004	-1.239	0.006	-1.514	0.007
7 points	-1.683	0.007	-1.566	0.009	-2.034	0.013
8 points	-1.787	0.007	-1.686	0.008	-2.089	0.012
9 points	-1.696	0.007	-1.646	0.009	-1.764	0.013
10 points	-2.183	0.011	-2.066	0.013	-2.651	0.023
11-20 points	-2.517	0.007	-2.419	0.008	-2.916	0.015
21-30 points	-2.713	0.026	-2.626	0.030	-3.257	0.093
31-150 points	-2.054	0.061	-1.996	0.071	-2.102	0.250
Policy and points group interactions:						
1 points	-0.293	0.006	-0.334	0.009	-0.244	0.008
2 points	-0.236	0.003	-0.269	0.004	-0.201	0.003
3 points	-0.243	0.003	-0.266	0.004	-0.229	0.004
4 points	-0.408	0.006	-0.447	0.008	-0.355	0.009
5 points	-0.451	0.005	-0.510	0.006	-0.338	0.007
6 points	-0.472	0.006	-0.517	0.007	-0.394	0.009
7 points	-0.693	0.009	-0.745	0.012	-0.559	0.016
8 points	-0.692	0.009	-0.746	0.011	-0.558	0.016
9 points	-0.786	0.010	-0.834	0.012	-0.698	0.017
10 points	-0.636	0.014	-0.703	0.017	-0.375	0.028
11-20 points	-1.100	0.008	-1.167	0.010	-0.803	0.018
21-30 points	-2.233	0.031	-2.285	0.036	-1.876	0.108
31-150 points	-4.655	0.070	-4.669	0.082	-5.062	0.285
Drivers	1,928,782		1,236,868		691,915	
Driver days	5,589,694,528		3,042,120,572		2,547,573,956	
SSR	2,754,998		1,949,301		805,694	

Tab. 2: Fixed effects regression models (all demerit-point levels)
Fixed effects regression coefficients after estimating driver-specific intercept coefficients. Samples are drawn by randomly selecting seventy per cent of the drivers. For readability, all coefficients and standard errors were multiplied by 1,000.

High-Point Drivers

We then studied the sample that is restricted to drivers with a history of getting tickets. We calculated balances of demerit points over rolling windows of two-year periods in the sample before the four-year window around the policy change. This way, we calculated the number of demerit points on each driver's record over the past two years, for all days from 01 April 2002 to 31 March 2006, which is the four-year period before the sample. Using this approach, we ensure that none of the tickets that were counted in those balances were included in the dataset for the regression model. We defined the sample of high-point drivers to include all drivers with six or more points during this pre-sample period.

Table 1 shows the regression results for the fixed effects model with cluster-robust standard errors. In this subsample, the slope coefficients are similar to those from the full sample, and the results are similar across male and female drivers. In this sample of actively speeding drivers, there is much less variation and the coefficients are more precisely estimated, as the standard errors are much smaller. Drivers tend to get fewer tickets as they accumulate demerit points and approach thresholds for harsher penalties, including license suspension. These effects are even stronger for drivers after the policy change. In contrast to the result for the sample of all drivers, in this sample the policy-demerit-point interactions are all statistically significant at conventional levels of significance. Thus, the policy effects differ significantly for drivers with demerit points than for those who currently have no demerit points but who did have many points in the past.

We calculated the F-statistic to test the restriction that the parameters are the same for both male and female drivers. The unrestricted sum of squared residuals was 619774. The restricted sum of squared residuals was 619775. The value of the F-statistic was 22, which corresponds to a p-value of nearly zero, since the F-statistic is much higher than the one percent critical value of 1.7242. This strongly suggests that male and female driving behaviour should be modeled separately.

Now consider the net effect of the policy indicator and the interaction of the policy indicator with the points group indicator, in Figure 2. First of all, the effects are very similar for males and females, aside from a slight difference among drivers with nine demerit points. Perhaps more striking is how narrow the confidence intervals are for drivers with a past record of demerit points. The policy effect is not very strong for drivers with three points or less but this relationship declines steadily for drivers with more demerit points. As with the full sample of drivers, the relationship takes a sharp downturn for drivers with more than ten points. The level of the effect is similar to that measure on the sample of drivers with all point levels: from a decrease of one ticket every three years to one ticket per year.

Again, to compare the difference in statistical significance with and without the cluster-robust standard error estimates, refer to the estimates in Table 4. For this sample, there is only slight variation in the magnitude of the standard errors, particularly for the demerit-point categories of drivers who tend to get a moderate number of tickets. The difference in standard errors only marginally affects the magnitude of the t-statistics, which are well over the critical values at conventional levels of significance. This suggests that the trials between different drivers, all of whom have had high demerit-point balances, can be treated

Sample	All Drivers		Male Drivers		Female Drivers	
Estimate	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Intercept	-0.000	0.000	0.000	0.000	-0.000	0.000
policy	0.126	0.008	0.116	0.007	0.161	0.004
Demerit points group indicators:						
1 points	-0.447	0.039	-0.444	0.035	-0.464	0.017
2 points	-0.530	0.017	-0.515	0.015	-0.591	0.008
3 points	-0.490	0.013	-0.500	0.012	-0.451	0.006
4 points	-0.911	0.028	-0.901	0.026	-0.961	0.012
5 points	-0.945	0.020	-0.920	0.018	-1.069	0.008
6 points	-0.868	0.018	-0.892	0.017	-0.769	0.008
7 points	-1.194	0.029	-1.186	0.027	-1.245	0.012
8 points	-1.356	0.027	-1.350	0.025	-1.395	0.011
9 points	-1.253	0.028	-1.299	0.026	-1.023	0.011
10 points	-1.756	0.042	-1.750	0.039	-1.810	0.014
11-20 points	-2.153	0.027	-2.144	0.025	-2.250	0.009
21-30 points	-2.378	0.123	-2.371	0.119	-2.531	0.024
31-150 points	-1.886	0.393	-1.902	0.383	-1.435	0.062
Policy and points group interactions:						
1 points	-0.321	0.048	-0.299	0.044	-0.402	0.021
2 points	-0.219	0.021	-0.218	0.019	-0.221	0.010
3 points	-0.297	0.017	-0.277	0.015	-0.380	0.007
4 points	-0.587	0.036	-0.572	0.033	-0.645	0.015
5 points	-0.553	0.026	-0.556	0.024	-0.530	0.011
6 points	-0.786	0.027	-0.762	0.024	-0.874	0.011
7 points	-1.042	0.041	-1.031	0.037	-1.094	0.016
8 points	-1.019	0.038	-1.009	0.035	-1.058	0.014
9 points	-1.300	0.042	-1.238	0.038	-1.619	0.016
10 points	-1.232	0.056	-1.226	0.053	-1.242	0.019
11-20 points	-1.685	0.033	-1.685	0.032	-1.627	0.010
21-30 points	-3.123	0.150	-3.120	0.145	-3.013	0.031
31-150 points	-5.586	0.502	-5.547	0.486	-6.534	0.092
Drivers		265,083		214,936		50,147
Driver days		387,028,979		313,813,645		73,215,334
SSR		619,775		530,409		89,366

Tab. 3: Fixed effects regression models, with CRVE (drivers with high demerit-point balances)

Fixed effects regression coefficients after estimating driver-specific intercept coefficients. Samples are drawn by randomly selecting seventy percent of the drivers. Standard errors were calculated using the cluster-robust covariance matrix estimator, clustering on the individual driver. For readability, all coefficients and standard errors were multiplied by 1,000.
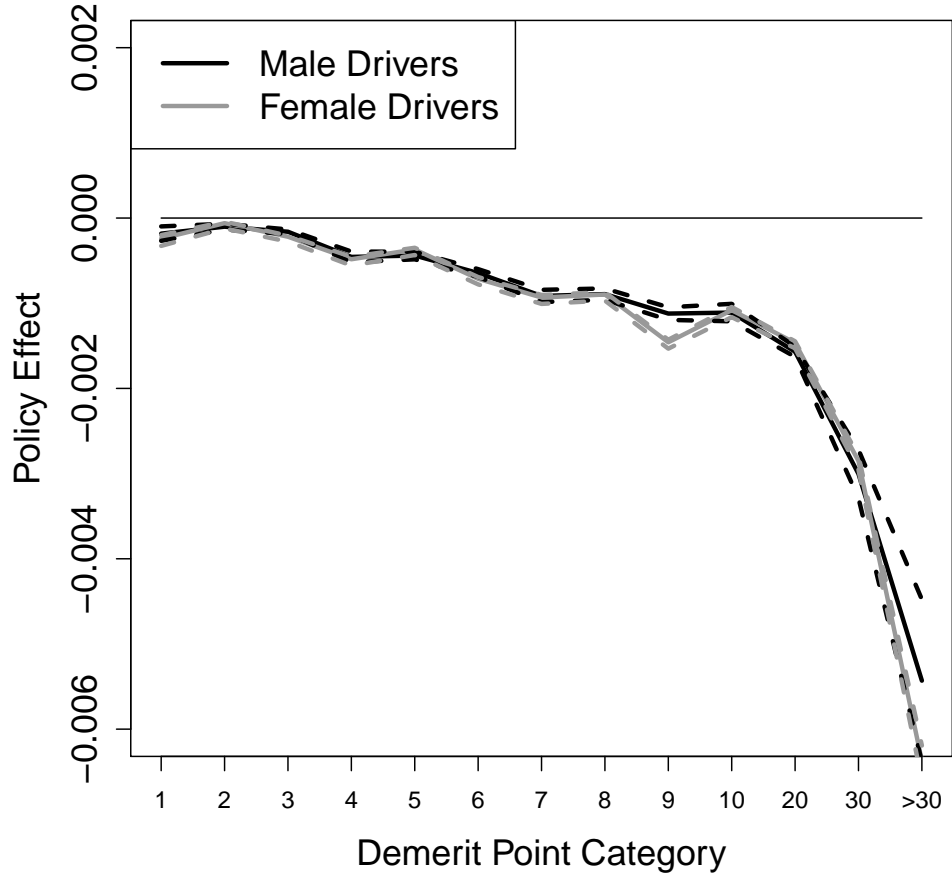
Fig. 2: Policy change and demerit points group interactions (drivers with high demerit points). Solid lines represent the policy effect for drivers for each demerit-point category, calculated as the sum of the policy indicator and the interaction of the indicators for the policy change and the demerit point category. Males drivers are represented by black lines and female drivers with grey. Dashed lines represent 95% confidence intervals, calculated using cluster-robust standard errors, clustered on individual drivers. All estimates were calculated by fitting a fixed effects regression model, with intercept coefficients for each driver.

Sample	All Drivers		Male Drivers		Female Drivers	
Estimate	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Intercept	-0.000	0.002	0.000	0.002	-0.000	0.004
policy	0.126	0.007	0.116	0.008	0.161	0.013
Demerit points group indicators:						
1 points	-0.447	0.027	-0.444	0.031	-0.464	0.056
2 points	-0.530	0.012	-0.515	0.014	-0.591	0.024
3 points	-0.490	0.010	-0.500	0.011	-0.451	0.019
4 points	-0.911	0.019	-0.901	0.021	-0.961	0.038
5 points	-0.945	0.014	-0.920	0.015	-1.069	0.028
6 points	-0.868	0.013	-0.892	0.014	-0.769	0.025
7 points	-1.194	0.019	-1.186	0.021	-1.245	0.041
8 points	-1.356	0.017	-1.350	0.019	-1.395	0.037
9 points	-1.253	0.018	-1.299	0.020	-1.023	0.038
10 points	-1.756	0.025	-1.750	0.028	-1.810	0.064
11-20 points	-2.153	0.016	-2.144	0.017	-2.250	0.041
21-30 points	-2.378	0.053	-2.371	0.056	-2.531	0.223
31-150 points	-1.886	0.122	-1.902	0.128	-1.435	0.569
Policy and points group interactions:						
1 points	-0.321	0.036	-0.299	0.041	-0.402	0.071
2 points	-0.219	0.016	-0.218	0.019	-0.221	0.031
3 points	-0.297	0.014	-0.277	0.016	-0.380	0.027
4 points	-0.587	0.025	-0.572	0.028	-0.645	0.051
5 points	-0.553	0.018	-0.556	0.021	-0.530	0.039
6 points	-0.786	0.019	-0.762	0.021	-0.874	0.040
7 points	-1.042	0.027	-1.031	0.030	-1.094	0.061
8 points	-1.019	0.025	-1.009	0.027	-1.058	0.057
9 points	-1.300	0.027	-1.238	0.030	-1.619	0.064
10 points	-1.232	0.035	-1.226	0.039	-1.242	0.091
11-20 points	-1.685	0.020	-1.685	0.022	-1.627	0.056
21-30 points	-3.123	0.068	-3.120	0.072	-3.013	0.289
31-150 points	-5.586	0.150	-5.547	0.157	-6.534	0.744
Drivers	265,083		214,936		50,147	
Driver days	387,028,979		313,813,645		73,215,334	
SSR	619,775		530,409		89,366	

Tab. 4: Fixed effects regression models (drivers with high demerit-point balances)
Fixed effects regression coefficients after estimating driver-specific intercept coefficients.
Samples are drawn by randomly selecting seventy per cent of the drivers. For readability,
all coefficients and standard errors were multiplied by 1,000.

as independent trials with little loss of accuracy in this subsample. This suggests a lower degree of heterogeneity among drivers who tend to get tickets.