UNIVERSIDAD DE SAN ANDRÉS

DEPARTAMENTO DE ECONOMÍA

MAESTRÍA EN ECONOMÍA

Assessing the impact of commuting on family violence

GASTÓN GARCÍA ZAVALETA

40915973

Mentor: Martín ROSSI

Buenos Aires,

19 de Junio, 2022

Tesis de Maestría en Economía de

Gastón GARCÍA ZAVALETA

"Evaluación del impacto de la duración del viaje al trabajo en la violencia familiar"

Resumen

Estudio el impacto de la duración del viaje al trabajo en la violencia familiar. Para evitar potenciales problemas de endogeneidad, exploto la variación exógena en el tiempo de viaje generada por los accidentes automovilísticos. Utilizando datos a nivel de condado de EE.UU., encuentro que cuando ocurre un accidente automovilístico que aumenta el tiempo de viaje disminuye el número de casos reportados de violencia familiar. Argumento que los mecanismos subyacentes son el cansancio físico y la asignación del tiempo.

Palabras clave: violencia familiar, accidentes de auto, tiempo de viaje, cansancio físico, asignación del tiempo.

"Assessing the impact of commuting on family violence"

Abstract

I study the impact of commuting time on family violence. To avoid potential endogeneity concerns, I exploit the exogenous variation in the commuting time provided by car accidents. Using US county-level data, I find that when a car accident that increases commuting time occurs, the number of family violence reported cases during the following hours decreases. I argue that the underlying mechanisms that may explain this relationship are physical tiredness and time allocation.

Keywords: family violence, commuting, car accidents, physical tiredness, time allocation.

Códigos JEL: D10, D910, R410.

1 Introduction

Commuting is a regular, unavoidable activity that absorbs an important part of personal time. According to the US Census Bureau, the average American spends 25.4 minutes commuting every day. Chatterjee et al. (2020) argue that commuting has potential objective and subjective impacts on the commuter during their journey, after their journey, and in the long term. Objective impact refers to the waste of time, the expenditure of money, and the physical effort; subjective impact is associated with the affective experience of commuting and its relationship with psychological wellbeing.

Previous studies report that commuting is the least satisfying of all daily activities (Kahneman et al., 2014; Mokhtarian et al., 2015; Lancée, Veenhoven and Burger, 2017). It also has significant spillover effects on other life domains. Time spent commuting affects time allocation in other activities. Longer commuting durations imply less time spent with family and friends (Christian, 2012). A relevant question is whether commuting only affects the quantity of these social interactions or whether it also impacts their quality. In particular, I am interested in the impact of commuting on family violence. By affecting emotions and stress, longer commuting might increase the number of reported family violence cases (Beland and Brent, 2018). However, by affecting time allocation and causing physical tiredness, it may reduce them. This paper provides evidence supporting the second hypothesis.

The literature to date on the consequences of commuting is extensive, although stronger evidence of causal relationships is needed. The impact of commuting on different life domains is hard to assess because of usual endogeneity problems. Following Beland and Brent (2018), I overcome this limitation by exploiting the exogenous variation in the commuting time provided by car accidents. According to the Bureau of Transportation and Statistics (BTS), there are approximately 18,510 car crashes per day and 6.75 million per year, only in the US. Using a countrywide traffic accident dataset (Moosavi et al., 2019), I find that the mean duration of the impact of a car accident on traffic flow is around three hours[1]. This suggests that, on average, 6.75 million times per year there is an exogenous and significant extension on someone's journey duration in the US.

[1] This is the mean value for the 1,516,064 car accidents reported in this database.

2 Data

2.1 Data on commuting

The database on commuting was obtained from the US Census Bureau and contains a summary of commuting flows between 2011 and 2015 for each combination of residence county and workplace county in the US.

A potential concern with this database is that it contains information for 2011 - 2015, while data on crime and car accidents comprises the period 2016 - 2019. Then, I must assume that commuting flows did not suffer significant changes between these two periods. Under this assumption, the average percentage of the economically active population (EAP) that commutes every day to the same place for the period 2011 - 2015 is a good proxy of the degree of exposure of a county's population to a car accident that occurs during rush hours in the road that connects it with the workplace county for the period 2016 - 2019. In order to ensure that this exposure is large enough to affect a county-level outcome (the number of family violence reported cases), I consider for my analysis only those counties where the proportion of the labor force that commutes every day to the same place is more than half. Table 1 shows that 33 counties in the US satisfy this condition.

2.2 Data on crime

The database on crime was obtained from the National Incident-Based Reporting System (NIBRS) of the Federal Bureau of Investigation (FBI) and contains high-frequency data on crime for each of the residence counties considered in the analysis. I filter family violence crimes by keeping only those where the offender and the victim are members of the same family, according to the report.

2.3 Data on car accidents

The source used for car accidents data is Moosavi et al. (2019). This is a countrywide traffic accident dataset, which covers 49 states of the United States, and currently comprises about 1.5 million accident records. I was able to find data on both car accidents and crime for 12 out of the 33 counties that satisfy the commuters over EAP restriction.[2]

[2] These counties are Bullitt (KY), Christian (MO), Dallas (IA), Effingham (GA), Fayette (TN), Fort Bend (TX), Logan (OK), Oldham (KY), Spencer (KY), Tipton (TN), Wagoner (OK), and Warren (IA).

For each of these 12 counties, I searched in Google Maps for the path that connects them with the workplace. Then, I filtered car accidents by keeping only those that occurred in any of these ways between 2016 and 2019. Finally, I kept only those affecting traffic flow between 3 PM and 7 PM. My database comprises 806 car accidents that satisfy these conditions. Figure 1 shows that most of these accidents occur between 4 PM and 6 PM, which is assumed to be the time when most of the workers return home.

A potential concern with this data is that there are cases where Google Maps provides more than one way to go from residence county to workplace. Figure 2 shows the alternative ways that connect Fayette, Tennessee, and Shelby, Tennessee. In this case, it sounds reasonable that if there is an accident in one way, commuters that have not yet started their journey could take the other. Then, car accidents may not significantly affect commuting time. For this reason, I also show results considering only those counties with a unique path that connects them with the workplace county.

3 Identification strategy and results

I am interested in estimating the impact of longer commuting durations on family violence. The identification strategy exploits the exogenous variation in the commuting time provided by car accidents that affect traffic flow during rush hours on the roads that connect residence counties with workplace counties. Formally, I estimate the following equation by OLS:

$$Y_{it} = county_i + \beta\, accident_{it} + \gamma X_t + \epsilon_{it} \tag{1}$$

where Y_{it} is the number of family violence reported cases in county i between 6 PM and 6 AM of day t,[3] $county_i$ is a county fixed effect, $accident_{it}$ is a dummy variable that takes 1 if there is at least one car accident that affects traffic flow during rush hours in the road suggested by Google Maps to go from county i to the workplace during day t, X_t is a set of controls and ϵ_{it} is an error term. Depending on the strategy and the specification, X_t includes weather controls,[4] day of the week dummies, month dummies, year dummies, or a different dummy for each particular day. The coefficient of interest is β.

Identification would be challenged in the presence of omitted variables that correlate with both car

[3] Here *day* refers to the hours following the commuting back home.

[4] Weather data was obtained from World Weather Online API, following the code provided by https://github.com/ekapope/WorldWeatherOnline. It includes sun hours, humidity, precipitations, and temperature.

accidents and family violence. For example, the weather may affect both variables, and hence its omission could bias estimates. However, the results remain robust to the inclusion of different sets of controls.

3.1 Inference Concerns

The final database comprises a panel with hourly data on crime and car accidents for different counties in the U.S. from 2016 to 2019. In this framework, it may not be reasonable to assume that observations are independent of each other, particularly those belonging to the same county. This dependence would invalidate traditional inference. A plausible remedy would be to cluster observations at the county level, allowing for dependence within county but assuming independence between observations of different counties. However, the validity of this approach requires that the number of clusters tends to infinity, while in this application the number of counties is just twelve. To deal with this concern, I will follow two alternative strategies.

Strategy 1 retains the OLS estimator for equation (1) and computes wild bootstrapped standard errors. This widely used approach, described in Cameron et al. (2008), has proven to have an acceptable performance even when the number of clusters is low. However, its theoretical validity relies on two strong assumptions: first, it requires stringent homogeneity across clusters, and second, it requires that the number of clusters tends to infinity. Both assumptions may seem inappropriate in this application, particularly the second one. For this reason, Strategy 2 applies the approximate randomization tests (ARTs) developed by Canay, Romano, and Shaikh (2017). This approach requires neither a large number of clusters nor homogeneity across clusters but requires the size of the clusters to be large and weak dependence within each cluster. These assumptions make the second strategy more suitable to the context of this study than the first one. Nevertheless, I find comparable results using both strategies.

Before moving to the estimation results, an important clarification about the second strategy should be done. By construction, ART-based confidence intervals are centered not on the estimated parameters of a full sample regression but on the simple average of the within-cluster estimates. This is the reason behind the difference between the reported coefficients for strategies 1 and 2.

3.2 Results

Table 2 reports the results for the two alternative estimation strategies, both for the full sample and the "One Way" sub-sample. For Strategy (1), I consider three different sets of controls. However, for Strategy (2), the *specific date-fixed effects* specification is not plausible, as ART-based inference requires within-cluster variability. For simplification purposes, I consider for this second strategy only the most conservative specification (i.e., the one that includes weather controls and day of the week, month, and year fixed effects).

The coefficient of interest is negative and statistically significant at the 5% level for all specifications and strategies when considering the full sample, indicating that the occurrence of car accidents that affect traffic flow during rush hours has a negative impact on the number of family violence reported cases. For the "One Way" sub-sample, the coefficient is also negative but only significant at 5% for Strategy (2) and the specific date-fixed effects specification of Strategy (1). For the rest of the specifications, it remains significant at 10%.

4 Placebo tests and robustness check

In order to ensure that results indeed have a causal interpretation, I run two placebo exercises. In the first one, I estimate equation (1) but use as dependent variable the number of family violence reported cases during different times of the day. As Table 3 shows, the only statistically significant coefficients are those associated with the crimes reported between 6 PM and 12 PM, both for Strategy 1 and Strategy 2. In other words, car accidents affecting traffic flow between 3 PM and 7 PM correlate with the number of reported family violence cases during the following hours but not with those reported during the rest of the day. Additionally, the lack of significance in the 12 PM - 6 AM specification suggests that the effect of car accidents on family violence is short-term.

In the second placebo exercise, I run seven different regressions, using as independent variable not only the accidents that occur during day t, but also three lags and leads of this variable. Column (1) in Table 4 shows the results when using Strategy 1, while Column (2) shows the results for Strategy 2. In both cases, the only significant coefficient at 5% is the one associated with accidents that occur on the same day as the reported crimes.

In addition to the placebo exercises, I run a robustness check that consists in repeating the main regressions but excluding from the analysis the last day of each month. The reason behind this check is that crime databases are usually subject to over-reporting issues at the end of the months. As Table 5 shows, results remain virtually unaffected by this transformation in the data.

5 Potential concerns and mechanisms

Previous literature suggests that causality could be reverse. For instance, Largarde et al. (2004) argue that marital separation is associated with an increased risk of a serious accident; in this case, it would be reasonable to expect a similar effect from family violence. According to this hypothesis, if there is an impact of violence on accidents, it should be positive. However, all the coefficients associated with the leads of the variable *Accident* are negative and non-significant at 5%. Even more, crimes reported in the morning show a negative and non-significant correlation with the car accidents occurring in the afternoon.

Results suggest that longer commuting time causes a decrease in family violence reported cases. These results are consistent with at least two possible mechanisms: time allocation and physical tiredness. The first one is mechanic: if people spend more time commuting, they have less time to spend with family (Cristian, 2012). Previous evidence suggests that time spent at home is positively associated with family violence (Gibbons, Murphy, and Rossi, 2021). The second one is associated with commuters' fatigue symptoms (Kageyama et al., 1998). If long commuting produces physical tiredness, it may reduce the probability of violent behaviors.

6 Conclusions

Assessing the impact of commuting on different life domains is difficult because of potential endogeneity concerns. The identification strategy I employ overcomes this limitation by using car accidents as a source of exogenous variation in commuting time. Results show a reduction in the number of family violence reported cases in the hours following a car accident that affects most commuters' journey duration when returning home. Placebo tests suggest that this result may have a causal interpretation.

My findings are opposed to the evidence provided by Beland and Brent (2018). However, this discrepancy might be due to demographic differences in the populations considered for the analysis. While the latter employs data from Los Angeles County, I use data from twelve primarily rural counties in different states in the US. This poses a challenge in terms of the external validity of these studies. Motivated by these discrepancies, further research should attempt to overcome this limitation by considering a more comprehensive sample that allows understanding the differences found in the estimated impact of commuting on family violence for populations with distinct demographic characteristics.

References

Beland, L. P., Brent, D. A. (2018). Traffic and crime. Journal of Public Economics, 160, 96-116.

Cameron, A. C., Gelbach, J. B. and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics, 90 414–427.

Canay, I. A., Romano, J. P. and Shaikh, A. M. (2017). Randomization tests under an approximate symmetry assumption. Econometrica, 85 1013–1030.

Chatterjee, K., Chng, S., Clark, B., Davis, A., De Vos, J., Ettema, D., Reardon, L. (2020). Commuting and wellbeing: a critical overview of the literature with implications for policy and future research. Transport reviews, 40(1), 5-34.

Christian, T. J. (2012). Automobile commuting duration and the quantity of time spent with spouse, children, and friends. Preventive Medicine, 55(3), 215-218.

Gibbons, M. A., Murphy, T. E., Rossi, M. A. (2021). Confinement and intimate partner violence. Kyklos, 74(3), 349-361.

Kageyama, T., Nishikido, N., Kobayashi, T., Kurokawa, Y., Kaneko, T., Kabuto, M. (1998). Long commuting time, extensive overtime, and sympathodominant state assessed in terms of short-term heart rate variability among male white-collar workers in the Tokyo megalopolis. Industrial health, 36(3), 209-217.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. Science, 306(5702), 1776-1780.

Lagarde, E., Chastang, J. F., Gueguen, A., Coeuret-Pellicer, M., Chiron, M., Lafont, S. (2004). Emotional stress and traffic accidents: the impact of separation and divorce. Epidemiology, 762-766.

Lancée, S., Veenhoven, R., Burger, M. (2017). Mood during commute in the Netherlands: What way of travel feels best for what kind of people?. Transportation Research Part A: Policy and Practice, 104, 195-208.

Mokhtarian, P. L., Papon, F., Goulard, M., Diana, M. (2015). What makes travel pleasant and/or tiring? An investigation based on the French National Travel Survey. Transportation, 42(6), 1103-1128.

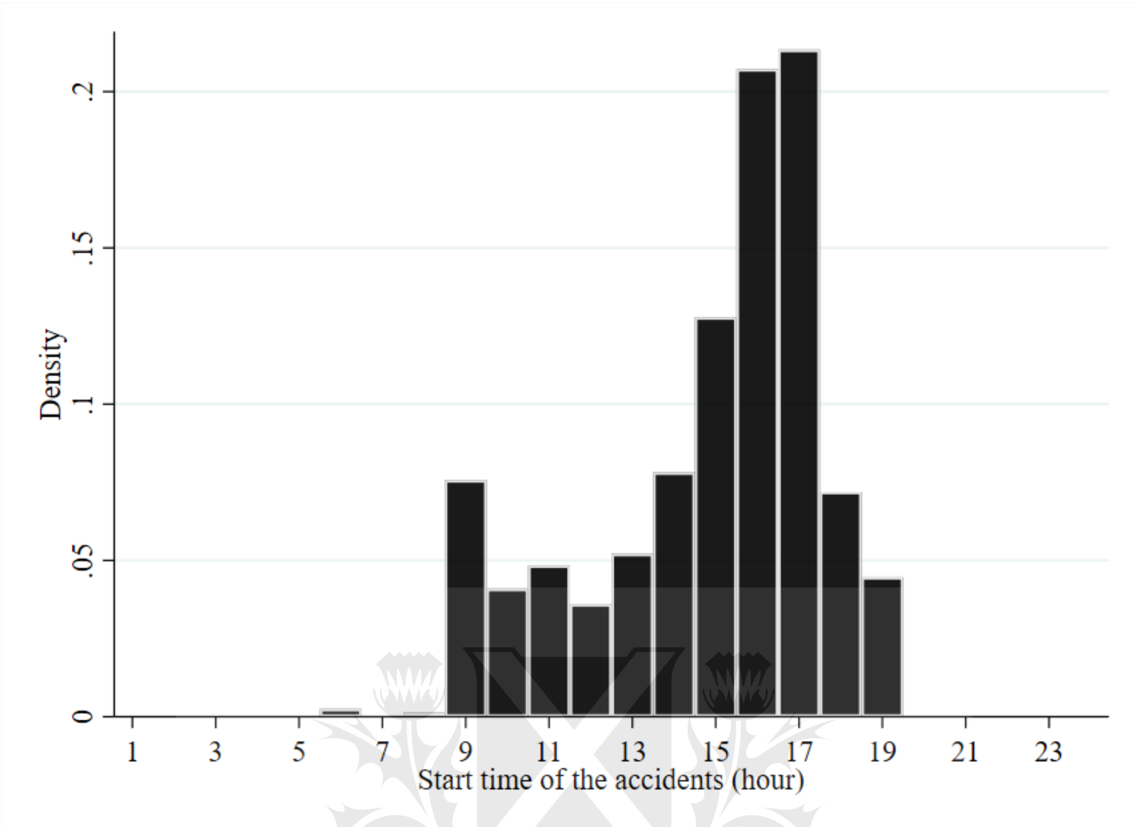
Moosavi, S., Samavatian, M. H., Parthasarathy, S., Ramnath, R. (2019). A countrywide traffic accident dataset.

Table 1: County commuting flows 2011 - 2015

	Working Place	Commuters	Labor Force	Commuters to Labor Force (%)
	(1)	(2)	(3)	(4)
Norton, Virginia	Wise, Virginia	1337	1739	76.88
Chattahoochee, Georgia	Muscogee, Georgia	1507	2087	72.20
Echols, Georgia	Lowndes, Georgia	1228	1844	66.59
Lee, Georgia	Dougherty, Georgia	8678	13967	62.13
Archer, Texas	Wichita, Texas	2500	4125	60.61
Fayette, Tennessee	Shelby, Tennessee	10734	17907	59.94
Wagoner, Oklahoma	Tulsa, Oklahoma	21524	36223	59.42
Lincoln, South Dakota	Minnehaha, South Dakota	17475	30651	57.01
Spencer, Kentucky	Jefferson, Kentucky	5233	9258	56.52
Bullitt, Kentucky	Jefferson, Kentucky	22263	39420	56.48
Oldham, Kentucky	Jefferson, Kentucky	17459	30938	56.43
Stanley, South Dakota	Hughes, South Dakota	1029	1845	55.77
Tipton, Tennessee	Shelby, Tennessee	15183	27547	55.12
Wakulla, Florida	Leon, Florida	7673	14013	54.76
Harris, Georgia	Muscogee, Georgia	8496	15520	54.74
Coryell, Texas	Bell, Texas	13420	24516	54.74
Effingham, Georgia	Chatham, Georgia	14660	26854	54.59
Russell, Alabama	Muscogee, Georgia	13016	24051	54.12
Logan, Oklahoma	Oklahoma, Oklahoma	11473	21461	53.46
Andrew, Missouri	Buchanan, Missouri	5260	9860	53.35
Lanier, Georgia	Lowndes, Georgia	1969	3740	52.65
Lonoke, Arkansas	Pulaski, Arkansas	17435	33138	52.61
Dallas, Iowa	Polk, Iowa	22927	43682	52.49
Menard, Illinois	Sangamon, Illinois	3464	6624	52.29
Warren, Iowa	Polk, Iowa	13953	26711	52.24
Edmonson, Kentucky	Warren, Kentucky	2411	4680	51.52
Sarpy, Nebraska	Douglas, Nebraska	46665	91004	51.28
Fort Bend, Texas	Harris, Texas	181752	356003	51.05
Meade, South Dakota	Pennington, South Dakota	6810	13421	50.74
Columbia, Georgia	Richmond, Georgia	34258	67851	50.49
Stanton, Nebraska	Madison, Nebraska	1769	3518	50.28
Storey, Nevada	Washoe, Nevada	931	1853	50.24
Christian, Missouri	Greene, Missouri	21447	42786	50.13

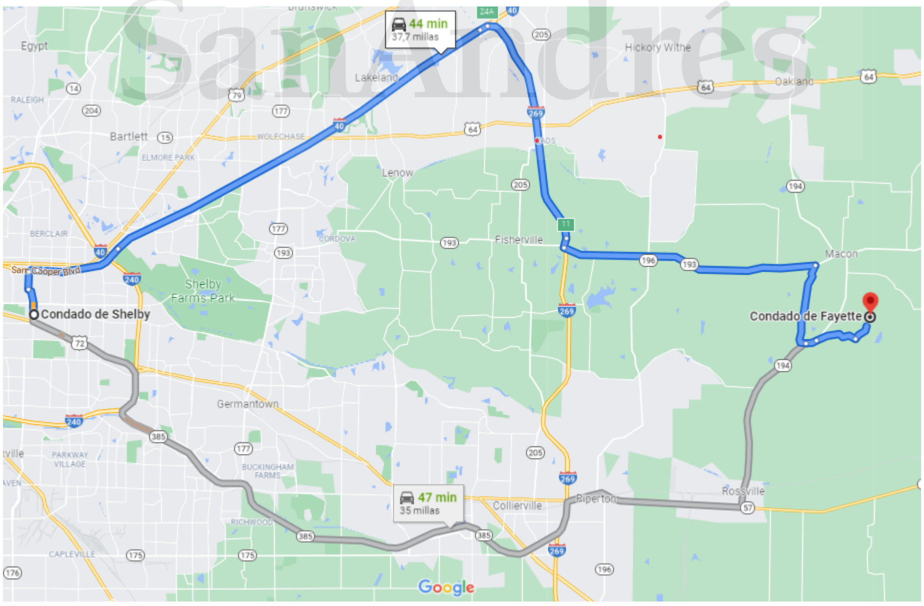
Notes: population and labor force data is from 2015. *Source:* US Census Bureau.

Figure 1: Car accidents by time of the day



Source: Moosavi et al. (2019)

Figure 2: Ways from Fayette, Tennessee to Shelby, Tennessee



Source: Google Maps. https://www.google.com/maps.

Table 2: The impact of commuting on family violence reported cases

	Family violence reported cases (6 PM - 6 AM)							
	Full Sample				One Way			
	Strategy 1			Strategy 2	Strategy 1			Strategy 2
	(1)	(2)	(3)	(1)	(1)	(2)	(3)	(1)
Accident	-0.276**	-0.274**	-0.283**	-0.084***	-0.313*	-0.311*	-0.332**	-0.097***
	(0.014)	(0.023)	(0.016)	(0.000)	(0.063)	(0.063)	(0.016)	(0.000)
County	Yes	Yes	Yes	No	Yes	Yes	Yes	No
DOW	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Month	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Year	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Weather	No	Yes	No	Yes	No	Yes	No	Yes
Specific Date	No	No	Yes	No	No	No	Yes	No
Mean of Dep. Variable	0.805	0.805	0.805	0.805	1.235	1.235	1.235	1.235
Number of Observations	14,974	14,974	14,974	14,974	8,400	8,400	8,400	8,400
R-Squared	0.413	0.414	0.447	.	0.384	0.385	0.442	.

*Significant at 10%; **significant at 5%; ***significant at 1%.
Strategy 1: estimates are computed using OLS. Wild bootstrapped p-values in parentheses.
Strategy 2: the coefficient represents the simple average of the estimated impact within each county, following Canay, Romano, and Shaikh (2017). ART-based p-values in parentheses.

Table 3: The impact of commuting on family violence by time of the day

	Family violence reported cases for different times of the day							
	6 AM - 12 AM		12 AM - 6 PM		6 PM - 12 PM		12 PM - 6 AM	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Accident	-0.056	-0.007	-0.065	0.097	-0.182**	-0.053***	-0.092	-0.031*
	(0.391)	(0.654)	(0.479)	(0.172)	(0.042)	(0.005)	(0.149)	(0.062)
Mean of Dep. Variable	0.283	0.283	0.492	0.492	0.555	0.555	0.250	0.250
Number of Observations	14974	14974	14974	14974	14974	14974	14974	14974
R-Squared	0.264	0.264	0.330	0.330	0.373	0.373	0.226	0.226

*Significant at 10%; **significant at 5%; ***significant at 1%.
All regressions control for weather conditions, day of the week, month, and year.
Columns (1): estimates are computed using OLS. Wild bootstrapped p-values in parentheses.
Columns (2): the coefficient represents the simple average of the estimated impact within each county, following Canay, Romano, and Shaikh (2017). ART-based p-values in parentheses.

Table 4: The effect of lags and leads of car accidents on family violence

	Family violence reported cases	
	(1)	(2)
Accident t-3	-0.081	0.022
	(0.162)	(1.000)
Accident t-2	-0.046	-0.034
	(0.776)	(0.360)
Accident t-1	-0.09	-0.023
	(0.806)	(0.580)
Accident t	-0.274**	-0.084***
	(0.023)	(0.000)
Accident t+1	-0.186	0.059
	(0.304)	(0.405)
Accident t+2	-0.165	-0.109*
	(0.109)	(0.083)
Accident t+3	-0.137*	-0.033
	(0.062)	(0.175)
Mean of Dep. Variable	0.805	0.805
Number of Observations	14938	14938
R-Squared	0.412	.

*Significant at 10%; **significant at 5%; ***significant at 1%.
All regressions control for weather conditions, day of the week, month, and year.
Column (1): estimates are computed using OLS. Wild bootstrapped p-values in parentheses.
Column (2): the coefficient represents the simple average of the estimated impact within each county, following Canay, Romano, and Shaikh (2017). ART-based p-values in parentheses.

Table 5: Robustness check - Excluding the last day of each month

	Family violence reported cases (6 PM - 6 AM)							
	Full Sample				One Way			
	Strategy 1			Strategy 2	Strategy 1			Strategy 2
	(1)	(2)	(3)	(1)	(1)	(2)	(3)	(1)
Accident	-0.274**	-0.272**	-0.283**	-0.080***	-0.317*	-0.315*	-0.340***	-0.095***
	(0.019)	(0.030)	(0.010)	(0.000)	0.063	(0.063)	(0.000)	(0.024)
County	Yes	Yes	Yes	No	Yes	Yes	Yes	No
DOW	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Month	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Year	Yes	Yes	No	Yes	Yes	Yes	No	Yes
Weather	No	Yes	No	Yes	No	Yes	No	Yes
Specific Date	No	No	Yes	No	No	No	Yes	No
Mean of Dep. Variable	0.801	0.801	0.801	0.801	1.228	1.228	1.228	1.228
Number of Observations	14482	14482	14482	14482	8124	8124	8124	8124
R-Squared	0.415	0.415	0.446	.	0.387	0.388	0.442	.

*Significant at 10%; **significant at 5%; ***significant at 1%.
Strategy 1: estimates are computed using OLS. Wild bootstrapped p-values in parentheses.
Strategy 2: the coefficient represents the simple average of the estimated impact within each county, following Canay, Romano, and Shaikh (2017). ART-based p-values in parentheses.